

## **W001: Abiotic Stress**

### **Harnessing Natural Variation to Identify Gene Variants and their Molecular Mechanisms that Confer Stress Resilience**

**Wolfgang Busch**, Salk Institute for Biological Studies, La Jolla, CA

Roots play a crucial role for plant survival and productivity. They forage the soil for nutrients and water, thereby providing the shoots with molecules that are essential for photosynthesis, growth and defense. Nutrient and water distribution in the soil fluctuates spatially and temporally, as does the presence of potential pathogens and toxic minerals. To function efficiently, roots therefore need to determine the biotic and abiotic properties of local soil environments and coordinate growth and development of the root system with nutrient and water uptake. Importantly, the optimal strategies might change in different habitats. To effectively confer increased stress resilience to plants via optimized root responses, we harness existing resilience mechanisms in natural populations of the model species *Arabidopsis thaliana* and try to understand the genetic and molecular mechanisms underlying adaptive responses to stresses. Importantly, natural variation allows for the identification of variants that already have undergone selection in the context of living organisms in their natural environments. Using a systems genetics approach that integrates high throughput phenotyping, genome wide association mapping and functional genomic approaches, we have generated a large atlas of root responses to different nutrient conditions and associated genetic variants. In this presentation, I will highlight two recent examples in which we have found genetic variants and uncovered molecular mechanisms that confer increased resistance to iron deficiency and iron toxicity respectively, as well an integrative approach for studying the basis of nutrient level interactions on root growth.

## **W002: Abiotic Stress**

### **New Tools for Dynamically Maximizing Crop Productivity**

**Sean Cutler**, University of California-Riverside, Riverside, CA

## **W003: Abiotic Stress**

### **Dissecting the QTLome for Osmotic Adjustment and Chlorophyll Fluorescence in Field Grown Durum Wheat**

Giuseppe Condorelli<sup>1</sup>, Maria Newcomb<sup>2</sup>, Eder Licieri Grolì<sup>1</sup>, Nicolas Zendonadi<sup>3</sup>, Marco Maccaferri<sup>4</sup>, Elisabetta Frascari<sup>4</sup>, Ebrahim Babaeian<sup>5</sup>, Markus Tuller<sup>5</sup>, Onno Müller<sup>3</sup>, Todd C. Mockler<sup>6</sup>, Nadia Shakoor<sup>6</sup>, Jeffrey White<sup>7</sup>, Rick Ward<sup>2</sup> and **Roberto Tuberosa**<sup>8</sup>, (1)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, (2)Maricopa Agricultural Center, University of Arizona, (3)Forschungszentrum Jülich, Germany, (4)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy, (5)Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, (6)Donald Danforth Plant Science Center, Saint Louis, MO, (7)US Arid Land Agricultural Research Center, USDA-ARS, (8)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, Bologna, Italy

Among the proxy traits involved in the adaptive response of wheat to drought and other abiotic stresses, information on the QTLome governing osmotic adjustment (OA) and chlorophyll fluorescence imaging (CFI) is very limited, mainly due to the difficulty to measure these traits on a scale suitable for GWAS. In this study, OA and CFI were measured in plots of 248 elite accessions of durum wheat grown (two replicates) under a fully automated Lemnatec Field Scanalyzer (LFS) and exposed to a progressively severe drought treatment starting 5 days prior to flowering which lasted for two weeks. Leaf relative water content (RWC), chlorophyll content (SPAD) and leaf rolling (LR) were also measured. CFI was measured (i) in the dark in four time series using the LFS camera and (ii) during daytime using a light-induced fluorescence transient (LIFT) sensor mounted on a manually pushed cart. The high variability in OA and CFI among the durum accessions resulted in high to medium repeatability ( $h^2 = 72.3$  and  $54.6\%$ , respectively). At the end of the drought treatment (mean leaf RWC =  $62.1\%$ ), OA and leaf RWC were positively correlated ( $r = 0.78$ ). Association mapping using flowering time as covariate revealed 15 QTLs for OA (global  $R^2 = 63.6\%$ ) as well as eight major QTL hotspots on chromosome arms 1BL, 2BL, 4AL, 5AL, 6AL, 6BL

and 7BS where a higher OA capacity was always positively associated with leaf RWC, SPAD and negatively associated with LR, hence indicating a beneficial effect of OA on the water status of the plant. Additionally, the comparative analysis with previous field trials showed concurrent effects for five of these hotspots on normalized difference vegetation index (NDVI), thousand kernel weight (TKW) and/or grain yield (GY), hence supporting OA as a valuable proxy for marker-assisted selection aimed at enhancing drought resistance in wheat.

#### **W004: Abiotic Stress**

##### **Overexpression of a Maize Transcription Factor in Maize Increases Grain Yield in the Field**

**Jeffrey Habben**, Corteva Agriscience, Johnston, IA

Increasing maize grain yield has been a major focus of both plant breeding and genetic engineering to meet the global demand for food, feed, and industrial uses. MADS-box transcription factors have been shown to regulate genes involved in controlling numerous plant growth and development characteristics. We have increased and extended the expression of a maize MADS-box transcription factor gene, *zmm28*, under the control of a moderate-constitutive maize promoter. The resultant transgenic maize plants have increased plant growth, photosynthesis capacity, and nitrogen utilization. Overall, these positive attributes are associated with a significant increase in grain yield relative to wild-type controls that is consistent across years, environments, and elite germplasm backgrounds. We conclude that alteration in expression of a single native maize gene in maize can improve both resource capture and resource utilization, resulting in a significant improvement in grain yield, the ultimate complex quantitative trait.

#### **W005: Abiotic Stress**

##### **Genomic and Epigenomic Bases of Transgressive Segregation: Using an Old Science to Create Novel Adaptive Phenotypes in Rice**

**Benildo G. de los Reyes**, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX

The genetic blueprint of the new generation of crops with minimal penalty to growth and productivity potentials under marginalized environments is an important question faced by agriculture in 21<sup>st</sup> century. The question pushes even further the frontier of biological complexity that modern plant breeding needs to conquer, beyond the achievements of the Green Revolution, marker-assisted selection, and transgenic technology. Genomics-enabled plant breeding must recognize that such level of complexity cannot be addressed by a reductionist approach. Any additional physiological gains similar to what has been optimized by natural selection must involve complex synergies that also require reconciliation with inevitable biological trade-offs.

Evolutionary theories recognize that genetic recombination under genome shock is an important driver of adaptive speciation, by virtue of the phenotypic novelties of rare wide-hybrids and recombinants, as also observed among transgressive segregants in plant breeding. In this presentation, the author will discuss recent findings on a transgressive population of rice for salt tolerance, to make a case that stress-adaptive developmental and physiological novelties involve intricate molecular synergies and network rewiring created by genome shock and epigenome confrontation. Modern views on the possible molecular underpinnings of transgressive phenotypes will be presented in context of the Omnigenic Theory for quantitative traits and gene regulation by DNA methylation and chromatin remodeling. Perspectives on how genomic and epigenomic modeling could harness a transgressive genome and epigenome to create the new generation of ecologically resilient crops will be presented as alternative to the more reductionist paradigms of functional genomics and genome editing.

#### **W006: Abiotic Stress**

##### **Next Generation Sequencing Technologies for Trait Dissection and Molecular Breeding Legumes**

**Annapurna Chitikineni**<sup>1</sup>, **Aamir W Khan**<sup>2</sup>, **Prasad Bajaj**<sup>1</sup>, **Vanika Garg**<sup>1</sup> and **Rajeev K Varshney**<sup>1</sup>,  
(1)ICRISAT, Hyderabad, India, (2)ICRISAT, Hyderabad, Telangana, India

Although crop improvement programs have made excellent progress in enhancing crop productivity and production, there is still a huge scope to fill the yield gap for majority of crops especially for tolerance to abiotic stress and resistance to biotic stresses in developing countries. Genomics-assisted breeding can help enhancing crop

productivity by enhancing precision and efficiency in the breeding programs. However, until recently, majority of the dryland crops have remained untouched with genomics revolution. Two key reasons for this situation include engagement of only few institutes and availability of limited resources at international level for research and development in these crops. With an objective to address these issues, the Sequencing and Informatics Services Unit, Center of Excellence in Genomics & Systems Biology (CEGSB) at ICRISAT is engaged in offering high-throughput genotyping and next-generation sequencing coupled with basic computational analysis to ICRISAT and its partners in developing countries on cost-to-cost basis. In addition, empowering national partners and knowledge dissemination in genome analysis and genomics-assisted breeding is another important activity at CEGSB. CEGSB, by providing sequencing/genotyping and computational genomics services has enabled scientists at ICRISAT and its partners to assemble genomes, catalogue genome variation in germplasm collection, trait mapping using QTL mapping, GWAS, QTL-Seq approaches and translate genome information in breeding. Furthermore, CEGSB by organizing 14 training courses, has trained 421 scientists from 14 different countries of Asia and Africa. Some examples in above mentioned areas in legumes namely chickpea, pigeonpea and groundnut will be presented.

## **W007: Advanced Computational Methods – Machine Learning, Containers, and Clouds**

### **Direct Computation on Phenotypic Descriptions for Novel Candidate Gene Prediction**

**Ian Braun** and Carolyn J. Lawrence-Dill, Iowa State University, Ames, IA

Natural language descriptions of plant phenotypes present in databases and the scientific literature are a rich source of information for biological research that seeks to untangle relationships between genes and observable phenotypes, such as plant health or size. The volume and unstructured nature of these text descriptions however necessitates a computational approach for leveraging them to predict gene-to-phenotype associations. We computationally translated descriptions of plant phenotypes into structured representations that can be processed to identify biologically meaningful associations. These representations include the EQ (Entity-Quality) formalism, which uses terms from biological ontologies to represent phenotypes in a standardized, semantically-rich format. Ontology terms are mapped to text descriptions with a combination of string-matching and word embedding algorithms. Our computationally produced representations of text descriptions also include numerical vectors, generated using either a bag-of-words approach or document embedding. We compared resulting phenotype similarity measures to those derived from manually curated data to determine the performance of each method. Computationally derived EQ and vector representations were comparably successful in recapitulating biological truth to representations created through manual EQ statement curation. Moreover, these computational methods for generating representations of phenotypes are scalable to large quantities of text because they require no human input. These results indicate that it is now possible to computationally and automatically produce and populate large-scale information resources that enable researchers to query phenotypic descriptions directly. Ongoing work to produce phenomics-focused text mining tools for the bioinformatics community and a resource for exploring the results of this work for the plant biology community is discussed.

## **W008: Advanced Computational Methods – Machine Learning, Containers, and Clouds**

### **With the Power of AI Comes Great Responsibility**

**Fernanda Foertter**, NVIDIA, Santa Clara, CA

Whole genome sequencing will become more prevalent in all applications of genomics as sequencing costs, methods and modalities improve. Fast and accurate assembly will be a necessity in order to make use of the ever increasing volume of high quality deeper sequencing data. Additionally recent developments in miniaturization of long read sequencers has given researchers the opportunity to sequence samples in situ, creating a need to compute on these data closer to the source, a workload traditionally reserved to a data center. To overcome this challenge, graphics processing units with the computational power for base calling and genome assembly are integrated into the sequencing

instruments. We will present work on how artificial intelligence (AI), software and hardware advances have sped up not only upstream analysis such as base calling and assembly, but also how these are giving rise to new tools for downstream analysis. This talk will share common misconceptions regarding deep learning, show examples of where it has been successfully applied to genomics, and tips for when and where it's appropriate to use it.

#### **W009: Advanced Computational Methods – Machine Learning, Containers, and Clouds**

##### **Predicting Gene Loss in Plants - Lessons Learned from Laptop-Scale Data**

**Philipp E. Bayer**, Jacqueline Batley and David Edwards, University of Western Australia, Perth, WA, Australia

Many biological datasets are 'laptop-scale' with less than a million rows and a few dozens columns. In this talk, aimed at biologists with minimal computational background, we show how to extract biological sense from such tables using scikit-learn and state-of-the-art gradient boosting algorithms such as XGBoost.

We discuss pitfalls and common problems when working with unbalanced biological data: over-reliance on metrics such as classification accuracy can hide important problems during model training. We show how to dig into the data and the model's predictions, how to figure out what went wrong, and how to improve the model's performance.

We then show how to use Shapley Additive Explanations to learn what the model has learned in order to learn more about the underlying mechanisms of gene presence/absence in plants.

#### **W010: Advanced Computational Methods – Machine Learning, Containers, and Clouds**

##### **Plenty of Fish - Identifying *Tilapia* Species from Images**

**Felix Shaw**, Earlham Institute, Norwich, United Kingdom

Protein production from aquaculture has steeply increased in the last three decades with tilapia (*Oreochromis* spp.) aquaculture production representing over 4.7 million tons in 2016. However the mis-identification of species has led to stock contamination and release of non-native species into water bodies across East Africa, resulting in hybridisation and local extinction of native species. Therefore species identification is critical for the sustainability of tilapia aquaculture practices. Our work seeks to identify these species from image data as well as other characteristics such as gender. It will run on a mobile phone in the field. This will allow farmers to select the correct species and sexes of fish to maximise yield whilst protecting wild populations by ensuring farmed populations may be correctly identified and properly contained.

#### **W011: African Orphan Crops**

##### **Aiming for Excellence in Training and Sustaining African Plant Breeders**

**Rita H. Mumm**, University of Illinois; African Plant Breeding Academy, Howard-Yana Shapiro, University of California, Davis, Davis, CA; Mars, Incorporated, McLean, VA, Eric Y Danquah, WACCI, University of Ghana, Accra, Ghana, Allen Van Deynze, Seed Biotechnology Center, University of California, Davis, Davis, CA, Richard Edema, MaRCCI, Uganda; Makerere University, Uganda, Enoch G. Achigan-Dako, Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Science, University of Abomey-Calavi, Abomey-Calavi, Benin, Walter P. Suza, Iowa State University, Ames, IA and Rufaro Madakadze, Alliance for a Green Revolution in Africa, Nairobi, Kenya

To achieve food and nutritional security and foster economic growth in Africa, a strong, innovative, multi-sector workforce is needed to collectively develop improved crop varieties that address the various demands of farmers, consumers, and other value chain stakeholders.

Preparing such a workforce starts with graduate education. In the past decade, a number of African institutions have stepped up to develop strong graduate programs in plant breeding to train Africans in Africa to stem the brain drain that has crippled the continent. These efforts have been supported through partnerships with international organizations including AGRA, World Bank, USAID SIL, the European Commission, and private sector companies.

A curriculum that builds such competency emphasizes the complex outcomes of a learning process (i.e. knowledge, skills, and attitudes to be applied by learners) rather than focusing only traditionally-defined subject matter. Apprenticeship is key to developing the “how-to” in rich detail to put concepts into action. Furthermore, external influences are important in promoting a creative mindset and in cultivating problem-solving skills, introducing students to the network of plant breeding professionals comprising a Community of Practice from which to draw immediately and into the future.

Sustaining such a workforce requires continuing education for plant breeders as scientific advancements progress at an unprecedented pace. Plant breeders must have the knowledge and ability to critically evaluate the benefit of deploying new technologies in the seed product pipeline to enhance the efficiency and effectiveness of the breeding process. Deploying technologies to maximal benefit requires access to tools for enablement.

A premier program for continuing education and professional development of African plant breeders is the African Plant Breeding Academy, organized by UC Davis. The Academy has trained 112 African scientists from 27 countries across Africa, 87% of whom are PhDs and 38% of whom are women, collectively working to improve over 105 crop species. The newly-formed African Plant Breeders Association aims to serve as an ongoing venue for continuing education and professional development. At its inaugural meeting in October 2019, the Association offered a rich program featuring innovative research as well as workshops for enhanced teaching techniques for professional development of university instructors.

The development of improved crop varieties to nourish Africa is an urgent need. The overarching response to this need can be most effectively supported and accelerated through global, multi-sector partnerships in plant breeding education and continued professional development.

## **W012: African Orphan Crops**

### **Development of a Genomics-Based Breeding and Improvement Programme for *Moringa oleifera* in Kenya**

**Prasad Hendre**<sup>1</sup>, Alice Muchugi<sup>2</sup>, David Odee<sup>3,4</sup>, Stephne Cavers<sup>4</sup>, Samuel Muthemba<sup>5</sup>, Robert Kariba<sup>5</sup>, Ann E. J. Yssel<sup>6</sup>, Huan Liu<sup>7,8</sup>, Min Liu<sup>8,9</sup>, Sunil Kumar Sahu<sup>8,9</sup>, Shiyu Chen<sup>10</sup>, Xun Xu<sup>8,11</sup>, Xin Liu<sup>8,12</sup>, Carrie Waterman<sup>13</sup>, Mike Olson<sup>14</sup>, Yves Van de peer<sup>15</sup>, Allen Van Deynze<sup>16</sup>, Anthony Simons<sup>5</sup>, Howard-Yana Shapiro<sup>10</sup> and Ramni Jamnadass<sup>17</sup>, (1)WORLD AGROFORESTRY (ICRAF), Nairobi, Kenya, (2)World Agroforestry (ICRAF), Nairobi, Kenya, (3)Biotechnology Laboratory, Kenya Forestry Research Institute (KEFRI), Kenya, (4)UK Centre for Ecology & Hydrology, United Kingdom, (5)World Agroforestry (ICRAF), Kenya, (6)University of Pretoria, South Africa, (7)BGI-Research, Shenzhen, China, (8)State Key Laboratory of Agricultural Genomics, China, (9)BGI-Shenzhen, China, (10)University of California, Davis, Davis, CA, (11)BGI-Shenzhen, Shenzhen, Guangdong, China, (12)Beijing Genomics Institute-Shenzhen, Shenzhen, China, (13)University of California, Davis, (14)National Autonomous University of Mexico, Mexico, (15)Ghent University, Ghent, Belgium, (16)University of California, Davis, CA, (17)World Agroforestry, Nairobi, Kenya

*Moringa (Moringa oleifera)* is a perennial shrubby tree originating from the foothills of the Himalayas in North-Western India and now widely distributed and cultivated across tropical and sub-tropical areas in Asia, Africa and Latin America. *Moringa* is an important multi-purpose nutrient-dense tree used as food, fodder, and medicine as a part of tropical and semi-tropical agroforestry landscapes. It is a fast growing, easily manageable tree with abundant production of edible leaves, flowers and fruits. The leaves are 27% protein by dry weight with essential amino acids and are rich in vitamins C, provitamin A, K, beta-carotene and minerals- calcium, iron, magnesium, manganese, and phosphorus with high dietary fiber. The leaves also contain a very high amount of antioxidants (polyphenol) and anti-inflammatory agents (isothiocyanate) with potential health benefits. Due to its agro-economic potential and high

nutrient contents like iron, it is considered as an important potential source to combat malnutrition, especially for women and children in developing countries. Thus, Moringa was prioritized by the African Orphan Crops Consortium to generate genomic resources and to develop a breeding program based on this information.

As a first step, a good quality genome sequence was generated using Illumina's short read sequencing technology with a total length of 216.8 Mb covering 79% of the estimated genome size. An evidence-guided electronic genome annotation indicated presence of 18,451 protein-coding genes with a high genome completeness as indicated by BUSCO evaluation, where ~89% of the core embryophyta genes were detected. This genome sequence is being improved and strengthened by long read sequencing technologies.

An analysis of genetic diversity in the ICRAF's Moringa gene bank collection (~ 400 accessions) using DArT-SNP markers indicated that African material comprised a narrow genepool. From the total analyzed Moringa genepool the material from Philippines was the most diverse (if scaled at 100%) followed by Ghanaian collection (83%), Malawian (77%), Haitian combined with Jamaican (62%), and East African (57%). ICRAF (World Agroforestry Centre, Nairobi, Kenya) along with the Kenya Forestry Research Institute (KEFRI, Nairobi, Kenya) and the UK Centre for Ecology & Hydrology (CEH, Edinburgh, UK) have jointly developed a programme to link the newly developed genomic resources with evaluation of quantitative genetic variation in key traits. Multisite and multi-environment field trials comprising 3,000 trees from 70 half-sib families sampled across Kenya and the global collections has been laid out. Phenotypic evaluation and genotyping will commence in 2020 and genome-wide association approaches will be used to assess trait-linked markers and to evaluate the potential for marker-assisted or genomic selection to improve the species for beneficial traits such as leaf and seed production, seed oil yield, foliar nutritional content.

### **W013: African Orphan Crops**

#### **Draft Genome Sequence of the African Eggplant**

**Damaris Achieng Odeny**, The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Nairobi, Kenya

The African eggplant (*Solanum aethiopicum*) is a nutritious traditional vegetable used in many African countries, including Uganda and Nigeria. Believed to have been domesticated in Africa from its wild relative, *S. anguivi*. *S. aethiopicum* has been routinely used as a source of disease resistance genes for several Solanaceae crops, including *S. melongena*. Lack of genomic resources has meant that breeding of *S. aethiopicum* has lagged behind other vegetable crops. We assembled a 1.02 Gb draft genome of *S. aethiopicum*, which contained predominantly repetitive sequences (76.2%). We annotated 37,681 gene models, including 34,906 protein-coding genes. Expansion of disease resistance genes was observed via two rounds of amplification of long terminal repeat retrotransposons, which may have occurred around 1.25 and 3.5 million years ago, respectively. By re-sequencing 65 *S. aethiopicum* and *S. anguivi* genotypes, 14,995,740 single nucleotide polymorphisms (SNPs) were identified, of which 41,046 were closely linked to disease resistance genes. Analysis of domestication and demographic history revealed active selection for genes involved in drought tolerance in both 'Gilo' and 'Shum' groups. A pan-genome of *S. aethiopicum* was assembled, containing 51,351 protein-coding genes; 7,069 of these genes were missing from the reference genome. The genome sequence of *S. aethiopicum* enhances our understanding of its biotic and abiotic resistance. The single nucleotide polymorphisms identified are immediately available for use by breeders. The information provided here will accelerate selection and breeding of the African eggplant, as well as other crops within the Solanaceae family.

### **W014: African Orphan Crops**

#### **Draft Genomes of Two Artocarpus Plants, Jackfruit (*A. heterophyllus*) and Breadfruit (*A. altilis*) Provide Insight into Starch Metabolism**

**Xin Liu**, Beijing Genomics Institute-Shenzhen, Shenzhen, China

### **W015: African Orphan Crops**

## **African-Led Genome Sequencing of Lablab and African Yam Bean Orphan Crops Genomes to Unveil Pathways Underlying Key Traits**

**Bernice Waweru**<sup>1</sup>, Jean-Baka K. Domelevo Entfellner<sup>2</sup>, Oluwasheyi Shorinola<sup>3</sup>, Peter Emmrich<sup>4</sup>, Chris Jones<sup>3</sup>, Isaac Njaci<sup>4</sup>, Beatus Lymo<sup>5</sup>, Davies Kaimenyi<sup>6</sup>, Mukani Moyo-Okoba<sup>7</sup>, Edwin Murungi<sup>8</sup>, Ermias Assefa<sup>9</sup>, Ahadi Birindwa<sup>10</sup>, Mary Gathoni Maranga<sup>11</sup>, Abubaker Muwonge<sup>12</sup>, Piuz Mugaga<sup>13</sup>, Hellen Nigussie<sup>14</sup>, Hassan Ali<sup>15</sup> and Henry Osamede<sup>16</sup>, (1)KALRO, Njoro, Kenya, (2)Biosciences eastern and central Africa, International Livestock Research Institute, Nairobi, Kenya, (3)ILRI, Nairobi, Kenya, (4)BecA-ILRI Hub, Kenya, (5)Nelson Mandela Institute of Science and Technology, Tanzania, United Republic of, (6)The Sainsbury Laboratory, United Kingdom, (7)International Potato Center, Kenya, (8)Egerton University, Kenya, (9)Ethiopian Institute for Agricultural Research, Ethiopia, (10)Evangelical University in Africa, Congo (The Democratic Republic of the), (11)Jagellonian University, Poland, (12)NaCCRI, Uganda, (13)Makerere University, Uganda, (14)Addis Ababa University, Ethiopia, (15)University of Nyala, Sudan, (16)University of Ibadan, Nigeria

Legumes form a crucial part of diets in many Sub-Saharan Africa countries due to their high seed protein content and their low cost when compared to animal-derived protein sources. Despite offering excellent opportunities for sustainable intensification of agriculture, many legumes, especially orphan legume crops have received little research and breeding attention, causing large yield gaps and wasted potential for addressing the challenges of food security and sustainability. Two of these orphan legumes are lablab (*Lablab purpureus*) and African yam bean (AYB; *Sphenostylis stenocarpa*). Rich in protein, macro and micro-nutrients, both legumes are important for their edible leaves and seed grains. The African Yam bean also uniquely produces edible tubers while the haulms of both are utilized as animal feed. Both legumes exhibit nitrogen-fixing ability and thrive in marginal soils under low-input farming systems. However, both legumes are still largely underutilized, mainly due to their hard seed coats and the presence of anti-nutritional factors reducing digestibility. Little genomic information is available to assist in efforts to understand the genetic architecture underlying important traits and unlock the full potential of these crops. This study reports the first draft genomes of the AYB and lablab based on third generation long reads using Oxford Nanopore sequencing. Several approaches were used to generate long read *de novo* assemblies from 5.1 million AYB reads and 31 million lablab reads. Assembly using Redbean yielded an assembly length of 841 Mb with N50 of 48,083 bp for AYB and 354 Mb with N50 of 562,331 bp for lablab. Flye gave an assembly of 653 Mb for AYB with an N50 of 409,006 bp and 369 Mb with N50 of 1,442,401 bp for lablab. We also performed hybrid assembly using Illumina short reads to improve the accuracy of the assemblies and this yielded an assembly length of 486 Mb with an N50 of 7,124 bp for AYB. Efforts are ongoing to improve the contiguity of these assemblies to achieve chromosome-scale assemblies using Hi-C mapping. We are also generating RNA-Seq data for functional annotation of the AYB genome to identify key pathways underlying important nutritional, adaptation and resource partitioning traits.

This is the first report of the AYB genome, and long read lablab genome sequenced, assembled, and analyzed entirely in Africa by a group of young bioinformaticians following an ambitious capacity building effort at the Biosciences East and Central Africa ILRI Hub in Kenya. Our work therefore highlights the opportunities presented by orphan crop genome sequencing efforts for capacity building in agricultural genomics and bioinformatics in Africa.

### **W016: A Global Vision for Crop Improvement and Food Security: Connecting the Dots**

#### **A Strategic Approach to Modernization of the Breeding Pipeline and Delivery of Genetic Gains**

**Nora Lapitan**, USAID Bureau for Food Security, Washington, DC

Feeding a rapidly growing population will require modernization of the global breeding pipeline to deliver a steady stream of improved crop varieties designed for a range of environmental and market conditions. A recognition of this need, especially in the context of the unique challenges confronted by institutions, breeders, producers, and consumers in under-resourced parts of the world, led to the development of a strategic approach to advancing CGIAR and developing-country national agricultural research organization breeding systems. Two new multi-donor funded programs were recently established to address these challenges. The Excellence in Breeding platform provides tools, services, and technical guidance to developing-country breeding systems, as well as technical oversight of Crops to End Hunger, an initiative aimed at accelerating and modernizing the development, delivery,

and widescale use of new crop varieties. Earlier this year, USAID funded a Crop Improvement Innovation Lab that will support and promote these efforts. Led by Cornell University, the Crop Improvement Innovation Lab will harness the expertise of scientists in the US and partner countries to develop new breeding tools, technologies, and methods, especially for improvement of sorghum, millet, legumes, and roots, tubers, and bananas. Together, these programs will work toward a unified, modern pipeline that will promote the use of new technologies, centralized breeding approaches, appropriate designs for product profiles, integrated biofortification, access to mechanization, better parental selection, germplasm testing, and improved data handling and management. The ultimate goal is to arrive at an aligned global effort to achieve genetic gains and improved variety turnover.

## **W017: A Global Vision for Crop Improvement and Food Security: Connecting the Dots**

### **Vision for the Feed the Future Innovation Lab for Crop Improvement**

**Stephen Kresovich**, Clemson and Cornell Universities, Clemson, SC

The vision for the Cornell University Management Entity (ME) for the Feed The Future (FtF) Innovation Lab for Crop Improvement (ILCI) is to serve as a support system for national agricultural research institutions (NARIs) in target regions to identify, develop, pilot, and transfer appropriate tools, technologies, and methods (TTMs), equipping them to deliver increased genetic gain and new varieties for key product profiles that advance economic growth, resilience, and nutritional development goals of the Global Food Security Strategy (GFSS). The ILCI will be dynamic, holistic and responsive, with high potential for immediate impact through a visionary and opportunistic ME coupled with multidisciplinary areas of inquiry — designated the Crop Improvement Toolbox. Leveraging the ongoing activities this consortium encompasses, the Toolbox will be ready for rapid deployment of TTMs and expertise to NARIs in targeted crops and geographies. We plan five areas of inquiry: 1) Priority Setting; 2) Trait Discovery; 3) Genomics; 4) Phenomics; and 5) Breeding Informatics. These areas of inquiry, together with NARI capacity development activities, will contribute to GFSS goals and fulfill the four objectives of the ILCI including: 1) TTMs that improve trait discovery and breeding efforts identified, developed, piloted, and transferred to NARIs; 2) strengthened capacity of NARIs to utilize TTMs for accelerated and improved breeding of locally-adapted crop varieties targeted to smallholders; 3) ILCI activities supported by leveraged resources and aligned efforts; and 4) ILCI activities and outputs coordinated across FtF crop improvement portfolio efforts. In complement, a competitive research portfolio will focus on dual goals of establishing centers of innovation in target regions for development and deployment of the Toolbox, as well as area of inquiry-specific investments to transfer cutting-edge TTMs to a wider range of NARIs and collaborating institutions. The ILCI will support inclusive agricultural growth that benefits of investment and productivity gains in ways that target low-income people in particular, thus leading to gains in terms of reductions in poverty and undernutrition and gains in resilience and will address key crosscutting priorities including gender equality; youth inclusion; nutrition and food safety; and resilience and risk management. A rigorous monitoring, evaluation, and learning plan has been designed to quantify the progress and impact of the proposed ILCI priorities and activities and measure contributions to the goals of individual programs and institutions in the network as well as the overall goal of the GFSS.

## **W018: A Global Vision for Crop Improvement and Food Security: Connecting the Dots**

### **The CGIAR Excellence in Breeding Platform and Crops to End Hunger Initiative**

**Biswanath Das**<sup>1</sup>, Michael Olsen<sup>1</sup>, Kelly Robbins<sup>2</sup> and Michael Quinn<sup>3</sup>, (1)CIMMYT, Nairobi, Kenya, (2)Cornell University, Ithaca, NY, (3)Excellence in Breeding, CIMMYT, Texcoco, Mexico

The CGIAR Excellence in Breeding platform was established in 2017 with the objective of modernizing CGIAR and NARs breeding programs to order to increase their rates of genetic gain and variety turnover on farmer's fields. The platform consists of 5 modules supporting end to end modernization of breeding pipelines through provision of tools, services and technical consultancy. Over the past 24 months most CG breeding programs and select NARs breeding programs in Kenya, Uganda and Ghana have been assessed and priorities identified for modernization. In 2019, the Crops to End Hunger (CtEH) initiative, was set up by major donors to fund modernization priorities that will have a transformational impact on priority CGIAR and NARs breeding programs with technical oversight provided by EiB.

## **W019: A Global Vision for Crop Improvement and Food Security: Connecting the Dots**

## **Cowpea Research in Burkina Faso: Progress, Challenges to Crop Improvement, and the Place of the Crop Improvement Innovation Lab Initiative**

**Teyoure Benoit Joseph Batieno**, Institut de l'Environnement et de Recherches Agricoles, Kamboinse, Burkina Faso; INERA, Ougadougou, Burkina Faso and Tignegre Jean-Baptiste, Sidibe Hamadou, Zongo Hamadou, Zida Serge, Poda Leandre, Coulibaly Soumabere, Bao-Lam Hyunh, Timothy Close, Philip Roberts, and Ouedraogo Jeremy

Cowpea (*Vigna unguiculata* [L.] Walp.) is one of the cheapest sources of protein for rural people in Burkina Faso. Harvested before the cereal crops, cowpea is considered as a “hungry-season” crop or poor’s meat. Its potential to address food and nutritional security in Burkina Faso and beyond is well established. However, the crop yield remains low. Efforts have been made to improve several traits including drought tolerance, and resistance to Striga and diseases. Progress in recent years has benefitted from projects funded by the US Agency for International Development (USAID; Legume Innovation Lab, Innovation Lab for Climate Resilient Cowpea), the Kirkhouse Trust SCIO, the Alliance for a Green Revolution in Africa (AGRA), and the CGIAR Generation Challenge Program (Tropical Legumes). A long-standing relationship between INERA and the University of California Riverside has been important in for improvement of the breeding methods in Burkina Faso, which combine farmers’ participatory variety selection and modern breeding tools. This has led to the development of important breeding lines currently under testing for release, and for prior release of the most popular varieties such as Tiligré and Komcallé. Through adoption of new varieties, together with improvements in farmer practices, yield has increased from 200kg/ha to 800 kg/ha and the mean production from 100,000 tonnes in the early 1980’s to 700,000 tonnes in recent years. However, efforts are still needed improvements in physical resources and a better workflow to improve genetic gain. This presentation will focus on achievements of the cowpea breeding program in Burkina Faso, the challenges, and some comments on how the new Crop Improvement Innovation Lab might be of assistance.

## **W020: A Global Vision for Crop Improvement and Food Security: Connecting the Dots**

### **The Evolution of a Revolution: Re-Designing Green Revolution Breeding Programs in Asia and Africa to Increase Rates of Genetic Gain**

**Joshua N. Cobb**, International Rice Research Institute, Los Baños, Laguna, Philippines, Mohammad Rafiqul, International Rice Research Institute, Bangladesh, Sanjay Kumar Katiyar, International Rice Research Institute (IRRI), India, Alexis Ndayiragije, International Rice Research Institute, Mozambique, Rosemary Murori, International Rice Research Institute, Tanzania, United Republic of, John Damien Platten, International Rice Research Institute, Los Baños, Philippines, Jerome Bartholome, International Rice Research Institute | CIRAD, Philippines and Parthiban Prakash, International Rice Research Institute, Los Banos, Philippines

As rice feeds nearly half of the human population, rice breeding is a critical focal point for achieving the UN Sustainable Development Goal of eliminating hunger and poverty by 2030 and to providing a sufficient quantity of safe and nutritious food to vulnerable populations in the developing world. However, despite dramatic improvements in understanding the genetic basis of complex traits in rice over the last 20 years, annual rates of genetic gain for yield and other important traits in most public rice breeding programs in Asia and Africa are extremely low. Understanding and manipulating the key drivers of genetic gain will be necessary for rice breeding programs to fully meet the expectations of the 21st century. Funded by the Bill and Melinda Gates foundation and in coordination with the CGIAR Excellence in Breeding Platform, the International Rice Research Institute (IRRI) aims to transform rice breeding by aligning IRRI’s international breeding efforts together with national public breeding programs (NARs programs) into collaborative regional breeding networks. These CGIAR-NARs breeding networks serve as a platform to deploy an integrated breeding model that combines modern genomic technologies with regional knowledge and testing capabilities to ensure that smallholder rice farmers have access to a steady stream of consistently improved, high yielding, locally adapted, and market-ready rice varieties.

## **W021: A Global Vision for Crop Improvement and Food Security: Connecting the Dots**

### **Delivering Genetic Gains to Smallholder Farmers in the Face of Climate Change**

**Gary Atlin**, Bill & Melinda Gates Foundation, Seattle, WA

Cropping systems in the developing world are changing rapidly due to population growth, system intensification due to market linkage, and climate change. In many areas, farmers are trying to increase yields and fit additional crops into rotations in the face of climate instability. A steady stream of new varieties, developed for current conditions and markets, is needed to support both agricultural transformation and adaptation to a changing climate, but most farmers in the developing world use varieties that are over 20 years old. This is largely due to a “stalled Green Revolution” resulting from national and international public sector breeding systems that rely heavily on visual selection and deliver very low rates of genetic gain.

CGIAR breeding programs and their national partners are embracing three major breeding system changes that will improve both the rate of genetic gain and the probability that new varieties will be adopted by farmers:

1. *The conceptualization of breeding as product development.* This has placed the intentional design of new varieties to meet the needs of farmers, processors, and consumers at the heart of the breeding process. Product profiles that guide cultivar development are assembled on the basis of market intelligence gleaned from men and women farmers and consumers as well as other value chain actors. Traits must be weighted in these profiles both on the basis of their contribution to value and on their variances and covariances.
2. *The quantitative optimization of breeding pipelines,* using quantitative genetic principles to maximize the rate of genetic gain delivered per year and per dollar. This requires separating the improvement of source populations from the extraction of commercial products from those populations, with greatly increased focus on optimization of the population improvement component. Pipelines are optimized by accelerating the breeding cycle, and increasing the accuracy of selection for breeding value using information from relatives gleaned from the application of G-BLUP at early stages of testing. Accelerated population improvement that uses some elements of “speed breeding” technology to advance 3 to 4 generations per year can complete breeding cycles in 2 or 3 years; in such programs high rates of gain can be achieved even in modestly sized populations that can be easily handled by small public breeding institutes and private companies.
3. *The implementation of systems for rapidly deploying new high-value haplotypes,* quickly bringing them from very low to high frequency in elite breeding populations, and thus allowing them to be used in forward breeding with minimal linkage drag.

Application of these three major advances in public sector breeding requires high levels of technical support, as well as access to low-cost and reliable outsourced genotyping and bioinformatics services. This support is being provided to CGIAR breeding programs and their national partners through the Excellence in Breeding Platform. Together, these changes have the potential to increase the rate of genetic gain delivered to farmers in the developing world 3- or 4-fold.

## **W022: Allele Mining**

### **SNP Markers for Panicle Architecture and Grain Traits Developed from GWA-QTL and Available for *japonica* Rice Improvement**

**Georgia C Eizenga**, Aaron K. Jackson, Melissa H. Jia, Jeremy D. Edwards and Anna M. McClung,  
USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR

Panicle architecture, grain size and grain weight are important yield component traits to consider when breeding rice (*Oryza sativa* L.). Having SNP markers for these traits would expedite breeding efforts through marker assisted selection. Previously, the Rice Diversity Panel 1 (RDP1), representing the five major rice subpopulations, was phenotyped for these yield related traits and genome-wide association mapping (GWA)-QTL were identified. Most southern U.S. cultivars are classified as *tropical japonica* and California cultivars as *temperate japonica*; thus, these subpopulations are of particular interest to U.S. breeders. To develop markers and dissect the variation underlying these GWA-QTLs for rice improvement, diverse *japonica* RDP1 accessions were selected as parents to develop biparental recombinant inbred line (RIL) mapping populations. All 276 progeny from the Estrela (admixture of *japonica*) × NSFTV199 (*tropical japonica*) cross have been evaluated for panicle architecture and grain traits. The QTL analyses with 256 RILs revealed 38 RIL-QTL which overlapped with the previously identified GWA-QTL and regions on chromosomes 3, 4, 5, 6, 7, 8 and 9 were selected for marker development. To develop markers, the sequence variation in regions surrounding the significant SNPs identified in GWA studies was assessed in rice

genomic databases to find optimum sites to target for marker development. Once developed, the markers were tested in the Estrela × NSFTV199 population for amplification and polymorphism, and in two additional biparental *japonica* populations developed from RDP1 accessions. Polymorphic marker data were analyzed in these populations to identify marker-trait associations. Six Estrela × NSFTV199 RILs with desirable panicle architecture traits along with acceptable grain size, maturity, plant height and grain yield were selected for evaluation in a replicated field trial conducted over two years. The targeted SNP markers corresponded well with the panicle architecture and grain size data from the field studies. These SNP genotypes will be used to select RILs with the best agronomic and grain quality traits that are suited for the U.S. market along with QTL for desirable panicle architecture traits. The selected RILs will be released as improved germplasm for U.S. breeders.

### **W023: Allele Mining**

#### **TE-Marker, a Transposon-Based Marker System for Identifying Alleles Associated to Agronomic Traits**

**Haidong Yan**, School of Plant and Environmental Sciences, Virginia Tech, USA, Blacksburg, VA, Aureliano Bombarely, Virginia Polytechnic Institute and State University, Department of Horticulture, Blacksburg, VA; Department of Bioscience University of Milan, Milan, VA, Italy, David Haak, Virginia Tech and Bingyu Zhao, Virginia Tech, Blacksburg, VA

Identification of genetic markers associated with agronomic traits is an essential strategy to improve crop performance. Single Nucleotide Polymorphism (SNP) marker system was widely utilized in Genome-Wide Association Study (GWAS). However, this system may generate spurious associations due to high Linkage Disequilibrium (LD) of the SNP markers. In this study, we proposed a genetic marker system called TE-marker using Transposable Elements (TEs) based on whole-genome shotgun sequencing data. We have used two different datasets to test our approach: 1- Reanalysis of the resequencing of 50 different rice accessions. This approach is able to cluster four populations (*Oryza sativa japonica*, *O. sativa indica*, *O. rufipogon*, *O. nivara*) with the admix accessions between them as the classical SNP marker approach. 2- Reanalysis of 176 *O. sativa japonica* accessions and their association to specific traits through GWAS. Our results showed that LD of TE markers decayed faster than of SNP markers. For the GWAS, TE markers could detect association peaks that were equivalent to the ones produced by SNP markers for some traits such as leaf blade width. Our approach pointed seven alleles nearby the peak regions with annotations related with photosynthesis, and auxin and gibberellin responsiveness that it could be associated with leaf development in the rice. TE-marker is a novel complementary approach to the classical SNP markers that it assists to reveal population structures and identify alleles associated with agronomic traits.

### **W024: Allele Mining**

#### **Dynamic Effects of Interacting Genes underlying Rice Flowering-Time Phenotypic Plasticity and Global Adaptation**

**Tingting Guo**<sup>1</sup>, Qi Mu<sup>1</sup>, Jinyu Wang<sup>1</sup>, Adam E. Vanous<sup>1</sup>, Akio Onogi<sup>2</sup>, Hiroyoshi Iwata<sup>3</sup>, Xianran Li<sup>1</sup> and Jianming Yu<sup>1</sup>, (1)Iowa State University, Ames, IA, (2)National Agriculture and Food Research Organization, Japan, (3)The University of Tokyo, Tokyo, Japan

Observed phenotypic variation in living organisms is shaped by genomes, environment, and their interactions. Rapidly accumulating genome sequence data have facilitated the identification of novel alleles of important genes underlying phenotypic variation. However, how alleles and combinations of alleles change effects across environments, resulting in phenotypic plasticity, is often unknown. Here, we demonstrate an analytical framework to dissect phenotypic plasticity and to associate haplotype plasticity with geographic distribution. First, we conducted phenotypic plasticity dissection in a rice genetic population. The observed flowering-time plasticity was systemically mapped to four known genes (*Hd1*, *Hd2*, *Hd5*, and *Hd6*). These genes, discovered for their photoperiodic response, differentially responded to temperature at the early growth stage to jointly determine flowering time. The effects of these plasticity genes were revealed with multiple reaction norms along the temperature gradient. With the integration of genomics and the temperature environmental index, accurate performance predictions were obtained. Second, we examined the accessions from the 3,000 Rice Genomes Project for allelic variation at the four flowering-time genes and constructed haplotypes at both individual-gene and multi-gene levels. The geographic distribution of haplotypes showed that haplotype combinations were preferentially

adapted to different temperature zones. The temperate zone was dominated by haplotypes more sensitive to temperature changes, while the tropical zone had a majority of less sensitive haplotypes. Our findings bridged the gap between phenotypic plasticity dissection and allele mining by integrating knowledge from genomics, cloning and function characterization, environment quantification, and predictive modeling.

## **W025: Allele Mining**

### **Allele Discovery in a Six-Row Barley Multiparent Population**

**Gary J. Muehlbauer**, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Characterizing and exploiting diverse germplasm collections is essential for continued crop improvement. Increasing the genetic diversity in the US Upper Midwestern spring six-row malting barley cultivars is especially important as they exhibit limited genetic diversity driven by the strict malting and brewing company requirements. One source of genetic diversity is the USDA National Small Grain Core Collection (NSGCC) which contains 2,417 barley accessions, composed primarily of landraces and cultivars/varieties from over 100 countries. To exploit the diversity in the NSGCC for barley improvement and allele mining, we developed a multiparent population by crossing 88 diverse six-row parents randomly selected from the barley NSGCC with the spring six row barley malting variety Rasmusson. This population, referred to as the Barley Recombinant Inbred Diverse Germplasm Population (BRIDG6), consists of 6,160 F<sub>5</sub>-derived recombinant inbred lines. Quantitative trait locus (QTL) mapping identified 23 flowering time QTL, among which seven were associated with previously identified flowering time genes including *HvPpdH1*. Exome capture sequencing of a subset of 78 of the BRIDG6 parents revealed 11 haplotypes at *HvPpdH1*, and accessions from Asia exhibiting previously undiscovered private alleles with both positive and negative effects. The large size of BRIDG6 often precludes screening the entire population for traits that are time consuming or expensive to score. Thus, to enable more efficient use of BRIDG6 for future QTL mapping studies several subsampling strategies were examined, resulting in an optimal sample size for the most robust QTL detection of least 50 families and 3,000 RILs. Taken together, this study demonstrates the utility of BRIDG6 for allele mining and barley improvement.

## **W026: Analysis of Complex Genomes**

### **Insights into Rye Biology and Triticeae Relationships based on a Chromosome-Scale Genome Assembly**

**Nils Stein**, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany and International Rye Genome Sequencing Consortium

Rye, *Secale cereale*, belongs to the Triticeae tribe in the *Poaceae*. Opposite to the global importance of its close relatives barley and wheat, rye is economically relevant mainly to the Northern European countries. It is produced as animal feed and biofuel crop but also for human food. Rye biology and life history is in contrast to wheat and barley for many aspects. Rye is a secondary domesticate which traveled to Europe as a weed in barley and wheat fields. It is a self-incompatible and out-crossing species with very good winter hardiness. Heterosis in rye is high and is exploited efficiently in CMS-facilitated hybrid breeding. Its diploid genome is estimated to comprise 7-8 Gigabases which is about 50% larger than the closely related diploid Triticeae genomes e.g. of barley (5 Gbp). An international consortium used now de novo short read sequencing-by-synthesis and assembly combined with high density genetic mapping, Hi-C analysis and optical mapping to produce a high quality chromosome-scale sequence assembly, which greatly facilitates the application of molecular genetic tools and strategies in research and breeding for crop improvement and for reaching a better understanding of the mechanisms underlying rye's distinct biological and genetic features.

## **W027: Analysis of Complex Genomes**

### **Improving Methods of Automatic Annotation of Plant Genomes**

Tomas Bruna, Alexandre Lomsadze and **Mark Borodovsky**, Georgia Institute of Technology, Atlanta, GA

Integration of different pieces of evidence, i.e. RNA-Seq reads and homologous protein mapping and *ab initio* derived sequence patterns is critical for accurate annotation of plant genomes. Complex plant genomes have large size, large number of repeats (transposable elements) as well as heterogeneous nucleotide composition.

It was shown that when general gene prediction methods were applied to plant genomes significant manual work is still needed to reach genome annotation with satisfactory accuracy. Construction of a fully automated method of annotation of novel complex genomes is still an open problem.

Earlier we have developed automated gene finding method with unsupervised training of statistical models employed in the algorithm. This type of approach generates many false positives in complex plant genomes with large volume of non-coding regions. We present a novel method of model training in which sets of coding and non-coding regions is selected based on mapping of transcriptome and protein data.

New algorithm, GeneMark-ETP+, integrates RNA-Seq short read alignments produced by VARUS as well as hints generated by protein mapping delivered by ProHint. This semi-supervised training approach was shown to generate more accurate annotation of principle isoforms of protein coding genes. A focus of the new method is selection of a highly reliable set of introns derived from RNA-Seq reads, proteins, and *ab initio* predictions to guide *ab initio* gene finder training and predictions. We demonstrated that this approach resulted in increase in accuracy of gene annotation in complex plant genomes.

## **W028: Analysis of Complex Genomes**

### **Long Read Sequence Technology Resolves Hidden Features among Brassica Genomes**

**Andrew G. Sharpe**, Global Institute for Food Security (U of S), Saskatoon, SK, Canada

Plant genome assembly has been developing rapidly with costs declining and scaffold size and genome coverage improving; however, with short read technologies, underlying contig size remains limited and it is inevitable that some genomic regions will not be captured and duplicated or repetitive regions are often collapsed. Concomitant with these improvements there is a growing appreciation that copy number variants, presence/absence variants and structural rearrangements have played an important role in the adaptation of phenotype. Long read sequencing technologies offer a unique opportunity to capture often elusive structural variation in genomes. To test applicability to polyploid species, a *de novo* genome assembly was generated for *Brassica nigra*, a paleohexaploid, using Oxford Nanopore Technologies (ONT) sequence reads. The resultant assembly was error corrected using Illumina short reads, and HiC and genotype data was added to generate pseudomolecules. The ONT assembly extended the original reference assembly by 59 Mb, covering ~89% of the expected genome size. The majority (85%) of the additional assembled sequence represented repetitive DNA, yet ~3,500 additional genes were added to the new assembly. The long-read assembly provided a novel insight into the repetitive genome structure, access to previously hidden genes, and could span non-recombinant regions.

## **W029: Analysis of Complex Genomes**

### **TILLING by Target Capture Sequencing (TbyTCS) to Improve Soybean Seed Composition Traits**

**Zhou Zhou**, Naoufal Lakhssassi and Khalid Meksem, Southern Illinois University at Carbondale, Carbondale, IL

Chemical mutagenesis emerges as a Genetically Modified (GM)-free strategy to produce large-scale allelic series in soybean for economically important trait improvement. Here we develop a high-throughput TILLING by Target Capture Sequencing (TbyTCS) technology coupled with single nucleotide polymorphisms (SNPs) identification bioinformatic tools to identify population-wide mutations in soybeans. Because of the robustness of SNP calling, this novel technology ensures high-quality yield of true mutations while removing the majority of false positives. Four Ethyl methanesulfonate (EMS) mutagenized soybean populations (over 4000 mutant families) have been screened for the presence of induced mutations in targeted genes. The mutation types and effects have been characterized for a total of 138 soybean genes in soybean seed composition, disease resistance, and other traits. By using TbyTCS, we discovered novel sources of soybean oil traits as well as protein and carbohydrate traits. EMS-induced mutations were identified and characterized for 19 genes within the fatty acid, protein, and carbohydrate

biosynthetic pathway, including the *GmKASIIA/B*, *GmSACPD-C/D*, *GmFAD2-1A/1B*, *GmFAD3A/B/C*, *GmSus*, *GmGy*, and *GmCG*. The TbyTCS technology provides an unprecedented platform for highly effective screening of polyploidy mutant populations and gene functional analysis. The obtained soybean mutants in this study are used in subsequent soybean breeding for improved seed composition traits.

### **W030: Analysis of Complex Genomes**

#### **Using Contiguous Long-Read Assemblies to Investigate Interspecific Macrosatellite Evolution and Speciation**

**Kevin R Bredemeyer**, Texas A&M University, COLLEGE STATION, TX and William J. Murphy, Texas A&M University, College Station, TX

Since their discovery as valuable genetic markers, repetitive sequences that differ in copy number between individuals and species, known as variable number tandem repeats (VNTRs), have received ever-increasing attention from geneticists. Macrosatellites are a class of VNTRs defined by particularly large repeat units, often several kilobases in length, high GC content and highly polymorphic copy number. Macrosatellites have been implicated in key cellular processes through transcriptionally or spatially controlled chromatin remodeling and are often regulated through direct DNA methylation that can be disrupted by significant deviations in copy number. We previously mapped a major effect X-linked locus associated with male hybrid sterility in an interspecific hybrid cat breed, and identified the macrosatellite *DXZ4* as a candidate hybrid sterility locus. In addition to copy number variability, *DXZ4* is an interesting candidate because of its implication in the maintenance of the structural conformation unique to the inactive X of female somatic cells. Together, these two features suggest a potential link between interspecies divergence and structural misregulation of the X chromosome during a male specific instance of X chromosome silencing known as meiotic sex chromosome inactivation, which leads to hybrid male sterility. However, due to the highly repetitive and polymorphic nature of macrosatellites like *DXZ4*, most of these regions are missing or misassembled in all but the highest quality finished genome assemblies. Our recent *de novo* assembly of the genomes of several cat species using long read sequencing, F1 Trio-binning, and HiC scaffolding has allowed completion of, and interspecific comparisons between, many of these complex genomic regions, providing crucial insight connecting rapid evolution of macrosatellites and their association with speciation.

### **W031: Analysis of Complex Genomes**

#### **Comparative Transcriptomics and Metabolite Profiling to Unravel Defense and Defensive Secondary Metabolite Formation in Tea Trichomes**

**Jian Zhao**, Anhui Agricultural University, Hefei, China and Penghui Li

State Key Laboratory of Tea Plant Biology and Utilization, College of Tea and Food Science and Technology, Anhui Agricultural University, Hefei, 230036,

#### **Abstract**

Tea plant (*Camellia sinensis*) leaf produces a wide array of secondary metabolites, including catechins, caffeine, and theanine as characteristic bioactive ingredients of teas, which make major contributions to tea flavors and health benefits. Trichomes on tea plant leaves produced many special flavors-contributing products and are regarded as an important quality trait, however, the underlying molecular bases for how tea unicellular trichomes are generated on plant leaf to synthesize specialized metabolites, and what's their physiological function in tea plant are not fully understood. Here, we integrated metabolome and transcriptome analyses on tea trichomes and trichome-removed leaves (TR-leaves) to gain insight into the formation and functions of trichomes as well as their synthesis of specialized metabolites. Metabolite profiling and RNA-Seq data strongly supported the defense functions of trichomes by highly expressing many defensive proteins/enzymes, or synthesizing and accumulating a complex defensive secondary metabolites, some of them like mono- and sesquiterpenes, contributing to the essential role of trichomes in determining tea flavors. Besides producing these characteristic tea secondary metabolites, tea plant trichomes also highly and specifically accumulated more defense metabolites and expressed more defensive genes, such as UV-protective flavonols and peptides, insect-toxic caffeine and proteinase inhibitors, and other volatiles, disease-resistant metabolites and proteins. Tea trichomes also specifically express defense-related germin-like proteins, chitinase, laccase, GST, high-light protective proteins, and LRR protein kinase family genes, which are

usually associated with plant defense against herbivore insects and pathogens and abiotic stresses. We functionally characterized CsMYB184, CsGL3, and CsWD40-repeat genes are highly associated with trichome formation and development with molecular and genetic tools. Antisense knockdown of CsMYB184 also reduced expression levels of gene involved in biosynthesis of volatiles, catechin, and caffeine, followed by reduced production of these metabolites in buds and young leaves. Nucleus-localized CsMYB184 activated the promoters of ANR and ANS and a tea plant GL2 homolog CsGL2 genes in a reporter activation assay. Association studies on trichome phenotyping and gene expression in various tea plant germplasm demonstrated a close relation between CsMYB184 expression level and trichome density and lengths. Our findings suggest that tea plants have evolved to use trichomes for defense and adaptation to adversary environments, also by enhanced metabolic capabilities for efficient production of diverse defensive metabolites.

## **W032: Animal Epigenetics**

### **Profiling the Immune Epigenome across Global Cattle Breeds**

**Jessica R. Powell**, The Roslin Institute, Edinburgh, United Kingdom

Understanding the variation between well and poorly adapted cattle breeds to local environments is essential for breeding cattle with improved climate and disease resistant phenotypes. While studies are beginning to characterise the genetic basis of cattle breed diversity, alternative mechanisms underlying breed-specific traits are largely unexplored. Changes at the chromatin level are of particular interest because of their potential role in disease. However, the tools and reference resources to study these changes across cattle breeds are almost entirely lacking. In this study, we have characterised the chromatin accessibility and DNA methylation landscapes genome-wide of seven immune cell types across three diverse cattle breeds. Holstein Friesian, N'Dama and Nelore cattle were selected to represent the European taurine, African taurine and indicine cattle lineages respectively. Gene expression data for Holstein Friesian cattle have also been generated.

We find extensive epigenetic divergence between the taurine and indicine cattle lineages across cell types. The unique cell type profiles enable the accurate deconvolution of complex cellular mixtures using digital cytometry approaches. Finally, we show distinct sub-categories of CpG islands based on their chromatin and methylation profiles that discriminate between classes of distal and gene proximal islands linked to discrete transcriptional states.

These data provide a comprehensive resource to help exploit the diversity among cattle breeds and improve cattle productivity for farming communities in low and middle income countries.

## **W033: Animal Epigenetics**

### **Systematic Discovery and Characterization of Chromatin States and Butyrate-Induced Variations for Cattle Genome Functional Annotation**

**Cong-jun Li**, ARS, USDA, Beltsville, MD

The lack of functional annotation of genomes hinders the biological interpretation for complex trait variation, domestication, and adaptive evolution in livestock. Here we established the first global map of regulatory elements (15 chromatin states) and defined their coordinated activities in cattle, through genome-wide profiling for six histone modifications, RNA polymerase II, CTCF-binding sites, DNA accessibility, DNA methylation, and transcriptome in Rumen Epithelial Primary Cells (REPC). We demonstrated that each chromatin state exhibited specific enrichment for sequence ontology, gene expression and methylation across tissues, trait-associated variants, eQTLs, selection signatures, and evolutionarily conserved elements, implying distinct biological functions. After butyrate (a key regulator for rumen development) treatments, we observed that the weak enhancers and flanking regions of transcriptional start sites (TSS) were the most dynamic chromatin states, occurred concomitantly with significant alterations in gene expression and DNA methylation, which was significantly associated with heifer conception rate and stature. Our results demonstrate the crucial role of functional genome annotation for understanding genome regulation, complex trait variation, and adaptive evolution in livestock. Using butyrate to induce the dynamics of the epigenomic landscape, we were able to establish the correlation among nutritional elements, chromatin states, gene activities, and phenotypic outcomes.

### **W034: Animal Epigenetics**

#### **Developmental and Allele-Specific Methylation Patterns in Fetal Liver of Pigs Derived from White Composite x Meishan Reciprocal Crosses**

**Ryan J. Corbett**<sup>1</sup>, Haibo Liu<sup>2</sup>, Dan Nonneman<sup>3</sup>, Timothy P.L. Smith<sup>4</sup>, Nancy E. Raney<sup>5</sup>, Christopher K. Tuggle<sup>6</sup> and Catherine W. Ernst<sup>5</sup>, (1)Genetics & Genome Sciences Graduate Program, Michigan State University, East Lansing, MI, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)USDA, ARS, USMARC, Clay Center, NE, (4)U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, (5)Department of Animal Science, Michigan State University, East Lansing, MI, (6)Iowa State University, Ames, IA

The liver is a major metabolic organ that influences numerous economically important phenotypes in swine. However developmental and allele-specific gene regulation, including that governed by DNA methylation, are understudied in pig liver. We performed whole-genome bisulfite sequencing (WGBS) of pig fetal liver collected from White Composite and Meishan reciprocal crosses at 30 and 70 days of gestation (dg;  $N=8,2/\text{stage}/\text{cross}$ ) in order to assess stage- and allele-specific methylation (ASM). WGBS read alignment and extraction of CpG methylation rates were performed using Bismark, and differential methylation analyses were performed using *methylKit*. We also performed allele-specific mapping of reads and ASM analyses, using SNPs identified from whole-genome sequencing data. Global CpG methylation rates ranged from 59.5-63.6% and were significantly higher in 70dg samples ( $p=0.01$ ). We identified 24,601 differentially methylated regions (DMRs; difference > 10%, FDR <  $1e-5$ ) between stages, 91% of which were hypermethylated at 70dg. DMRs were enriched in gene promoters; 1956 promoter-hypermethylated genes (70dg vs. 30dg) were enriched for GO terms related to early development, while 676 promoter-hypomethylated genes were enriched for lipid and glucose metabolism terms, suggesting decreased and increased transcription of genes involved in these processes, respectively. 529 regions exhibited ASM between White and Meishan alleles, and these were enriched in genes associated with lipid metabolism. Lastly, 430 regions exhibited ASM between maternal and paternal alleles, including regions in the *IGF2* and *IGF2R* gene clusters that are known to exhibit genomic imprinting. This work provides novel insight into epigenetic regulation during pig liver development and has identified genomic regions subject to breed- and parent-specific regulation.

### **W035: Animal Epigenetics**

#### **Hitting the Mark: Characterizing Four Histone Modifications in Ovine Liver and Spleen**

**Kimberly M. Davenport**<sup>1</sup>, Alisha T. Massa<sup>2</sup>, Michelle R. Mousel<sup>3</sup>, Maria K. Herndon<sup>4</sup>, Stephen N. White<sup>3</sup>, Suraj Bhattarai<sup>5</sup>, Stephanie McKay<sup>5</sup>, Timothy P.L. Smith<sup>6</sup>, Jacky K. Carnahan<sup>7</sup>, Noelle Cockett<sup>8</sup>, Brenda M. Murdoch<sup>1</sup> and the Ovine FAANG Project, (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)Washington State University, Pullman, WA, (3)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (4)Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, (5)University of Vermont, Burlington, VT, (6)USDA, ARS, USMARC, Clay Center, NE, (7)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, (8)Utah State University, Logan, UT

Although the ENCODE project has defined regulatory elements in humans, very little is known in sheep. The functional annotation of the sheep genome, including the location of gene regulatory elements, is essential for understanding the potential mechanisms that may influence economically important traits. The objective of this study is to identify the locations of gene regulatory elements in sheep by characterizing histone modifications in two tissues, liver and spleen. Tissue samples were collected from two female and two castrated male sheep and flash frozen in liquid nitrogen. Chromatin immunoprecipitation (ChIP) was conducted for four histone marks; H3K4me3, H3K27ac, H3K4me1, and H3K27me3 known to be related to active or repressed chromatin states of gene regulatory elements. ChIP sequencing libraries were prepared and sequenced to >60 million reads each. Quality control parameters were performed with FastQC and Trim Galore, and high-quality reads were mapped to Oar\_rambouillet\_v1.0 with Bowtie2. Peaks were called for narrow marks using MACS2 and broad marks using SICER with false discovery rates of 0.05 and 0.01, respectively. Similarity between animals were examined using a Spearman correlation. Peaks were compared between tissues and sexes to identify states of conservation and differences. Chromatin states were characterized by implementing a Hidden Markov Model with 15 states in

ChromHMM. The study identified genomic positions enriched for these four histone marks in sheep, establishing the likely boundaries of regulatory elements involved with gene regulation in these two important tissues. These results will aid in future identification of regulatory mechanisms that influence economically important traits.

### **W036: Animal Epigenetics**

#### **Discovery of Regions of Genome Regulation in the Horse**

**Jessica L. Petersen**<sup>1</sup>, Nicole B. Kingsley<sup>2</sup>, Sichong Peng<sup>2</sup>, Colin Kern<sup>3</sup>, Catherine Creppe<sup>4</sup>, Erin N. Burns<sup>5</sup>, Huaijun Zhou<sup>3</sup>, James N. MacLeod<sup>6</sup>, Tomasz Szmatoła<sup>7</sup>, Ewelina Semik-Gurgul<sup>7</sup>, Tomasz Ząbek<sup>7</sup>, Ted Kalbfleisch<sup>8</sup>, Rebecca R. Bellone<sup>9</sup>, Carrie J. Finno<sup>10</sup> and Equine FAANG Consortium, (1)University of Nebraska-Lincoln, Lincoln, NE, (2)University of California, Davis, Davis, CA, (3)Department of Animal Science, University of California, Davis, CA, (4)Diagenode, Liège, Belgium, (5)University of California - Davis, Davis, CA, (6)Department of Veterinary Science, University of Kentucky, Lexington, KY, (7)National Research Institute of Animal Production, Balice, Poland, (8)University of Louisville, Louisville, KY, (9)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (10)UC Davis, Davis, CA

The horse occupies a unique role in animal genomics serving as an agricultural and companion animal species. In addition, the horse is an exceptional model for studies of clinically and economically important phenotypes. Genomic resources for the horse are robust; however, a comprehensive atlas of markers of genome regulation is lacking. The goal of our collaboration, as part of the Functional Annotation of ANimal Genomes initiative, is to characterize elements of genome regulation between the sexes and across tissues. Successes to date include characterization of the transcriptome (poly-A and small RNA) of 50 tissues from two Thoroughbred mares. Those data (<https://www.ebi.ac.uk/ena/data/view/PRJEB26787>) were utilized in annotation of the new equine reference genome (EquCab3). Four histone modification marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) have been mapped in ten tissues of each horse, resulting in over 1.8 million newly annotated regulatory elements, each covering from 0.6 to 5.1% of the genome. CCCTC-insulator marks have been explored in eight tissues, with an average of 53,000 CTCF-binding sites identified per tissue. ATAC-seq has been optimized in two tissues providing further information on open chromatin. Finally, to better understand tissue-specific patterns of DNA methylation, reduced representation bisulfite sequencing was completed on the eight prioritized tissues. These millions of newly annotated genomic features are being used in studies of equine-specific phenotypes as well as in cross-species studies of genome evolution. Ongoing work is extending data collection and genome annotation in tissues from the mares as well as in two Thoroughbred stallions, which will enable comparisons between sexes.

### **W037: Animal Epigenetics**

#### **Identification of Epigenetic Markers Predictive of Late Embryonic Mortality in Bovine Milk**

**Katy Kesler**<sup>1,2</sup>, Jessie Zenchak<sup>1</sup>, Kelly Sporer<sup>2</sup>, Casey Droscha<sup>2</sup> and Todd Byrem<sup>2</sup>, (1)Michigan State University, (2)CentralStar Cooperative

Reproductive efficiency is the most important factor in determining producer profitability. Yet, embryonic mortality contributes 56% of reproductive failure. Industry standards of pregnancy diagnostics include transrectal palpation, ultrasound and measurement of pregnancy-associated glycoprotein (PAG) levels via ELISA; however, no diagnostic can predict late embryonic mortality (LEM), one major contributor to reproductive failure. Developmental processes and epigenetic factors tightly control the timing and magnitude of gene expression. Circulating epigenetic factors such as microRNAs (miRNAs) to determine embryogenic fidelity is at the forefront of modern molecular diagnostics as in non-invasive prenatal testing. Circulating miRNAs in serum and milk are reliable non-invasive biomarkers of animal physiology due to their stable, sensitive, and specific nature. We hypothesized that LEM-specific miRNAs in milk are present, predictive and robust biomarkers of embryonic mortality. Milk represents a non-invasive and economical diagnostic medium for producers. This work will aid in the discovery of milk-based biomarkers predictive of LEM equipping producers with a novel diagnostic that delivers enhanced knowledge about pregnancy status empowering them to make more profitable breeding decisions. MicroRNAs are extracted using an optimized semi-automated protocol and miRNAs are verified using RT-qPCR. Candidate miRNAs are being profiled via RNA-seq and validated using RT-qPCR. MicroRNA-148a and miR-26a were characterized as potential endogenous controls due to their uniform presence during the lactation cycle. MicroRNA-222 and miR-25 were

screened as LEM candidates as proof of principle of dynamic physiologic targets. Physiologic miRNAs are present and informative in milk. Discovery of milk based epigenetic biomarkers will create a platform for diagnostics that can deliver improved knowledge about cow physiology. This work is contributing to the development of a new paradigm in dairy diagnostics; reforming the value added to a milk sample for dairy producers and practitioners.

### **W038: Animal Epigenetics**

#### **Diet Impacts the Avian Epigenome for Generations**

**Christopher M. Ashwell**, North Carolina State University, Raleigh, NC

### **W039: Animal Epigenetics**

#### **Prenatal Transportation Stress Alters Genome-Wide Methylation of DNA in Leukocytes from Brahman Bull Calves**

**Brittini P. Littlejohn**<sup>1,2</sup>, Deborah M. Price<sup>1,2</sup>, Don A. Neuendorff<sup>1</sup>, Jeffery A. Carroll<sup>3</sup>, Rhonda C. Vann<sup>4</sup>, Penny K. Riggs<sup>2</sup>, David G. Riley<sup>2</sup>, Charles R. Long<sup>1</sup>, Thomas H. Welsh, Jr.<sup>2</sup> and Ronald D. Randel<sup>1</sup>, (1)Texas A&M AgriLife Research & Extension Center, Overton, TX, (2)Department of Animal Science, Texas A&M University, and Texas A&M AgriLife Research, College Station, TX, (3)USDA-ARS, Livestock Issues Research Unit, Lubbock, TX, (4)Mississippi Agricultural and Forestry Experiment Station, Mississippi State University, Raymond, MS

The role of the prenatal epigenome in shaping postnatal outcomes is largely unknown in the bovine species. We compared DNA methylation in leukocytes from prenatally stressed (PNS) relative to Control calves. Mature Brahman cows were transported (n = 48) for 2-h periods at 60, 80, 100, 120, and 140 ( $\pm$  5) d of gestation or maintained as non-transported Controls (n = 48). Reduced representation bisulfite sequencing was used to assess differential methylation in leukocytes from a subset of 28-d-old bull calves (n = 7 PNS and 7 Control) born to Transported and Control dams. Samples from PNS calves contained 16,128 CG, 226 CHG, and 391 CHH sites that were differentially methylated relative to Control calves (C = cytosine; G = guanine; H = either adenine, thymine, or cytosine). Of the differentially methylated CG sites, 1,205 were located within promoter regions, 3,103 within introns, 1,260 within exons, and 10,560 within intergenic regions. Of those sites, 7,407 were hypermethylated ( $\geq$  10% more methylated than Controls;  $P \leq 0.05$ ) and 8,721 were hypomethylated ( $\geq$  10% less methylated than Controls;  $P \leq 0.05$ ). Because increased DNA methylation within gene promoter regions has been typically reported to suppress transcriptional activity, differentially methylated (difference  $\geq$  10%;  $P \leq 0.05$ ) CG sites located within promoter regions were used to predict alterations to biological pathways via pathway analysis (Ingenuity). In PNS calves, 113 pathways were identified as altered ( $P \leq 0.05$ ). Among these were pathways related to behavior, stress response, and immune function. Furthermore, PNS calves from the entire population of calves born to Transported and Control dams exhibited more excitable temperaments, increased circulating concentrations of cortisol, and differential innate immune response to an endotoxin challenge. In support of our hypotheses, differential methylation of genomic DNA was detected in PNS calves, which was correlated with phenotypic differences observed in the larger population of calves in this study. Results suggest DNA methylation as a mechanistic basis for prenatal programming. Better understanding mechanisms by which prenatal programming alters the developmental trajectory of biological systems *in utero* presents novel opportunities to improve livestock phenotypes.

### **W040: Animal Genomics and Adaptation to Climate Change**

#### **Genetic Factors Associated with Changes in Feeding Behavior Due to Elevated Temperature**

**Gary A. Rohrer**, USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, Clay Center, NE, Amanda J. Cross, DNA Genetics, Columbus, NE, Brittney N Keel, USDA, ARS, USMARC, Clay Center, NE, Tami Brown-Brandl, University of Nebraska-Lincoln and Joseph Cassady, South Dakota State University, Brookings, SD

Heat stress has negative impacts on pork production, particularly in the grow-finish phase. During heat stress events, feeding behavior of pigs is altered to reduce heat production. However, not all animals respond similarly to elevated ambient temperatures. To determine if genetic factors were associated with differences in feeding behavior at

different environmental temperatures, feeding behavior was studied year-round in a barn containing 6 pens of 40 pigs/pen. Pigs were placed in the barn at 8 weeks of age and removed after 12 weeks on the study. All pigs (n = 1653) were produced by sows from a common population (Landrace-Duroc-Yorkshire composite) and sired by Duroc, Landrace or Yorkshire boars. Pen assignments ensured uniform numbers of male and female pen mates for each breed of sire. Days were partitioned into categories based on their maximum temperature humidity index (THI): “Normal” (THI < 23.33°C), “Alert” (23.33°C < THI < 26.11°C), “Danger” (26.11°C < THI < 28.88°C) and “Emergency” (THI > 28.88°C). All pigs tended to reduce their feeding behavior during late afternoon and increase feeding behavior during the early evening when THI category exceeded Normal. Females had a greater reduction in feeding time during late afternoon hours on days with higher THI values than males. Breed of sire differences were also observed as Duroc-sired and Yorkshire-sired pigs increased feeding times in the mid-morning as well as early evening, whereas Landrace-sired pigs only increased feeding behavior in early evening when THI exceeded Normal. To avoid population stratification effects in genome-wide association studies (GWAS), phenotypic data were adjusted for breed of sire and sex prior to conducting a GWAS using genotypic data from ~60,000 SNP markers in GenSel. Candidate genes within regions identified by the GWAS include heat shock proteins and immune function. Differences in feeding behavior of grow-finish pigs due to increased temperatures were observed along with evidence that these differences are controlled by genetics. Selection for pigs that are more tolerant to elevated ambient temperature should enhance production and animal well-being.

USDA is an equal opportunity provider and employer.

#### **W041: Animal Genomics and Adaptation to Climate Change**

##### **Understanding of Genomic Properties and Development of Useful Genomic Resources from Polar Organisms: Antarctic Blackfin Icefish Genome Reveals Adaptations to Extreme Environments**

**Jin-Hyoung Kim**, Korea Polar Research Institute, Incheon, South Korea and Unit of Polar Genomics of KOPRI

Potential impacts and risks on global climate change and warming are more highlighted. In teleost, Notothenioids have evolved successfully to adapt to specific Antarctic marine ecosystems. Icefishes (suborder Notothenioidei; family Channichthyidae) are the only vertebrates that lack functional haemoglobin genes and red blood cells. Here, we report a high-quality genome assembly and linkage map for the Antarctic blackfin icefish *Chaenocephalus aceratus*, highlighting evolved genomic features for its unique physiology. Phylogenomic analysis revealed that Antarctic fish of the teleost suborder Notothenioidei, including icefishes, diverged from the stickleback lineage about 77 million years ago and subsequently evolved cold-adapted phenotypes as the Southern Ocean cooled to sub-zero temperatures. Our results show that genes involved in protection from ice damage, including genes encoding antifreeze glycoprotein and zona pellucida proteins, are highly expanded in the icefish genome. Furthermore, genes that encode enzymes that help to control cellular redox state, including members of the *sod3* and *nqo1* gene families, are expanded, probably as evolutionary adaptations to the relatively high concentration of oxygen dissolved in cold Antarctic waters. In contrast, some crucial regulators of circadian homeostasis (*cry* and *per* genes) are absent from the icefish genome, suggesting compromised control of biological rhythms in the polar light environment. The availability of the icefish genome sequence will accelerate our understanding of adaptation to extreme Antarctic environments.

#### **W042: Animal Genomics and Adaptation to Climate Change**

##### **Microbiome of Heat-Stressed Layer Hens**

**John C. Hsieh**<sup>1</sup>, Michael E. Persia<sup>2</sup>, Carl J. Schmidt<sup>3</sup> and Susan J. Lamont<sup>1</sup>, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Virginia Polytechnic Institute and State University, Blacksburg, VA, (3)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Increased frequency of extreme temperature events from climate change present a challenge to animal agriculture. During periods of elevated ambient temperature, both animal welfare and productivity might be negatively impacted. For poultry production, feed consumption is negatively correlated with increased ambient temperature. We hypothesize that exposure to high ambient temperature will alter the gut microbiota composition for chickens.

We designed a 4-week experiment with 80 24-week-old laying hens that were divided evenly between the thermoneutral group (TN, 23°C, n=40) and the high temperature group (HT, 35°C for 8 hours daily, n=40) to study changes of the cecal microbiome. Cecal medullary contents were collected from 8 random birds from TN and HT groups at 5 time periods after the start of HT group's exposure to heat: 3 hour (0WK), 1 week (1WK), 2 weeks (2WK), 3 weeks (3WK), and 4 weeks (4WK). Amplicons of the V1-V3 region of the 16S rRNA gene were generated from DNA extracted from the cecal contents. Sequencing of the amplicons were performed with the Illumina MiSeq platform to generate 300-bp paired-end reads. Sequences were processed with the *mothur* MiSeq SOP followed by statistical analysis in R with *phyloseq* and *microbiome* packages.

An average of over 100,000 unique sequences per sample were assigned to the 20,295 OTUs found. Compositional analysis found that an overwhelming majority (>95%) of the relative abundance was from the *Firmicutes* and *Bacteroidetes* phyla. Further examination of these phyla showed that the *Firmicutes* phyla was mostly represented by members from the *Ruminococaceae* and *Lachnospiraceae* families and the *Bacteroidetes* phyla was mostly represented by members from the *Rikenellaceae* family. All 3 families are typically found in animal gut microbiome samples and have a commensal relationship with the host. Analysis of the alpha diversity showed a significant difference between TN and HT groups at 1WK ( $p = 0.021$ ) and 4WK ( $p = 0.05$ ), however, ordination plots did not show distinct clusters between samples from the TN and HT groups. Further group-level comparisons based on beta diversity using permutational analysis of variance (PERMANOVA) were performed for the 1WK and 4WK samples. PERMANOVA only found significant difference between TN and HT groups at 1WK ( $p = 0.035$ ) and not at 4WK ( $p = 0.229$ ). The coefficients of the 1WK PERMANOVA indicates that the difference between TN and HT is most highly driven by members from the genus *Alistipes* of the family *Rikenellaceae*. Many members from the family *Ruminococaceae* (n=8) also made the list of top 20 most influential OTUs driving the difference between TN and HT as predicted by the 1WK PERMANOVA.

From this study, we were able to detect significant differences in the composition of cecal microbiota between TN and HT groups after 1 week, suggesting the cecal microbiome did respond to treatment of heat exposure. We were also able to identify key families of bacteria that changed in abundance at the 1-week time point, but the level of resolution from 16S amplicon sequencing was not sufficient to fully resolve the genus/species of the bacteria.

## **W043: Animal Genomics and Adaptation to Climate Change**

### **Detection of Selection Signatures among Chicken Population under Different Climatic Conditions**

**M. Walugembe**<sup>1</sup>, Francesca Bertolini<sup>2</sup>, Chandraratne M.B. Dematawewa<sup>3</sup>, Ahmed R. Elbeltagy<sup>4</sup>, Carl J. Schmidt<sup>5</sup>, Max F. Rothschild<sup>1</sup> and Susan J. Lamont<sup>1</sup>, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)National Institute of Aquatic Resources, Denmark, (3)University of Peradeniya, Faculty of Agriculture, Department of Animal Science, Peradeniya, Sri Lanka, (4)Animal Production Research Institute, Dokki, Cairo, Egypt, (5)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Changes in climate, particularly heat with drought or high humidity, are a major challenge to livestock performance. Animal survival resulting from both adaptation to the local environmental conditions and from artificial selection for production traits are expected to leave selection signatures in the genome. The aim of this study was to identify selection signatures that may be associated with the adaptation of indigenous chickens from two different climatic regions (Sri Lanka = Tropical; Egypt = Arid) and in non-indigenous chickens that derived from human migration events to the generally tropical State of São Paulo, Brazil. Analyses were conducted using fixation index (Fst) and hapFLK analyses. Chickens from Brazil, Sri Lanka and Egypt were genotyped using the Affymetrix Axiom®600k Chicken Genotyping Array. Pairwise Fst analyses among countries did not detect major regions of divergence between chickens from Sri Lanka and Brazil, with ecotypes/breeds from Brazil appearing to be genetically related to Asian-Indian (Sri Lanka) ecotypes. Several differences, however, were detected in comparisons of Egyptian with either Sri Lankan or Brazilian populations, and common regions of difference were detected on chromosomes 2, 3 and 8. The hapFLK analyses for the three separate countries suggested unique regions that are potentially under selection on chromosome 1 for all three countries, on chromosome 4 for Sri Lankan, and on chromosomes 3, 5, and 11 for the Egyptian populations. Some of identified regions under selection with hapFLK analyses contained genes such as TLR3, SOCS2, EOMES and NFAT5, all of which are associated with immune function. This study,

therefore, suggests that immune function may be an important biological function that aids in adaptation mechanisms in response to arid and tropical environments.

#### **W044: Animal Genomics and Adaptation to Climate Change**

##### **Poultry Genomic Projects in Low and Middle Income Countries**

**Androniki Psifidi**<sup>1</sup>, Olivier Hanotte<sup>2</sup>, Lonneke Vervelde<sup>3</sup>, Nick Sparks<sup>4</sup>, Joshi Chaitanya<sup>5</sup>, Damer Blake<sup>1</sup>, Georgios Banos<sup>6</sup>, Fiona Tomley<sup>1</sup> and One Health Poultry Hub consortium, (1)Royal Veterinary College, University of London, Hatfield, United Kingdom, (2)LiveGene – CTLGH, International Livestock Research Institute, Addis Ababa, Ethiopia, (3)The Roslin Institute and R(D)SVS, University of Edinburgh, United Kingdom, (4)Scotlands' Rural College, United Kingdom, (5)College of Veterinary Science and Animal Husbandry, Anand, India, (6)The Roslin Institute and Scotlands' Rural College, Midlothian, Scotland

High population growth, urbanisation and rising incomes have led to increased demands for animal protein. Poultry meat and eggs are now the leading global source of protein for humans and demand is predicted to increase more in the coming decades. A major challenge is to achieve sustainable poultry expansion whilst reducing risk to animal and human health. In Low and Middle Income Countries (LMICs), where rates of intensification in poultry production are highest, vaccinations and biosecurity measures are commonly sub-optimal. Up to 60% of poultry stock can die due to predation and infectious disease. Breeding for increased resistance to zoonotic and other important poultry pathogens, as well as for beneficial microbiome composition with relevance to competitive exclusion, offers an attractive strategy to control infectious disease. We have been studying the genetics of infectious disease in African and Asian indigenous and tropically adapted commercial chickens. Our initial results have been encouraging since a low to moderate but significant heritability has been estimated and several quantitative trait loci have been identified for immune, disease and microbiome traits. Ongoing studies, which will be discussed during this presentation, aim to inform the best strategies to control infectious disease in LMICs.

#### **W045: Apiaceae**

##### **Trait Identification and Genomic Database Development for Carrot (*Daucus carota*) Improvement**

**Philipp W. Simon**, USDA, ARS and University of Wisconsin, Madison, WI, Shelby L. Ellison, University of Wisconsin-Madison, Madison, WI and Douglas Senalik, USDA, ARS, Madison, WI

Carrot is the most economically important and genetically well-characterized member of the Apiaceae, and more than 13,000 germplasm accessions are held in global collections, but that germplasm is not adequately phenotyped, and molecular and bioinformatics tools to incorporate important traits during cultivar development are not well-developed. To begin to remedy this shortcoming, the USDA-NIFA-SCRI project 2016-51181-25400 “Identifying phenotypes, markers, and genes in carrot germplasm to deliver improved carrots to growers and consumers” is phenotyping ~700 accessions of cultivated carrot from the USDA-NPGS germplasm collection and ~50 publically-developed inbreds and populations are being phenotyped for 12 traits identified by stakeholder input as important for future cultivar development. Several phenotyped plants from each accession are being resequenced, and the CarrotOmics database is being developed. To date, new germplasm sources to improve *Alternaria* leaf blight, root-knot nematode, cavity spot, and bolting resistance, stand establishment, drought tolerance, flavor, color and nutritional value have been identified. Novel candidate genes for carotenoid and anthocyanin pigments have been discovered and phylogenetic relationships of the Apiaceae have been clarified. Phenotyping tools to characterize the genetic architecture of shoot and root morphology and growth are being developed and methods to more accurately predict traits in germplasm that has not been phenotyped are being evaluated. Beyond germplasm phenotyping and genotyping activities, the project team is developing breeding pools enriched for favorable alleles of key traits, evaluating the bioefficacy of carrots that vary in nutrient composition, and assessing the economic impact of traits newly introduced to carrot. Well-characterized germplasm and genomic tools to facilitate future carrot improvement will benefit future growers and consumers.

#### **W046: Apiaceae**

## **Improved Hybrid *de novo* Genome Assembly, Gene Prediction and Annotation of Carrot (*Daucus carota*)**

**Massimo Iorizzo**<sup>1</sup>, Hamed Bostan<sup>2</sup>, Shelby L. Ellison<sup>3</sup>, Douglas Senalik<sup>4</sup>, Philipp W. Simon<sup>4</sup> and Julien Curaba<sup>2</sup>, (1)Plants for Human Health Institute, Department of Horticultural Science, North Carolina State University,, Kannapolis, NC, (2)Plants for Human Health Institute, North Carolina State University, NCRC, 600 Laureate Way, Kannapolis, NC, 28081, USA, (3)University of Wisconsin-Madison, Madison, WI, (4)USDA, ARS and University of Wisconsin, Madison, Madison, WI

The release of the carrot genome sequence (v2.0) in 2016 has rapidly enhanced molecular and genomic research for this species, changing the nature of research in carrot biology. Research is shifting more towards extensive genetic screening for genome-wide association analysis and functional genomics. However, despite the high-quality of the current carrot genome assembly v2.0 release, improvements are needed due to the multiple challenges associated with short read sequencing data used to develop it. Here, an improved genome assembly, gene prediction and annotation of carrot DH1 (v3.0) is presented. The new genome assembly sequence covers 440 Mb, with a contig N50 >6Mb, all assembled into 9 pseudomolecules/chromosomes and 2 organellar genomes. Compared with the previous assembly (v2.0), the v3.0 assembly includes about 11% (54 Mb) of novel nucleotide sequence, >21% (>100 Mb) new sequences anchored at the chromosome level, and represents a >193 fold increase in contig N50. Using a combination of IsoSeq and Illumina transcriptome data, 36,216 gene models were predicted, with >4,000 additional gene models as compared to the previous gene prediction. Taking advantage of the IsoSeq full-length high-quality transcripts, >6,000 mis-predicted and partial gene models in v2.0 were identified and manually curated. Finally, a comprehensive catalog of alternative splicing (AS) events in the carrot DH1 was obtained, and efforts to identify tissue specific isoforms and AS events are ongoing.

## **W047: Apiaceae**

### **Genome-Wide Association Analysis of Carotenoids in Carrot**

**Shelby L. Ellison**, University of Wisconsin-Madison, Madison, WI, Phil Simon, USDA, ARS and University of Wisconsin, Madison, Madison, WI and Douglas Senalik, USDA, ARS, Madison, WI

Carotenoids are 40-carbon isoprenoid molecules that produce the red, yellow and orange pigmentation found in nature. Carotenoids play an essential role in light capture and photoprotection in plants. In humans, pro-vitamin A carotenoids, such as beta-carotene, are converted to vitamin A, which is critical for maintaining healthy vision, immune response, and cellular communication and differentiation. Non-pro-vitamin A carotenoids, including lycopene and lutein, have been associated with increased antioxidant activity, conferring benefits such as the prevention of age-related macular degeneration and prostate cancer. Orange carrots accumulate large quantities of alpha- and beta-carotene, while yellow and red carrots acquire significant quantities of lutein and lycopene, respectively. While root color can be scored and selected with high-throughput, low-cost phenotypic methods within breeding programs, it remains to be well established what the genetic loci to target for selection would be for orange, red, and yellow carrots. Carotenoid content from over 700 cultivated carrots from the US National Plant Germplasm System was assessed via HPLC to perform a genome-wide association study (GWAS) to identify genomic regions conferring carotenoid accumulation. Phenotypic evaluation of pigmentation was also determined by colorimetry and several strong correlations between phenotypic methods were identified. Assessment of pigmentation via colorimetry may offer a fast and inexpensive method for high throughput phenotyping in the future. Marker-trait associations were observed for all carotenoids evaluated, including previously verified and novel associations. The results pave the way for creating a robust marker-assisted selection strategy for enriched pigmentation in carrot breeding programs.

## **W048: Apiaceae**

### **Characterization of the Tendency for Bolting among Carrot Germplasm Accessions**

**Allen Van Deynze**<sup>1</sup>, Theresa Hill<sup>2</sup>, Armando Garcia-Llanos<sup>1</sup>, Shelby L. Ellison<sup>3</sup>, William Matthews<sup>1</sup>, Philip A. Roberts<sup>4</sup>, Douglas Senalik<sup>5</sup> and Phil Simon<sup>5</sup>, (1)University of California, Davis, CA, (2)Department of Plant Sciences, Seed Biotechnology Center, University of California, Davis, California,

USA, Davis, CA, (3)University of Wisconsin-Madison, Madison, WI, (4)University of California, Riverside, CA, (5)USDA, ARS and University of Wisconsin, Madison, Madison, WI

Modern, carrot cultivars are biennials, requiring exposure to cold temperatures in order to flower. Initiation of flowering, referred to as bolting, is an undesirable trait for growers and consumers as it stimulates fiber development in the root, rendering the crop unmarketable. Wild accessions, an invaluable source of genetic diversity, can provide traits that address critical needs of the carrot industry such as resistance to biotic and abiotic stress, reliable stand establishment, and improved consumer quality. However, the wild carrot germplasm includes annual, early flowering plants that are unsuitable for commercial cultivation. Understanding the genetic basis of bolting in carrot will enable breeding strategies that leverage the diversity of wild carrot accessions. We have characterized bolting frequencies within the first season of cultivation for 695 accessions, representing the breadth of carrot genetic diversity. Bolting was scored across multiple locations, including stress environments, using a randomized complete block design with two replications. Thirty to 40% of accessions bolted in at least one trial. Among bolting accessions, one-half were environment-specific and bolting frequency had a continuous distribution indicating quantitative genetics with a strong environmental interaction. A single gene has been identified as having a major role in controlling carrot floral induction but this study indicates additional genes underlie the trait. A genome wide association study will be reported on including candidate genes to bolting in carrot.

## **W049: Apiaceae**

### **Abundance and Insertional Polymorphism of Carrot Mites and Demography of *Daucus carota***

Alicja Macko-Podgorni, Katarzyna Stelmach, Kornelia Kwolek and **Dariusz Grzebelus**, University of Agriculture in Krakow, Krakow, Poland

Miniature inverted repeat transposable elements (MITEs) are small non-autonomous DNA transposons ubiquitous in plant genomes, mobilized by their autonomous relatives. *Stowaway* MITEs are derived from and mobilized by elements from the *mariner* superfamily. They constitute a significant fraction of the carrot genome, however there is lack of comprehensive analysis of variation caused by *Daucus carota* *Stowaway* MITEs (*DcStos*), their association with genes and putative impact on the genome evolution.

We report on 14 families of carrot *Stowaways*, *DcStos*, jointly occupying ca. 0.5% of the host genome. We systematically mined 31 genomes of wild and cultivated *D. carota*, which yielded 18.5 thousand copies showing a remarkable insertion site polymorphism. The genomic distribution of *DcStos* differed with respect to the origin of host populations corresponding with the four major groups of *D. carota* (wild European, wild Asian, eastern cultivated, western cultivated). We showed that *DcStos* were associated with genes and occurred most frequently in 5' and 3' UTRs. Individual families differed in their propensity to reside in particular segments of the genic region. Most importantly, *DcSto* copies in the 2kb up- and downstream regions were more frequently associated with genes encoding transcription factors, suggesting their possible functional impact. More than 1.5% of all *DcSto* insertion sites comprised different copies in exactly the same position in different host genomes, indicating the existence of insertional hotspots. The *DcSto7b* family was much more polymorphic than the remaining families in the cultivated carrot. We showed a line of evidence pointing at its activity in the course of carrot domestication and identified *Demar1* as an active carrot *mariner* element and a possible source of the transposition machinery for *DcSto7b*. *DcSto* intron length polymorphisms (DcS-ILPs) detected substantial genetic diversity and, showing considerable discrimination power, may be exploited as a tool for germplasm characterization and analysis of genome relationships. *DcSto* insertions mined from eastern cultivated carrots were usually much less frequent than those mined from the reference genome, possibly reflecting a bottleneck at the origin of the western carrot gene pool.

## **W050: Application of Genetics and Genomics for Large-scale Breeding Programs**

### **Developing Durable Disease Resistance**

**T Lynne Reuber**, Two Blades Foundation, San Mateo, CA

Plant disease is an important limitation on agricultural production, causing average losses of 11-30% average global losses each year. Genetic disease resistance is the most environmentally sustainable solution, but it is challenging to create durable resistance since single modes of action are rapidly overcome by pathogen evolution. The explosion of genomic information is allowing us to mine the genetic potential of wild crop relatives, but these resources must be

managed and deployed sustainably. One potential solution is the deployment of multiple independent resistance genes stacked at a single locus.

2Blades Foundation works with a network of international collaborators to discover and deploy new sources of resistance for important crop diseases. Our work in soy, wheat, and potato will be discussed.

## **W051: Application of Genetics and Genomics for Large-scale Breeding Programs**

### **Data Driven Plant Breeding**

**Christian R. Werner** and John M. Hickey, University of Edinburgh, The Roslin Institute, Edinburgh, United Kingdom

Breeding is increasingly viewed as an integrative science that has data science at its core. Advances in the fields of genomic prediction, genome sequencing, genome editing and the use of simulation to evaluate alternative breeding systems offer huge potential to drive step changes in crop performance. Untapping the potential of these technologies requires their development and deployment in concert to build upon a platform and culture of data science.

In this presentation, two concepts of data-driven breeding to improve the efficiency of breeding programs will be outlined.

Firstly, the concept of a “two-part” strategy for the deployment of genomic selection in crop breeding will be presented. The “two-part” strategy reorganises plant breeding programs into two separate but connected parts: (i) a population improvement part which rapidly improves the average performance of a population; and (ii) a product development part which extracts and disseminates products to farmers. Stochastic simulation has shown that the two-part strategy has the potential to more than double the rate of genetic gain achieved by breeding programs.

Secondly, the concept of Promotion of Alleles by Genome Editing (PAGE) in crop breeding will be presented. PAGE recognises that many important traits in crop production are controlled by thousands of individual genetic variants, each with small effects on overall performance. Stochastic simulation has shown that PAGE, driven by modest amounts of genome editing, has the potential to more than double the rate of genetic gain achieved by breeding programs. To deploy PAGE a process of genetic variant discovery that has sufficient throughput is required. The concept of “Allele Testing”, a process that integrates several steps to achieve the discovery of genetic variant in sufficient throughput will be outlined.

## **W052: Application of Genetics and Genomics for Large-scale Breeding Programs**

### **A New Paradigm for Beef Cattle Genetic Improvement**

**Fabiana Mokry**, Genus ABS, Deforest, WI, Scott Newman, Genus plc, Ziqing Weng, ABS GLOBAL, DeForest, WI and Matthew Cleveland, Genus plc, Hendersonville, TN

The beef industry faces the challenge of meeting growth for global beef demand through a supply chain designed to minimize increases in land, feed and waste products (while accounting for animal welfare) using a combination of advanced recording, analytical, and molecular technologies while increasing beef supply chain profitability. One of the key components to implementation relies on development and application of IoT technologies to automate day-to-day farm operations, multi-dimensional data collection, and compilation of novel phenotypes that, when combined with other technologies, are essential for improving beef production efficiency. The automated reality of IoT technologies brings new challenges, requiring development of artificial intelligence algorithms (i.e., machine learning, neural networks) and computing capabilities (i.e., cloud storage and computing) to harness the power of multi-dimensional data. Another developing technology with high impact in modern beef production is blockchain and smart contracts for animal traceability from farm to end consumer, transparency, and data sharing. There is also progress at the molecular and biological levels through sexed semen technologies, gene editing, gut microbiome and host genetics relationships, low cost/low density genotyping and sequencing methodologies that combined with IoT are essential for accelerating genetic gain. The key to meet growing global beef demand relies on harnessing the power of all these new technologies to accelerate genetic gain, develop target genetics for efficiency and

sustainability according to local resources, disseminate improved genetics, and ensure profitability and sustainability along the beef supply chain.

### **W053: Application of Genetics and Genomics for Large-scale Breeding Programs**

#### **Development of Cibus' Trait Machine™ to Efficiently Apply Gene Editing**

**Andrew Walker**, Cibus US LLC, San Diego, CA

An increasing number of traits are being discovered and developed in plants using gene editing and genomics. Simple traits in plants orchestrated by a small number of genes can be managed through backcross breeding, but more complex traits mediated by many genes cannot easily be backcrossed into new elite breeding lines. The concept of a Trait Machine™ is defined as the optimized process to efficiently develop complex traits directly in elite plant lines, which accelerates plant breeding and commercialization. The components of a Trait Machine might include robust tissue culture protocols, efficient gene editing and molecular screening, robotics, accelerated plant growth cycles in controlled environments, and phenotyping. Cibus has developed all components of a Trait Machine for canola and is applying it to develop multiple complex traits directly in elite canola lines.

### **W054: Application of Genetics and Genomics for Large-scale Breeding Programs**

#### **Use of Advanced Technologies to Digitally Engineer Seed Products at Corteva Agriscience**

**Radu Totir**, Corteva, Johnston, IA

All commercial breeding programs used by Corteva Agriscience to create seed products for all our customers, across crops and geographies, are designed to optimally leverage advanced technologies within routine high throughput integrated genetic evaluation systems. Digital engineering techniques are key to enable the seamless integration of cross-functional applied science to generate and convert into actionable knowledge high volume data streams that enable Corteva breeders to: a) understand and characterize the targeted population of genotypes under active selection b) perform high speed and throughput genetic evaluation conditional on the targeted population of environments c) accelerate “on growing site” performance improvements conditional on the targeted population of traits under selection; with the ultimate goal to develop superior seed products that enrich the lives of growers and consumers across the globe. Examples of leveraging digital engineering to integrate genomic blueprint knowledge with biological and environmental conditions for seed product design will be discussed. Comments will be made regarding the need for digital engineering guided cross-functional applied science solutions to solve challenging problems in production agriculture.

### **W055: Application of New Genomic Tools and Techniques in Arthropods**

#### **Application of New Genomic Tools and Techniques in Arthropods**

**Marce Lorenzen**, William Klobasa, Ordomb Huot, Nathaniel Grubbs, Dorith Rotenberg and Anna Whitfield, North Carolina State University

The ability to utilize state-of-the-art genomic tools can open the door to a wide range of new pest control solutions, including the use of gene drive. However, there are considerable barriers to deploying these tools in a new species. For example, despite the power of CRISPR/Cas9-based genome editing, published reports of confirmed genome editing in hemipteran species are rare. Here, we will describe a method for editing a new hemipteran species, the corn planthopper, *Peregrinus maidis*. While we had an excellent “guidebook” to follow – CRISPR/Cas9-based genome editing of the brown planthopper, *Nilaparvata lugens* – the process of deploying this technology in a new planthopper species was still challenging. Therefore, we plan to offer a number of tips and tricks to aid others working with hemipteran species. We started with the development of genomic and transcriptomic resources to identify promoters for *P. maidis*. The next steps involved developing protocols for harvesting and microinjecting precellular embryos, which are critical for Cas9-based genome editing, since DNA and proteins cannot easily cross cell membranes. The later steps cover Cas9-based genome editing itself, including selection of target genes. However, while Cas9-mediated knockout of the pivotal eye-color gene, *white (w)* was successful in generating *P. maidis* embryos that lacked eye-spot pigmentation, only embryos that still possessed eye-spot pigmentation hatched. Since this suggests that *w* may play a vital role in *P. maidis* development, we tested another eye-color gene, *cinnabar (cn)*. Unlike *w*, loss-of-function mutations in *cn* had no apparent impact on development, allowing us to

establish a red-eyed *P. maidis* colony. Our current work focuses on a more difficult challenge, that of editing the genome of the bird cherry-oat aphid, *Rhopalosiphum padi*, during clonal reproduction. Importantly, these are the first steps towards a much broader goal: bringing game-changing genomic tools to bear on understanding and controlling these important agricultural pests.

### **W056: Application of New Genomic Tools and Techniques in Arthropods**

#### **BAPC-Assisted CRISPR/Cas9 System: Targeted Delivery into Adult Ovaries for Heritable Germline Gene Editing (Hemiptera)**

**Wayne Hunter**, USDA-ARS, Fort Pierce, FL and John Tomich, Kansas State University, Manhattan, KS

Production of heritable gene edits in Hemiptera species has been challenging. Presented is a method using BAPCtofect™ to improve the delivery of CRISPR components into adult Asian Citrus Psyllid, *Diaphorina citri*, by direct microinjection into females near ovaries to produce heritable germline gene editing. Branched Amphipathic Peptide Capsules™ (BAPC™)(Phoreus Biotech, <https://phoreusbiotech.com/>) improved delivery of CRISPR components, plasmids, and dsRNA into psyllids (50% success versus 36% using commercially produced materials). *D. citri*, is a vector of bacterial plant pathogens, *Liberibacter asiaticus* that results in citrus tree-decline and death. Reduction of psyllid populations is key to management of the disease, Huanglongbing. Using this method multiple heritable gene Knockouts, KO were produced and detected in G1 and G2 offspring post injection of G0 adult females. The KO of thioredoxin & Vermillion, produced genetic, phenotypes (white eye), with significantly reduced fecundity (Treated 3 eggs vs. Control treated 128 eggs). Reduced adult longevity (Treated 8 vs 20d in Control). BAPCtofect-assisted CRISPR delivery opens new strategies to manage insect vectors, especially within the Hemiptera. Psyllid population reduction will help protect citrus, reduce bacterial spread and increase sustainable citrus production [Genomic Data open source at: [www.citrusgreening.org](http://www.citrusgreening.org)]. Funding in part: USDA-NIFA 2014-70016-23028. Developing an Infrastructure and Product Test Pipeline to Deliver Novel Therapies for Citrus Greening Disease. Hunter, W.B., Gonzalez, M.T., Tomich, J. 2018. Preprint: BAPC-assisted CRISPR/Cas9 System: Targeted Delivery into Adult Ovaries for Heritable Germline Gene Editing (Arthropoda: Hemiptera). bioRxiv 2018. <http://dx.doi.org/10.1101/478743>

### **W057: Application of New Genomic Tools and Techniques in Arthropods**

#### **Novel Trans-Complementing Split-Gene Drive System Provides Flexible Application for Safe Laboratory Investigation**

**Valentino M. Gantz**, Cell and Developmental Biology, University of California, San Diego, La Jolla, CA

CRISPR-based gene drives spread through populations bypassing the dictates of Mendelian genetics, offering a population-engineering tool for tackling vector-borne diseases, managing crop pests, and helping island conservation efforts. Current technologies raise safety concerns for unintended gene propagation. We address this by splitting the two drive components, Cas9 and gRNAs, into separate alleles to form a novel trans-complementing split-gene-drive (tGD). We demonstrate the ability of the tGD to promote super-Mendelian inheritance of two separate transgenes in the fruit fly *Drosophila melanogaster*. Lastly, we take advantage of the bi-component nature of the tGD to optimize individual transgenes and investigate the biology of component inheritance during the gene drive process.

### **W058: Application of New Genomic Tools and Techniques in Arthropods**

#### **The ABCs of using CRISPR in Non-Model Organisms**

**Fu-Chyun Chu** and Aaron T. Dossey, All Things Bugs LLC, Oklahoma City, OK

The idea of insects as sustainable food ingredients has been the focus of a new emerging industry in recent years. As with any crop, insects can be engineered via gene editing and other means to provide more desirable crop phenotypes. Our goal is to develop methods for using CRISPR/Cas9 to genetically engineer the two most commonly mass-produced edible insects: the yellow mealworm (*Tenebrio molitor*) and the house cricket (*Acheta domestica*). We tested both knock-out and knock-in methods, and compared the efficiency of both methods, within and between

species. We targeted *vermilion* genes in both species to test the knock-out efficiency, and used DNA constructs carrying enhanced green fluorescent protein (EGFP) genes driven by species-specific promoters for the knock-in tests. Tests were done with either one or three single guide RNAs (sgRNAs). For the *T. molitor* EGFP knock-in DNA construct, we used the promoter from the muscle actin gene, so expected to see a muscle expression pattern. When we screened G<sub>0</sub> eggs and 1<sup>st</sup>-instar larvae, we found that 60% of the one-sgRNA treatment individuals had somatic expression of EGFP in muscles, but only 15% from the three-sgRNA treatment had EGFP expression. Interestingly, 43% of eggs which only received the EGFP knock-in DNA construct (no sgRNA) also had EGFP expression, but this expression was transient and mostly outside of the muscles. Overall, ~39% of G<sub>0</sub> crosses from either sgRNA treatment produced white-eyed progeny. However, the knock-in efficiency was substantially different between treatments: 20% of the one-sgRNA G<sub>0</sub> crosses produced EGFP positive G<sub>1S</sub>, but no EGFP positive G<sub>1S</sub> were found in three-sgRNA or no-sgRNA treatments. For *A. domesticus*, we used the *Polyubiquitin* promoter to drive EGFP expression. Unfortunately, very few G<sub>0S</sub> (< 1%) expressed EGFP: only 1 from the one-sgRNA treatment, and 7 from the three-sgRNA treatment. However, nearly all G<sub>0</sub> crosses from both treatments produced white-eyed progeny. Checking for the knock-in in *A. domesticus* G<sub>1S</sub> was more difficult than expected, requiring use of both molecular and fluorescent screening to confirm the knock-in. Despite this difficulty, over 70% of G<sub>0</sub> crosses from either treatment produced EGFP knock-ins. These results show that we have successfully used the CRISPR/Cas9 system to generate 2-marker phenotypes in two insect species in single experiments by knocking out an eye-color gene while knocking in a fluorescent marker. Our successful experiments also demonstrate species-specific differences in knock-out and knock-in efficiencies that should be considered in future experiments. By showing these two marker system in *A. domesticus* and *T. molitor*, we demonstrate the potential to engineer insects for beneficial phenotypes.

## **W059: Application of New Genomic Tools and Techniques in Arthropods**

### **Knock Down of Imidacloprid Resistant Genes in Colorado Potato Beetle, *Leptinotarsa decemlineata* (Chrysomelidae: Coleoptera)**

Muhammad Nadir Naqqash, Ayhan Gökçe and Allah Bakhsh, Nigde Omer Halisdemir University, Nigde, Turkey

Colorado potato beetle, *Leptinotarsa decemlineata* Say (coleoptera: chrysomelidae), is the important pest of potato all over the world. This insect pest is resistant to more than 50 active compounds belonging to various chemical groups. Resistance management of imidacloprid resistant Colorado potato beetle (CPB) with eco-friendly method of RNA interference (RNAi) mediated gene silencing was tested under laboratory conditions. Three important imidacloprid resistance conferring genes viz. belonging to cuticular protein (CP), cytochrome p-450 monooxygenases (p-450) and glutathione synthase (GST) were amplified between two promoters of L4440 vector in *Escherichia coli* strain HT-115. Feeding bio-assays were conducted on imidacloprid resistant CPB lab colony by applying bacteria expressing dsRNA on potato leaflets. Significantly higher mortality was observed in all instars when they were exposed to dsRNA targeting CP. The mortality rate in 1<sup>st</sup> instar was 99.97% with CP-dsRNA as it was 67.38±0.22% in 2<sup>nd</sup> instar after 6 days of exposure. About 50.60±0.16% and 8.65±0.89% mortality was observed in 3<sup>rd</sup> and 4<sup>th</sup> instar larvae fed on CP-dsRNA for 3 days. Survival rate of insects exposed to CP-dsRNA decreased to 0.00±0.00%, 2.21±0.88%, 13.30±0.60% and 47.35±0.19% in 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> instar before reaching the adult stage. Significantly lower weight was calculated in 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> instar larvae exposed to dsRNA targeting CP. Synergism of RNAi with imidacloprid was conducted on the resistant 2<sup>nd</sup> instar which revealed that survival rate decreased to 0.00±0.00% in case of GST and CP treatments after applying reduced dose of imidacloprid. These results revealed that RNAi targeting CP, p-450 and GST enzymes could be useful tool in management of insecticide resistant CPB populations.

## **W060: Aquaculture**

### **A Long Reads-Based *de-novo* Assembly of the Rainbow Trout Arlee-Line Genome**

Guangtu Gao, USDA-ARS-NCCCWA, Kearneysville, WV

Although the most recent version of the rainbow trout genome assembly from the Swanson line has greatly improved the genome reference and is reliable for genes' prediction, it contains 420,055 spanned gaps and 7,839 unspanned gaps (GCA\_002163495.1). Hence, there is still a need to improve the contiguity and completeness of the reference assembly, which is now possible with long-read DNA sequencing technologies. Currently, we are also

working towards generating a rainbow trout “pan-genome” reference that will better represent the genetic diversity in this species. The Arlee doubled haploid YY male line has a different genetic background from the Swanson line. It was originated from a domesticated strain that was originally collected from the northern California coast. For the Arlee genome assembly, we generated 111x genome coverage in long-read sequence data using the PacBio Sequel system. The read length distribution has N50 of ~33 kb and an average read length greater than 20 kb. Contigs were assembled using the Canu pipeline and consensus sequence was error-corrected using two iterations of Arrow with the PacBio reads followed by one iteration of FreeBayes using Illumina paired-end reads. The Canu assembly contained 1,591 contigs with an N50 contig length of 9,835,815 bp, which is a major improvement in contiguity compared to the current Swanson assembly. The assembly was further improved with a Bionano optical map and Hi-C proximity ligation sequence data to produce super-scaffolds and correct mis-joined scaffolds. This improved the assembly to a total of 2.34 Gb in 919 scaffolds with an N50 length of 47,542,702 bp. The range of the scaffolds’ length distribution after Bionano and Hi-C was 16,956 bp – 90,526,592 bp. A BUSCO analysis detected 96.6% of conserved Actinopterygii gene content in this assembly. We are currently using the rainbow trout high-density genetic map to guide chromosomal alignment of scaffolds.

## **W061: Aquaculture**

### **Differences in Global Gene Expression Response to Dermo Disease Among Eastern Oyster Families Reveal Mechanisms of Resistance**

**Dina A. Proestou**, USDA ARS National Cold Water Marine Aquaculture Center, Kingston, RI

Dermo disease, caused by the protozoan parasite *Perkinsus marinus*, negatively impacts wild and cultured Eastern oyster populations, yet our knowledge of the mechanistic bases for parasite pathogenicity and the Eastern oyster’s response to it is limited. To better understand host responses to the parasite and identify molecular mechanisms underlying disease-resistance phenotypes, we experimentally challenged two families exhibiting divergent Dermo-resistance phenotypes with the parasite, generated global expression profiles using RNAseq and identified differentially expressed transcripts between control and challenged oysters from each family at multiple time points post-parasite injection. The susceptible and resistant families exhibited strikingly different transcriptomic responses to the parasite over a 28-day time period. The resistant family exhibited a strong, focused, early response to *P. marinus* infection, where many significantly upregulated transcripts were associated with the biological processes “regulation of proteolysis” and “oxidation-reduction process.” *P. marinus* virulence factors are mainly comprised of proteases that facilitate parasite invasion and weaken host humoral defenses, thus host upregulation of transcripts associated with negative regulation of proteolysis is consistent with a Dermo-resistant phenotype. In contrast, the susceptible family mounted a very weak, disorganized, initial response to the parasite. Few transcripts were differentially expressed between control and injected oysters, and no functional enrichment was detected among them. At the final 28 d time point 2450 differentially expressed transcripts were identified and were associated with either “G-protein coupled receptor activity” (upregulated) or “microtubule-based process” (downregulated). A handful of protease inhibitors were differentially expressed between control and injected susceptible oysters, but this function was not enriched in the susceptible data set. The differential expression patterns observed in this study provide valuable insight into the functional basis of Dermo resistance and suggest that the timing of expression is just as important as the transcripts being expressed.

## **W062: Aquaculture**

### **DNA Methylation Dynamics in Atlantic Salmon (*Salmo salar*) after being Challenged with High Temperature and Moderate Hypoxia**

**Anne Beemelmans**<sup>1</sup>, L. Ribas<sup>2</sup>, D. Anastasiadi<sup>3</sup>, J. Moraleda-Prados<sup>2</sup>, F.S. Zanuzzo<sup>1</sup>, M.L. Rise<sup>1</sup> and A.K. Gamperl<sup>1</sup>, (1)Department of Ocean Sciences, Memorial University of Newfoundland, Canada, (2)Institut de Ciències del Mar, ICM-CSIC, Spain, (3)The New Zealand Institute for Plant and Food Research, New Zealand

The marine environment is predicted to become warmer and more hypoxic over this century, and these conditions may become a challenge for cultured Atlantic salmon by negatively affecting their growth, immunology and welfare. DNA methylation mediates phenotypically plastic responses in gene expression that can potentially facilitate acclimatization responses. Thus, we measured DNA methylation from salmon that were subjected to: *i*)

control conditions (normoxia, 12°C); *ii*) an incremental increase in temperature (12°C to 20°C, at 1°C per week) and then held at 20°C for 4 weeks; or *iii*) the former temperature regimen in combination with moderate hypoxia (~70% air saturation). DNA methylation levels were measured at CpG sites within a ~500 bp region (Promotor, 5'UTR, Exon, Intron) of six important liver biomarker genes (*cribp*, *jund*, *pkd3*, *prdx6*, *serpinh1*, and *ucp2*). Considering both experimental groups, we found 12 CpGs (out of 94 total) across the six genes that were differentially methylated when exposed to 20°C for 3 days, whereas only 6 CpGs from three genes (*jund*, *prdx6* and *ucp2*) were affected after 4 weeks at 20°C. At both time points, we uncovered distinct DNA methylation profiles for fish of each treatment group, suggesting that high temperature and moderate hypoxia were inducing different CpG methylation changes in the liver of salmon. Further, we report significant relationships between CpG methylation and the mRNA expression of these genes that are complex and dynamic. These changes in DNA methylation may be an important regulatory mechanism allowing Atlantic salmon to quickly respond to new environmental challenges associated with global warming.

## **W063: Aquaculture**

### **Applications of Genomics to Expedite Genetic Improvement in Aquaculture Species**

**Ross D. Houston**, The Roslin Institute and R(D)SVS, Edinburgh, United Kingdom

The potential to grow aquaculture production via informed use of selective breeding and genomic technologies is huge, due to the relatively recent domestication and high fecundity of most species. In the more advanced and high value aquaculture sectors such as Atlantic salmon, genomic selection is routinely applied to increase selection accuracy and therefore cumulative genetic gain. This has been enabled by the development of high density SNP arrays and genotyping by sequencing technologies. To translate these benefits to many other aquaculture sectors, lower cost solutions are required, such as combined-purpose low density SNP panels for concurrent parentage assignment, imputation-based genomic selection, and potentially also marker-assisted selection for major QTL.

In parallel, high quality annotated reference genomes and functional genomic assays to profile transcriptional regulation can be utilised to prioritise putative causative variants in genomic regions associated with traits of economic interest. Such variants can potentially be harnessed to improve the accuracy of genomic prediction, and persistency of that accuracy in more distant relatives. Genome editing (e.g. CRISPR/Cas9) can be used to demonstrate the causality of these variants, and also has potential for the 'introgression' of favourable alleles from other strains or species, or the informed generation of *de novo* alleles, including via the application of genome-wide CRISPR screens.

This presentation will describe some examples of applied genomic and genome editing research aiming to take steps towards improvements in aquaculture breeding and production, with a focus on disease resistance.

## **W064: Aquaculture**

### **Genome of the Soft-Shell Clam and its Transmissible Cancer**

**Samuel FM Hart**, University of Washington, SEATTLE, WA

Cancer is normally an evolutionary dead-end—neoplastic cells that arise and evolve within an organism either regress or kill their host, and the death of the host marks the death of the cancer lineage. However, in some cases, neoplastic cells develop the ability to spread from individual to individual, turning from conventional cancers into clonal contagious cancer lineages. The natural transmission of cancer cells has been observed in two mammals (Tasmanian devils and dogs), and we have found that leukemia-like diseases in at least five bivalve species are due to the horizontal spread of clonal cancer lineages. One lineage affects soft-shell clams (*Mya arenaria*) along the east coast of North America and is ultimately fatal in most clams, contributing to the depletion of this commercially harvested species in many areas. We are currently assembling a reference genome for the soft-shell clam using PacBio sequencing combined with HiC data. Using draft reference genomes, we are investigating genomic changes in the evolution of this unique cancer lineage, including SNPs, structural variation, and copy number variation. We have found a retrotransposon, *Steamer*, which is expressed and amplified in genomic DNA of the contagious cancer lineage, expanding from 2-10 copies per haploid genome in normal animals to >100 in neoplastic cells. These new

integration events and other genomic changes have likely played a role in oncogenesis and continued evolution of the cancer with its hosts.

## **W065: Aquaculture**

### **Identification of Additional Single Nucleotide Polymorphisms associated with Resistance to Bacterial Cold Water Disease in Rainbow Trout using Whole Genome Resequencing**

Sixin LIU<sup>1</sup>, Guangtu Gao<sup>2</sup>, Roseanna Long<sup>2</sup>, Jason P. Evenhuis<sup>2</sup>, Gregory D. Wiens<sup>2</sup>, Kyle E. Martin<sup>3</sup> and Yniv Palti<sup>2</sup>, (1)USDA-ARS-NCCCWA, KEARNEYSVILLE, WV, (2)USDA-ARS-NCCCWA, Kearneysville, WV, (3)Troutlodge, Inc., Sumner, WA

Bacterial cold water disease (BCWD), caused by *Flavobacterium psychrophilum*, is a major disease in rainbow trout (*Oncorhynchus mykiss*). Previously, we have reported two major QTL associated with BCWD resistance on chromosomes Omy8 and Omy25. The objectives of this study were to identify additional single nucleotide polymorphisms (SNPs) associated with resistance to BCWD using whole genome resequencing, and to identify candidate genes for BCWD resistance. We conducted two rounds of pool-seq analysis in the Troutlodge odd-year May spawning population. For the first round of pool-seq, we pooled the DNA of parents based on their QTL haplotypes and the BCWD survival phenotypes of their offspring. In the second round of pool-seq, we pooled parental DNA samples solely based on BCWD phenotypes to avoid bias due to haplotype pre-selection. Over 10 million SNPs were identified in each round of pool-seq. Based on the first round of pool-seq, new SNPs showing significantly different allele frequencies between the two pools were used to genotype the 2015 Troutlodge May spawning population, and 26 SNPs associated with the BCWD resistance were validated. Candidate genes for the Omy08 QTL have also been identified after examining the functional annotation of the validated SNPs. Based on the second round of pool-seq, additional SNPs potentially associated with the two BCWD QTL have been identified, and we are currently evaluating those SNPs using the 2015 Troutlodge May spawning population.

## **W066: Aquaculture**

### **National Center for Biotechnology Information's (NCBI) Genome Annotation Resources for Aquaculture Species**

Nuala A. O'Leary, NCBI/NLM/NIH, Bethesda, MD

NCBI's Eukaryotic genome annotation pipeline ([ncbi.nlm.nih.gov/genome/annotation\\_euk/](http://ncbi.nlm.nih.gov/genome/annotation_euk/)) incorporates genomic, transcript, and protein sequence records, including RNA-seq data available in SRA, to provide comprehensive annotations of public genome assemblies submitted to NCBI's Assembly resource ([ncbi.nlm.nih.gov/assembly](http://ncbi.nlm.nih.gov/assembly)). To date, this pipeline has been used to annotate more than 570 eukaryotic genomes across diverse array of taxa that includes 79 fish species. Among these annotated genomes are numerous economically important aquaculture species such as carp, salmon, tilapia, shrimp and oyster. The annotations provided by this pipeline are available in various NCBI resources, including Reference Sequence (RefSeq) sequence databases, Gene, BLAST databases, FTP and in NCBI's Genome Data Viewer. All genome annotations produced by this pipeline are in scope for manual curation by the RefSeq curation group. Curators correct sequence or feature annotation errors that are identified by quality assurance tests, generate additional splice variants, and add feature annotation and data attributes. This presentation will describe some of the computational and manual curation procedures used in NCBI's genome annotation process and provide guidance on how the resources can be accessed and utilized by the aquaculture research community including a search functionality to find homologous gene sets.

## **W067: Aquaculture**

### **Genomics Technologies in Genomic Selection, and Case Studies**

Kahlil Lawless, Illumina, na, QC, Canada

Genomic selection has had a substantial impact on productivity and profitability over the last 10 years in species where it has been adopted by breeders. This evolution in breeding methodology was made possible by both scientific and technological advances. We will review success stories in plants and animals such as cattle and wheat,

provide a technology update from Illumina on genotyping methodologies including microarrays and next generation sequencing, and explore what implementation of Genomic Selection may mean for aquaculture breeding.

### **W068: Aquaculture**

#### **Evidence for Locally Adapted Phenotypes of Steelhead in the Columbia River**

**Shawn Narum**, Columbia River Inter-Tribal Fish Commission, Hagerman, ID

Anadromous species such as steelhead trout (*Oncorhynchus mykiss*) undergo long-distance migrations across geographical regions that consist of highly heterogeneous habitats. This may lead to local adaptation and signatures of adaptive variation associated with phenotypic traits that may be distinct across broad aquatic landscapes. Extensive genomics research in steelhead has revealed strong signals of local adaptation in steelhead throughout the Columbia River. Genome scans have identified that balancing selection maintains variation for phenotypic traits such as arrival timing to spawning grounds (premature vs. mature) and age-at-maturity (1-ocean vs. 2-ocean) in steelhead throughout the Columbia River. Genes of major effect have been identified for these two traits and development of markers from these candidate genes enable monitoring of phenotypic and genetic variation in natural populations or hatchery-reared stocks. This is a promising approach to maintain a broad portfolio of phenotypic diversity in steelhead that can buffer against exploitation and increase species persistence in disturbed ecosystems.

### **W069: Aquaculture**

#### **QTL and Joint-Association Mapping Reveal Loci Associated with Acute Low Salinity Tolerance in the Eastern Oyster (*Crassostrea virginica*)**

**Alexandra J McCarty**, University of Maryland Center for Environmental Science, Cambridge, MD, Standish K Allen, Virginia Institute of Marine Science and Louis Plough, Univ. Maryland Center for Environmental Science, Cambridge, MD

Oyster aquaculture is increasing globally, but varying environmental parameters pose challenges to farmers. In the Maryland-portion of the Chesapeake Bay, and in other estuarine and coastal systems, low salinity regions and large freshwater inflow events leave oysters exposed to salinities < 3 ppt for periods of time, which decreases oyster productivity and results in subsequent mass mortalities. While low salinity regions may hinder growth, they are enticing to farmers because they provide a refuge from common oyster diseases. A breeding program for extreme low salinity (< 3) waters would allow aquaculture to expand further into these reduced-disease areas, but the genetic framework underlying this trait is currently unknown. To determine the heritability of acute low salinity tolerance, we exposed fifty half-sibling families to an acute low salinity exposure (< 3) and determined that this trait, mortality at low salinity, is indeed heritable ( $h^2 = 0.4$ ). To determine genomic regions underlying this tolerance, we exposed four recombinant oyster families and individuals from a wild population to a similar acute low salinity exposure and generated genome-wide SNP data via ddRADseq techniques, genotyping all individuals (both alive and dead). We created QTL maps for each family and the wild population, as well as a joint-association map for both day-to-death and survival (live versus dead). Results indicate an important region on chromosome 1 related to survival at acute low salinity. Moderate heritability estimates and the identification of a highly significant QTL provide good support for the potential success of a low salinity breeding program.

### **W070: Aquaculture**

#### **Genomic Selection in Aquaculture: From Wild Stock to Advanced Breeding Schemes in *Penaeus monodon***

**Herman Raadsma**, University of Sydney, Camden, NSW, Australia and M.S. Khatkar<sup>1,6</sup>, J.L. Guppy<sup>2,6</sup>, N.M. Wade<sup>3,6</sup>, R.N. Huerlimann<sup>2,6</sup>, T.H. Noble<sup>3,6</sup>, A. Foote<sup>3,6</sup>, M. Hassan<sup>1,6</sup>, N. Khalilisamani<sup>1,6</sup>, D. Donovan<sup>4,6</sup>, K. Siemering<sup>5,6</sup>, M. J. Sellars<sup>3,6</sup>, J.A. Cowley<sup>3,6</sup>, G.J. Coman<sup>3,6</sup>, K.R. Zenger<sup>2,6</sup>, and D.R. Jerry<sup>2,6</sup>

Technological advances have made the use of large scale genomic and phenomic information a reality for many species. Here we discuss a large body of research, the technical advances, practical requirements and commercial

applications that have made genomic selection feasible in the black tiger shrimp, *Penaeus monodon*, under commercial conditions. The use of low-cost genome sequencing has enabled cost-effective genotyping on a large-scale and is of particular value for species without a reference genome or access to commercial genotyping arrays. We present the pitfalls and offer solutions to genotyping by sequencing approach, the building of appropriate genetic resources including a draft sequence of a reference genome, and to undertake genomic selection from first-hand experience. The potential to capture large-scale commercial phenotypes based on image analysis and artificial intelligence through machine learning, as inputs for the calculation of genomic breeding values is discussed. We highlight the advantages of the application of genomic selection over traditional aquatic breeding programs through being able to accurately predict complex polygenic traits including disease resistance; increasing rates of genetic gain; minimizing inbreeding; and negating potential limiting effects of Genotype by Environment interactions. Further practical advantages of genomic selection, through the use of large-scale communal mating and rearing systems are highlighted as well as presenting rate-limiting steps which impact on attaining maximum benefits from adopting genomic selection. Finally we introduce complex bio-economic modelling of *P. monodon* under intensive culture to direct breeding programmes to incorporate the appropriate breeding objectives for sustained profitable farming of this species.

### **W071: Aquaculture**

#### **Stress Regulation and Tolerance in Shrimp: The Transcriptomic Response to Ammonia Exposure in the Black Tiger Shrimp, *Penaeus monodon***

**Sarah Berry**, James Cook University, Bribie Island, QLD, Australia

Ammonia is extremely toxic in aquaculture and can accumulate rapidly in high-intensity shrimp production systems as a byproduct of metabolism and the breakdown of uneaten feed. Increased ammonia can impair growth, lower immunocompetence and disrupt numerous physiological functions including osmoregulation. This study exposed black tiger shrimp (*P. monodon*) to elevated ammonia for 72 hours and subsequent transcriptomic analysis of gill tissue identified 538 genes differentially expressed under ammonia stress, 86 upregulated and 452 downregulated. Upregulated genes were related to detoxification and elimination of xenobiotics, as well as the repair of cellular damage. Downregulated genes were associated with reduced feeding behaviors and functions of the shrimp innate immune system. Sequences obtained through RNASeq were then subjected to Regulatory Impact Factor (RIF) analysis to determine genes of high regulatory importance, which then together with the differentially expressed genes were used to construct a gene co-expression network using a Partial Correlation and Information Theory (PCIT) algorithm. These analyses identified several key potential “master regulator” genes associated with ammonia stress tolerance and performance in shrimp. These results may be applicable to selective breeding programs to enable a faster and more cost-effective identification of ammonia tolerant shrimp for production.

### **W072: Aquaculture**

#### **Impacts of Functional Feed Ingredients on Mucosal Immunity and Microbiota of Atlantic Salmon and Potential Implications for Sea Lice Resistance**

**Jacob W Bledsoe**, University of Idaho, Hagerman, ID

Aquaculture production of Atlantic salmon is worth US\$10.5 billion globally and one of the greatest challenges is losses to ectoparasitic infections from sea lice *Lepeophtheirus salmonis*. Historically, lice infections were managed by chemical therapeutants, though recently lice have shown resistance to most approved drugs. There is now interest in alternative treatments such as functional feeds, which are thought to alter the host's inflammatory response to ectoparasites or interfere with chemical signaling required for lice virulence. A twelve-week feeding trial was conducted at the National Cold Water Marine Aquaculture Center (USDA-ARS) with post-smolt Atlantic salmon (358 g ± 17) to evaluate effects of functional feeds on growth performance, microbiota, mucosal immunity, and lice resistance. Treatments included: (1) Control (Ctrl), (2) Ctrl + Coconut oil (98% lipid replacement) (CO), (3) Ctrl + 0.4% mannan-oligosaccharides (MOS), and (4) Ctrl + CO + MOS (CO-MOS). Fish growth was measured at six and twelve weeks, and all diets showed acceptable growth with no significant differences by treatment. Gut, gill, and skin microbiota were sampled from three fish tank<sup>-1</sup> (N = 18 diet<sup>-1</sup>) and characterized by 16S rRNA gene sequencing. Tissue samples and peripheral blood leukocytes were also collected for gene-expression analysis of immune biomarkers. Fish (N = 72 diet<sup>-1</sup>) were then subjected to common-garden challenges (4-hour static bath; 100

*L. salmonis* fish<sup>-1</sup>) followed by 14 days of observation, with no significant impact of diet on lice density. Microbiota and host gene-expression results across the three body sites and four dietary treatments will be presented.

### **W073: Aquaculture**

#### **Genomic Selection: From SNP Chips to Whole-Genome Sequence Data**

**Daniela Lourenco**<sup>1</sup>, Shogo Tsuruta<sup>1</sup>, Brian Bosworth<sup>2</sup>, Geoff Waldbieser<sup>2</sup>, Yniv Palti<sup>3</sup> and Ignacy Misztal<sup>1</sup>, (1)University of Georgia, Athens, GA, (2)USDA-ARS Warmwater Aquaculture Research Unit, (3)USDA-ARS-NCCCWA, Kearneysville, WV

Genomic selection has become the new standard in animal breeding and genetics over the past decade since its first implementation. The dairy industry was the first to benefit from this technology, and the aquaculture, poultry, and other livestock industries have since trailed the same path. Genomic information has been shown to enhance the rate of genetic gains by increasing the accuracy of selection and decreasing the generation interval. The current research in breeding and genetics has been focusing on finding new methods to make a better use of the genomic information and exploring new sources of data. One method that has been extensively used is the single-step genomic BLUP (ssGBLUP), which combines phenotypes, pedigree, and genomic information in one single analysis. This method has been implemented for commercial genetic evaluation of several livestock species, as well as some aquaculture species. With the recent availability of sequence data for several species, a common issue is whether this type of information can add benefits beyond what medium density SNP chips are providing. I will review our recent application of genomic selection in catfish and rainbow trout, and provide insights on the implementation of this technology in other industries. Finally, perspectives on the use of sequence data for the identification of causative variants and subsequent application for genomic predictions will be summarized to highlight when it can be beneficial for genetic improvement purposes.

### **W074: Aquaculture**

#### **Genomic Data Mining Tools for Domesticated Animal Species**

**Christine G. Elsik**<sup>1</sup>, Justin J. Le Tourneau<sup>1</sup>, Md Shamimuzzaman<sup>1</sup>, Deborah A. Triant<sup>1</sup>, Amy T. Walsh<sup>1</sup> and Darren Hagen<sup>2</sup>, (1)University of Missouri, Columbia, MO, (2)Oklahoma State University, Stillwater, OK

We have been developing a genomic data mining warehouse called FAANGMine (<http://faangmine.org>) and will present it as an example of a resource that can be adapted by the Aquaculture research community. FAANGMine is a data mining warehouse for species of interest to the Functional Annotation of Animal Genomes (FAANG) Consortium. Based on the InterMine data warehousing platform, FAANGMine integrates data from a variety of sources, such as reference genome assemblies, genes, proteins, protein families and domains, orthologs, pathways, gene ontology, QTL, variation, and publications. An assortment of search tools enables researchers without scripting skills to create and export customized annotation datasets merged with their own research data for use in downstream analyses. Built-in query templates provide starting points for data exploration, while the QueryBuilder tool supports construction of complex queries. The List Analysis and Genomic Regions search tools execute queries based on uploaded lists of identifiers and genome coordinates, respectively. Data can be exported in a variety of formats, including gff, fasta, json and tab-delimited files. Used in combination, the FAANGMine data mining tools provide robust support for meta-analysis and the integration of large datasets.

### **W075: Aquaculture**

#### **DNA Methylation and Transcriptomic Changes Involved in Atlantic Salmon Sexual Maturation**

**James W. Kijas**, CSIRO Animal, Health and Food Science, St Lucia, Australia

Atlantic salmon farming promotes growth in conditions which mean animals may complete sexual development at weights below harvest size. This can lead to a reduction in productivity and prompted us to investigate the biological mechanisms that control the timing of sexual maturation. We performed a time course experiment, whereby animals were manipulated with photoperiod before tissues were collected across the time window when animals commence sexual development. We performed whole genome bisulfite sequencing of three salmon tissues (pituitary, ovary and

liver) at both the beginning and end of the experiment, to take a first look at the patterns of DNA methylation and examine how they change in response to the onset of an important life history trait. Comparison across timepoints revealed 6,373 differentially methylated regions (DMRs), of which approximately 50% were located within genes (DMGs). The ovary underwent the most profound remodelling, with a strong bias towards increased methylation levels (hyper-methylation). at the final timepoint We also performed deep transcriptomic profiling (RNA-seq) of the same tissues to explore the relationship between methylation changes and gene expression. Weak correlation was observed considering all available genes, suggesting methylation may not be the key epigenomic regulator of global expression in the context of our experiment. However, we found a significant overlap between DMGs and differentially expressed genes in the ovary. Taken together, our results suggest chromatin remodelling genes play a role in the commitment of animals to the sexual maturation pathway. They also open the way for the identification of functional variants that can be used in advanced breeding approaches to boost productivity in Atlantic salmon farming.

## **W076: Aquaculture**

### **Influence of Ocean Acidification on DNA Methylation Patterns in Geoduck**

**Shelly A. Trigg**, University of Washington, Seattle, WA

To investigate acclimatization to ocean acidification through epigenetic mechanisms we examined genome-wide DNA methylation in early-stage juvenile geoduck clams that underwent a series of exposure experiments. Compared to ambient conditions, juvenile geoduck displayed decreased shell size in two low pH conditions (pH ~7.4 and pH ~7.0). When returned to ambient conditions for several months, the initial exposure to low pH resulted in compensatory growth with juveniles from the two initial low pH treatments being larger compared to those from the initial ambient pH condition. Interestingly, juvenile clams retained treatment specific regions of differential methylation after 125 days being removed from the initial low pH treatment. This is one of the first indications that DNA methylation could serve as a persistent epigenetic mark with a potential to influence subsequent response to environmental stress. Differentially methylated regions within specific genes and transposable elements shed light on putative functional roles of DNA methylation and mechanisms of intragenerational acclimatization. Taken together these data suggests that acclimatization to ocean acidification can result in benefits to geoduck growth, with exposure memory providing a mechanism for environmental hardening.

## **W077: Arabidopsis Informatics**

### **Plant Regulomics: A Data-Driven Interface for Retrieving Upstream Regulators from Plant Multi-Omics Data.**

**Yi-Jing Zhang**, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

High-throughput technology has become a powerful approach for routine plant research. Interpreting the biological significance of high-throughput data has largely focused on the functional characterization of a large gene list or genomic loci that involves the following two aspects: the functions of the genes or loci and how they are regulated as a whole, i.e. searching for the upstream regulators. Traditional platforms for functional annotation largely help resolving the first issue. Addressing the second issue is essential for a global understanding of the regulatory mechanism, but is more challenging, and requires additional high-throughput experimental evidence and a unified statistical framework for data-mining. The rapid accumulation of 'omics data provides a large amount of experimental data.

We here present Plant Regulomics, an interface that integrates 19 925 transcriptomic and epigenomic data sets and diverse sources of functional evidence (58 112 terms and 695 414 protein-protein interactions) from six plant species along with the orthologous genes from 56 whole-genome sequenced plant species. All pair-wise transcriptomic comparisons with biological significance within the same study were performed, and all epigenomic data were processed to genomic loci targeted by various factors. These data were well organized to gene modules and loci lists, which were further implemented into the same statistical framework. For any input gene list or genomic loci, Plant Regulomics retrieves the upstream factors, treatments, and experimental/environmental

conditions regulating the input from the integrated 'omics data. Additionally, multiple tools and an interactive visualization are available through a user-friendly web interface.

Plant Regulomics is available at <http://bioinfo.sibs.ac.cn/plant-regulomics>.

### **W078: Arabidopsis Informatics**

#### **Chromosome-Level Assemblies of Multiple Arabidopsis Genomes Reveal Hotspots of Rearrangements with Altered Evolutionary Dynamics**

**Wen-Biao Jiao** and Korbinian Schneeberger, Max Planck Institute for Plant Breeding Research, Cologne, Germany

We report chromosome-level, reference-quality assemblies of seven *Arabidopsis thaliana* accessions selected across the global range of this predominately inbred plant. Each genome revealed between 13-17 Mb rearranged and 5-6 Mb novel sequence introducing copy-number changes in ~5,000 genes, including ~1,900 genes which are not part of the current reference annotation. Analyzing the collinearity between the genomes revealed ~350 regions (4.1% of the euchromatin) where accession-specific tandem duplications destroyed the syntenic gene order between the genomes. These hotspots of rearrangements were characterized by the loss of meiotic recombination in hybrids within these regions and the enrichment of genes implicated in biotic stress response. Together this suggests that hotspots of rearrangements are governed by altered evolutionary dynamics as compared to the rest of the genome, which are based on new mutations and not on the recombination of existing variation, and thereby enable a quick response to the ever-evolving challenges of biotic stress.

### **W079: Arabidopsis Informatics**

#### **Exploring Arabidopsis and Brassicaceae Genomes with the Genome Context Viewer**

**Andrew Farmer**, National Center for Genome Resources (NCGR), Santa Fe, NM

The Genome Context Viewer (GCV) is a web-based multi-genome browser that dynamically associates and aligns genomic regions on the basis of their shared functional content. The visualizations and interactive features of GCV facilitate within and cross-genomic comparisons with respect to high level conserved and variable content such as presence/absence and copy number variations of genic features, inversions, translocations and other structural variation at both micro- and macrosyntenic levels. An instance of GCV has been established at <https://gcv-arabidopsis.ncgr.org> to incorporate annotated genomes from the Arabidopsis research community as well as other closely related Brassicaceae species. We will provide examples of how GCV can be used to facilitate traversal among genomes and to federate resources providing more in depth information about those genomes.

### **W080: Arabidopsis Informatics**

#### **Identification of Novel Translated ORFs using Enhanced Super-Resolution Ribosome Profiling in Arabidopsis**

**Larry Wu** and Polly Hsu, Michigan State University, East Lansing, MI

Most open reading frames (ORFs) in Arabidopsis are predicted by computational annotation pipelines, which designate mRNAs with ORFs greater than 100 amino acids as coding and assume the longest ORF in a given mRNA to be the main ORF. More recently, ribosome profiling has been introduced as an experimental approach to identify novel small translated ORFs. Here, we report an enhanced ribosome profiling method that provides high coverage in addition to strong 3-nt periodicity. With this improved method, we uncovered many small translated ORFs in different categories: uORFs (upstream ORFs), dORFs (downstream ORFs), oORFs (overlapping ORFs), miORFs (i.e., ORFs encoded by microRNA precursors) and sORFs (i.e., small ORFs encoded by presumed non-coding RNAs). uORF-containing genes are enriched for signal transduction and transcription factors, while dORF-containing genes are enriched for mRNA translation. The existence of translated oORFs suggests that alternative reading frames in a main coding sequence can encode more than one peptide in plants. The conservation of oORF/main ORF pairs between Arabidopsis and tomato implies the biological importance of both types of peptides. Additionally, many pseudogenes are translated, and some protein-coding genes are annotated as pseudogenes due to

errors in the TAIR9 genome sequence. In conclusion, an improved ribosome profiling method with strong 3-nt periodicity and high coverage provides a high-definition map of the plant translatoome. Our new data enable the discovery of novel peptide-encoding ORFs and facilitate the improvement of genome annotation in Arabidopsis and other plants.

### **W081: Arabidopsis Informatics**

#### **The Bio-Analytic Resource for Plant Biology – Thalemine and 2019 Updates**

**Sylva Donaldson**, University of Toronto / CAGEF, Toronto, ON, Canada

Since November 2019, the BAR is hosting the Arabidopsis instance of InterMine, called ThaleMine, formerly at Araport.org. This is a searchable online resource for integrating and analyzing Arabidopsis data. As part of the roll-out on the BAR, the Thalemine framework and content was completely updated. In addition to this useful addition to the BAR, new eFP Browsers are now available, including a single cell root RNA-seq data set. Work is currently being done on a gene regulatory net explorer, and updating the Arabidopsis Interactions Viewer. Last, work is progressing on an “intelligent assistant” for plant biology.

### **W082: Arabidopsis Informatics**

#### **Genome-Wide Identification of Splicing Quantitative Trait Loci (sQTLs) in Diverse Ecotypes of *Arabidopsis thaliana***

**Muhammad Waqas Khokhar**, Canterbury Christ Church University, CANTERBURY, United Kingdom

Alternative splicing (AS) of pre-mRNAs contributes to transcriptome diversity and enables plants to generate different protein isoforms from a single gene and/or fine-tune gene expression during different development stages and environmental changes. Although AS is pervasive, the genetic basis for differential isoform usage in plants is still emerging. In this study, we performed genome-wide analysis in 666 geographically distributed diverse ecotypes of *Arabidopsis thaliana* to identify genomic regions [splicing quantitative trait loci (sQTLs)] that may regulate differential AS. These ecotypes belong to different microclimatic conditions and are part of the relict and non-relict populations. Although sQTLs were spread across the genome, we observed enrichment for *trans*-sQTL (*trans*-sQTLs hotspots) on chromosome one. Furthermore, we identified several sQTL (911) that co-localized with trait-linked single nucleotide polymorphisms (SNP) identified in the Arabidopsis genome-wide association studies (AraGWAS). Many sQTLs were enriched among circadian clock, flowering, and stress-responsive genes, suggesting a role for differential isoform usage in regulating these important processes in diverse ecotypes of Arabidopsis. In conclusion, the current study provides a deep insight into SNPs affecting isoform ratios/genes and facilitates a better mechanistic understanding of trait-associated SNPs in GWAS studies. To the best of our knowledge, this is the first report of sQTL analysis in a large set of Arabidopsis ecotypes and can be used as a reference to perform sQTL analysis in the Brassicaceae family. Since whole genome and transcriptome datasets are available for these diverse ecotypes, it could serve as a powerful resource for the biological interpretation of trait-associated loci, splice isoform ratios, and their phenotypic consequences to help produce more resilient and high yield crop varieties.

### **W083: Arabidopsis Informatics**

#### **Improvements and New Features at the Arabidopsis Information Resource (TAIR)**

**Sabarinath Subramaniam**, Phoenix Bioinformatics, Fremont, CA

The Arabidopsis Information Resource (TAIR) is a genetic and molecular database focused on the model plant *Arabidopsis thaliana*. TAIR was established in 1999 and continues to provide the most comprehensive and current set of plant gene function data to the research community. This presentation will discuss the latest features and functionalities at TAIR, including upgraded sequence analysis tools, gene function information added from the published literature, and community resource pages. Recently, TAIR deployed the latest version of the JBrowse genome browser to replace the JBrowse viewer that was provided by Araport. The TAIR JBrowse instance not only reproduces the tracks available at Araport but also integrates new community submitted tracks based on recently released public data sets. This talk will feature a demo of the TAIR JBrowse viewer functionalities.

## **W084: Arthropod Genomics and Genome Engineering**

### **Overview of Genomic Tools to Support Boll Weevil Eradication Programs**

**Lindsey Perkin**<sup>1</sup>, Charles P.-C. Suh<sup>2</sup>, Tyler Jay Raszick<sup>3</sup>, Xanthe A. Shirley<sup>4</sup>, Greg A. Sword<sup>3</sup> and Raul Ruiz-Arce<sup>5</sup>, (1)Agricultural Research Service, USDA, College Station, TX, (2)USDA-ARS Southern Plains Agricultural Research Center, College Station, TX, (3)Texas A&M University, College Station, TX, (4)USDA APHIS PPQ, College Station, TX, (5)USDA APHIS Science & Technology, Edinburg, TX

The boll weevil, *Anthonomus grandis grandis* Boheman, has been eradicated from >95% of the cotton acreage in the U.S., however, active and post-eradication programs continue to operate and rely on pheromone traps to detect weevil populations. Herein, we highlight application of four molecular tools that are currently being used or developed to support boll weevil eradication programs. Initially, a pipeline has been established to determine the geographical origin of boll weevils captured in re-infested areas. Secondly, a diagnostic assay is under development to distinguish boll weevils from *A. g. thurberi* and other weevil species commonly captured in traps. Accurate identification of weevil species and respective geographical origin is important to avoid unnecessary remedial action. An additional technique is underway to estimate the relatedness of boll weevils. This information will determine if re-infestations resulted from mass dispersal of weevils or if a small number were carried anthropogenically into the area followed by local reproduction. Lastly, RNA-seq efforts are ongoing to identify gene targets involved in pheromone production and diapause which may be developed into gene disruption technologies. Implementation and success of these tools will be improved with access to a high-quality genome assembly. In partnership with the Ag100Pest Initiative, sequencing of a single boll weevil will be assembled and annotated. It is anticipated that these molecular tools, supported by a genome reference, will provide eradication programs with information that can shape new policy, regulations and lead to novel pest management solutions.

## **W085: Arthropod Genomics and Genome Engineering**

### **Genomic Insights into the Seasonal Biology of the European Corn Borer Moth**

**Erik B. Dopman**, Tufts University, Department of Biology, Medford, MA, Brad S. Coates, USDA-ARS Corn Insects & Crop Genetics, Ames, IA and Genevieve M. Kozak, University of Massachusetts-Dartmouth

Increases in the annual number of generations can lead to faster evolution, population growth, and reduced extinction risk during rapid climate change. For insects, both environmental and genetic factors influence the number of generations per year, but little is known about the molecular mechanisms or evolutionary causes. We focus on addressing these questions using the European corn borer moth (*Ostrinia nubilalis*), an herbivorous pest that is capable of producing four or more generations per year. In this talk, I will discuss how we use a diversity of ecological, behavioral, and genomic approaches to uncover the mechanisms underlying the evolution of generation number. Our results highlight a role for adaptive responses to seasonal climate through shifts in biological clocks.

## **W086: Arthropod Genomics and Genome Engineering**

### **Population Structure and Impact on Resistance to Transgenic Crops in Fall Armyworm (*Spodoptera frugiperda*) from Diverse Locations**

**Katrina Schlum**<sup>1</sup>, Juan Luis Jurat-Fuentes<sup>2</sup>, Kurt Lamour<sup>2</sup>, Caroline Placidi de Bortoli<sup>2</sup> and Scott J. Emrich<sup>3</sup>, (1)University of Tennessee, Department of Genome, Science & Technology, Knoxville, TN, (2)University of Tennessee, Department of Entomology and Plant Pathology, Knoxville, TN, (3)University of Tennessee, Department of Electrical Engineering and Computer Science, Knoxville, TN

The fall armyworm (*Spodoptera frugiperda*, J.E. Smith) is a highly polyphagous agricultural pest with long-distance migratory behavior threatening food security worldwide. This pest has a host range of >80 plant species but preferentially feeds on corn and has developed resistance to multiple pesticides. Specifically, field populations of *S. frugiperda* in North and South America have evolved practical resistance to transgenic corn producing the Cry1Fa insecticidal protein from the bacterium *Bacillus thuringiensis* (*Bt*). The mechanism and allele linked to resistance against Cry1Fa in Puerto Rico was identified as a 2-bp insertion in an ATP binding cassette subfamily C2 (*ABCC2*) gene in *S. frugiperda* (*SfABCC2*). The goal of this project was to survey fall armyworm genomes from the USA,

Brazil, Argentina, Kenya and Puerto Rico to determine population structure and how it may affect resistance evolution. Based on whole genome sequencing of 51 *S. frugiperda* samples, our analysis suggests there is no clear population structure based on BUSCO gene or k-mer based trees and principal component analysis. In contrast, two subpopulations were found based on a mitochondrial gene-based tree, supporting the continued use of mitochondrial gene markers (*COI* and *Tpi*) for surveying *S. frugiperda* populations. Our research is the largest diverse collection of *S. frugiperda* whole genome sequences characterized to date and provides a foundational resource for surveying *S. frugiperda* populations and resistance candidate alleles.

### **W087: Arthropod Genomics and Genome Engineering**

#### **Disruption of Genes Important for Flight, Feeding and Femaleness in the Disease Vector Mosquito, *Aedes aegypti***

**Zach Adelman**, Texas A&M University, College Station, TX

The development of cost-effective and simple reagents for programmable gene editing such as CRISPR Cas9 brings with it unprecedented opportunity to alter the genomes of organisms important to both public health and agriculture. Simultaneously, the realization of effective gene drive using CRISPR/Cas9 has created a technically viable pathway for the permanent modification of wild populations. Such beneficial modifications may include pathogen resistance, alterations in feeding behavior, restoration of insecticide susceptibility, shortening of lifespan, and gender bias. Using the arboviral vector *Aedes aegypti*, we are exploring genetic strategies to alter sex ratios, ablate female flight, and induce fatal indigestion upon bloodfeeding. Key to each of these approaches is the identification and characterization of critical genes underlying each of these important phenotypes.

### **W088: Arthropod Genomics and Genome Engineering**

#### **CRISPR/Cas9 Gene Editing Confirms the Function of a *Bacillus thuringiensis* Resistance Receptor Gene**

**Jeff Fabrick**<sup>1</sup>, Xianchun Li<sup>2</sup>, Yves Carriere<sup>2</sup> and Bruce Tabashnik<sup>2</sup>, (1)USDA-ARS, Maricopa, AZ, (2)University of Arizona

Crops genetically engineered to produce insecticidal proteins from the bacterium *Bacillus thuringiensis* (Bt) kill some of the world's major insect pests while simultaneously reducing the use of insecticide sprays, enhancing biological control, and increasing farmer profits. However, the evolution of pest resistance threatens the benefits of transgenic Bt crops. Transgenic cotton producing two Bt proteins (Cry1Ac and Cry2Ab) continues to be used against several key pests and has played a critical role in the eradication of the pink bollworm, *Pectinophora gossypiella*, in the U.S. However, the pink bollworm remains problematic in many other cotton growing areas of the world, including India where field-evolved practical resistance to Bt cotton is widespread. We previously identified mutations that disrupt the same ATP-binding cassette transporter gene (*PgABCA2*) in both a laboratory-selected strain from Arizona and in field-selected populations from India. Although our previous results showed genetic linkage between *PgABCA2* and resistance to Cry2Ab, an actual causative role in resistance for such mutations has not been established. Here, we used CRISPR/Cas9 gene editing to create gene knockouts in the *PgABCA2* gene in a Bt-susceptible pink bollworm strain and show that such mutations are sufficient to cause resistance to Cry2Ab. Hence, gene editing can provide practical means to accelerate efforts in discovery and attributing function to mutations in genes involved in pesticide resistance.

### **W089: Avian Genomics - Going Wild!**

#### **Avian Genomics in Animal Breeding: Do we still need model organisms?**

**Alain Vignal**, GenPhySE, INRAE, ENVT, ENSAT, 31326, Castanet Tolosan, France

Chicken is by far the most widely used bird in animal breeding and is also a model organism that has long been used for scientific observations, typically in embryology. This position at the crossroads of the breeding industry and of basic research is part of the reasons for which chicken was amongst the first vertebrates sequenced at the dawn of the era of large genomes analyses, with a draft genome published only three years after the human one. The other reasons were that it was the only representative of the bird lineage at the time having available genomics resources

such as genetic maps or large-insert DNA libraries. Just like a lot of the understanding of bird biology stemmed from using chicken as model, new insights into the specificities of bird genomes, such as the karyotype organization into macrochromosomes and microchromosomes, could be investigated in great detail.

Since then, largely due to the advent of the second-generation parallel sequencing and third-generation long-read methods, more bird species were sequenced and at the time of writing, 163 assemblies are available with various levels of quality in the NCBI genome database. These include other birds bred as agricultural species, such as turkey, duck, quail or guinea fowl.

In animal breeding, the genomic information is mainly used for QTL detection, marker-assisted or genomic selection and for a deeper understanding of biological mechanisms. Also, much had been done in poultry, to take advantage of the large collection of phenotypic variants present either in commercial or in fancy breeds. For instance, a number of genes causing coloration and ornamental variation were identified first in chicken, thanks for all the genomics tools available, and later on in other species, usually and in the absence of a reference genome, by directly testing the candidate genes found in chicken.

With the advent of high quality and low-cost sequencing, the number of available genomes will expand rapidly and their quality will increase. However, does this mean model organisms will not be needed anymore?

## **W090: Avian Genomics - Going Wild!**

### **Avian Ecological Epigenetics: Pitfalls and Promises**

**Bernice Sepers** and Kees van Oers, The Netherlands Institute of Ecology (NIOO-KNAW) and Wageningen University & Research (WUR), Wageningen, Netherlands

Epigenetic mechanisms can alter gene expression without structural modifications of the nucleotide sequence and are increasingly recognized as important mechanisms that can generate phenotypic diversity. Most of our current knowledge regarding the origin and role of epigenetic variation comes from research on plants or mammals, often in controlled rearing conditions. Epigenetic research on birds in their natural habitats is still in its infancy, but is needed to answer questions regarding the origin of epigenetic marks and their role in phenotypic variation and evolution. Therefore, I will present a recent review in which my colleagues and I examined the potential for studying epigenetic variation in natural bird systems. During this presentation, I will provide insights into the origin of epigenetic variation, the relationship between epigenetic variation and trait variation, and the possible role of epigenetic variation in adaptation to changing environments. Furthermore, I will address the pitfalls and promises of avian ecological epigenetics. In addition, I will suggest some topics I believe need to be addressed to develop the field of wild avian epigenetics further, which will be mainly related to (early) environmental causes of variation in DNA methylation. Finally, I will explain how I tackle the pitfalls and explore the promises of avian ecological epigenetics during my PhD project at the NIOO-KNAW (the Netherlands). Therefore, I will elaborate on how I study (early) environmental effects on DNA methylation and exploratory behaviour in wild great tits (*Parus major*) and show some preliminary results.

## **W091: Avian Genomics - Going Wild!**

### **The Hidden Information of Unmapped Reads in a Wild Songbird**

**Veronika N. Laine**, The Netherlands Institute of Ecology (NIOO), Wageningen, Netherlands

A widely used approach in next-generation sequencing projects is the alignment of reads to a reference genome. Despite methodological and hardware improvements which have enhanced the efficiency and accuracy of alignments, a significant percentage of reads frequently remain unmapped. Usually, unmapped reads are discarded from the analysis process, but significant biological information and insights can be uncovered from these data especially in the case of “missing genes” in birds where total number of uniquely identified avian coding genes is considerably smaller than for other tetrapods. Here, we explored the unmapped DNA (normal and bisulfite treated) and RNA sequence reads of the great tit (*Parus major*) reference genome individual. We aligned the unmapped contigs to BLAST database where many of the aligned contigs showed sequence similarity to different bird species but also to genes that were absent in the great tit reference assembly. Furthermore, there were also contigs that

represented known *P. major* pathogenic species. Most interesting were several species of blood parasites such as *Plasmodium* and *Trypanosoma*. Our analyses revealed that meaningful biological information can be found when exploring unmapped reads. In this talk I will also propose strategies to aid the capture and interpretation of this information from unmapped reads and how to hunt down the missing avian genes.

## **W092: Avian Genomics - Going Wild!**

### **The Role of Genomics in Conserving Biodiversity during the Sixth Mass Extinction**

**Sadye Paez**, The Rockefeller University, New York, NY

Approximately 1 million plant and animal species are threatened with extinction based on estimated proportions of currently threatened species, with extinction rates for birds (as well as mammals, amphibians, and reptiles) as fast or faster than previous extinction events. Further, habitat loss in the 25 biodiversity ‘hotspots’ suggest that birds and mammals are most at risk in this current extinction event. This human induced so-called “sixth mass extinction” is detrimental to humans and the planet. The impact on biodiversity, defined as all forms of life on our planet, their ecosystems, and their associated genetic information, is expected to be catastrophic, and many of these losses are likely to trigger cascading secondary extinctions in trophic networks. Population-level adaptive genetic variation is key to enabling species to adapt to changing environments, to respond to the pressures of natural selection, and to minimize the accumulation of deleterious mutations. Genomics offers a powerful and unique tool to characterize this biodiversity and genetic variation at unprecedented levels of detail, both among and within species, and is a potential complement and enhancement to traditional biodiversity conservation measures. However, conservation practitioners have historically been able to sample only small fractions of the genome, directly or indirectly, and until recently only short DNA reads have been widely accessible, hampering whole-genome approaches for all but a few model species. Informing and reversing this current mass extinction will benefit from error-free, chromosome-level reference genome assemblies, such as those from the Vertebrate Genomes Project, which currently has 22 curated avian reference genome assemblies (that are part of the 69 total curated of 124 in progress error-free, chromosome-level reference genome assemblies currently publicly available here: <https://vgp.github.io>).

## **W093: Avian Genomics - Going Wild!**

### **Conservation Genomics of the Critically Endangered Red Siskin (*Spinus cucullata*)**

**Paul B. Frandsen**<sup>1</sup>, Haw Chuan Lim<sup>2</sup>, Rebecca Dikow<sup>3</sup>, Brian Coyle<sup>4</sup>, Kathryn Rodriguez-Clark<sup>4</sup>, Pavel Dobrynin<sup>4</sup>, Klaus-Peter Koepfli<sup>5</sup>, Warren E. Johnson<sup>6</sup> and Mike Braun<sup>4</sup>, (1)Brigham Young University, Provo, UT, (2)George Mason University, Fairfax, VA, (3)Smithsonian OCIO Data Science Lab, Washington, DC, DC, (4)Smithsonian Institution, Washington, DC, (5)Center for Species Survival, Smithsonian's National Zoological Park and Conservation Biology Institute, Washington, DC, (6)The Walter Reed Biosystematics Unit, Museum Support Center MRC-534, Smithsonian Institution, Suitland, MD

The red siskin (*Spinus cucullata*) is a critically endangered South American bird threatened by trapping and habitat loss. Threat reduction efforts are underway, but full recovery will likely require future reintroductions. Aviculturists have traditionally hybridized red siskins with the common canary (*Serinus canaria domestica*), and past backcrosses may have also introgressed canary DNA into captive birds that are phenotypically indistinguishable from wild red siskins. To address concerns of cryptic introgression, as well as potential inbreeding, we are creating genome-scale molecular tools to support a well-informed conservation-breeding and reintroduction program that aims to restore siskin populations across their native range. We sequenced and assembled a *de novo* reference genome using a hybrid assembly approach (MaSuRCA) with long (PacBio) and short (Illumina) reads. This resulted in a genome with relatively high contiguity (contig N50 = 699 kb) and gene completeness as assessed via BUSCO (94.5% complete). We annotated the new reference genome using a pipeline consisting of RepeatModeler, RepeatMasker, BLAT, and Augustus. We also performed additional low-coverage re-sequencing of nine individuals to better understand the genetic diversity, population structure, and evolutionary history of the red siskin, and to design markers for analysis on historical samples and individuals used in a nascent captive breeding program. Population genetics analysis revealed two distinct populations from Guyana and Venezuela. We carried out site-frequency spectrum analysis of the two populations to reveal population divergence history and used a sliding-window approach to reveal genetic diversity, linkage disequilibrium and signatures of selection across the genomes. These

resources will be used for *in situ* and *ex situ* breeding management and population monitoring and to verify genetic purity of any birds selected as founders for conservation breeding. Additionally, this work provides important opportunities for comparative genomics that may yield valuable insights for conservation of other endangered and highly bottlenecked species.

#### **W094: Avian Genomics - Going Wild!**

##### **Landscape Level Genetic Consequences of Game-Farm Mallard Releases in North America: Domestic Mallards Gone-Wild!**

**Philip Lavretsky**, University of Texas at El Paso, El Paso, TX

Along with changing habitat, the direct release of domesticated individuals into the wild is a practice used world-wide to augment wildlife populations. I test between possible outcomes of human-mediate secondary contact using genomic techniques and at both historical and contemporary time scales for North American mallards. First, I apply a bait-capture array targeting thousands of markers to century-old (1842-1915) and contemporary (2009-2010) mallards. Next, I sequence thousands of nuclear ddRAD-seq markers, along with a mitochondrial marker across contemporary wild mallards throughout their North American range, as well as known game-farm and other feral “park” mallards. Visualizing population structure and estimating individual ancestry, I conclude the existence of extensive introgression between game-farm released and a North American wild mallards, with rates increasing over time. Geographically, I determine the highest rates of hybridization and introgression in eastern North America (i.e., Atlantic Flyway), where historically, game-farm mallard stocking has been most intense. Together, historical and contemporary results confirm that the intensive stocking practices of game-farm mallards conducted for the last century has fundamentally changed the genetic integrity of North America’s wild mallard population. In fact, it becomes of great interest to ask whether the iconic North American mallard is declining in the wild due to introgression of maladaptive traits from these domesticated forms.

#### **W095: Banana Genomics**

##### **Global Epidemic of Panama Disease on Banana is caused by a New Fungal Species Originating from Southeast Asia**

Nadia Ordóñez<sup>1</sup>, Nani Maryani<sup>1</sup>, Fernando Garcia-Bastidas<sup>2</sup>, David Torres Sanchez<sup>1</sup>, Carolina Aaguilera Galvez<sup>1</sup>, Xiaoqian Shi-Kunne<sup>1</sup>, Giuliana Nakasato Tagami<sup>1</sup>, Pauline Robert<sup>1</sup>, Juliette Quintero Vargas<sup>3</sup>, Emilio Arévalo-Peñaranda<sup>3</sup>, Jorge Vargas<sup>4</sup>, Harold JG Meijer<sup>1</sup>, Rene Hofstede<sup>2</sup>, Alexander Wittenberg<sup>2</sup>, Anker Sørensen<sup>2</sup>, **Michael F Seidl**<sup>5</sup> and Gert HJ Kema<sup>1</sup>, (1)Wageningen University & Research, Wageningen, Netherlands, (2)KeyGene, Wageningen, Netherlands, (3)The Colombian Agricultural and Livestock Institute, Bogotá, Colombia, (4)Centro de Investigaciones del Banano, Medellín, Colombia, (5)Utrecht University, Utrecht, Netherlands

The spatiotemporal origins and factors impacting the dissemination of plant pathogenic fungi remain elusive for many epidemics. Bananas are the world’s most popular fruit and represent crucial food commodities. Global banana production is dominated by Cavendish monocultures, the remedy to manage the Panama disease epidemic in Central America in the last century. Here, we discuss the most recent results of our ongoing efforts to study the diversity and dissemination of the causal agents of Panama disease worldwide. We genotyped a global collection of ~550 *Fusarium* isolates and traced the origins of Panama disease to Southeast Asia, bananas’ center of origin and a biodiversity hotspot for banana-infecting *Fusarium* species. While the previous epidemic was caused by a suite of genetically diverse *Fusarium* species, we show that the current epidemic that devastates Cavendish and many other varieties is caused by a single clone of the genetically distinct new species *F. odoratissimum*, commonly referred to as tropical race 4 (TR4). We generated chromosome-level genome assemblies of TR4 and two related *Fusarium* strains using the Oxford Nanopore PromethION platform, which was complemented by whole-genome re-

sequencing efforts aiming to capture the worldwide diversity of banana-associated *Fusarium* isolates. Comparative genomics revealed rapidly evolving chromosomal regions that occur specifically in TR4, which will enable us to study the molecular mechanisms underlying TR4 aggressiveness in the future. The rapid spread of TR4 into banana-growing regions in Asia, Middle East, Africa, and Latin America seriously threatens worldwide banana production, which has significant impact on food and fruit production. Understanding the spatiotemporal origins and evolution of the ongoing Panama disease pandemic is critical for developing sustainable control measures and to prevent further international dissemination.

## **W096: Banana Genomics**

### **Advancements Toward Molecular Breeding and Genomics in the ‘Breeding Better Bananas’ Project in East Africa**

**Allan Brown**<sup>1</sup>, Brigitte Uwimana<sup>2</sup>, Moses Nyine<sup>3</sup>, Ivan Kabiita Arinaitwe<sup>4</sup>, Mohamed Hussein Mpina<sup>5</sup>, Marnik Vuylsteke<sup>6</sup>, Elizabeth Aitken<sup>7</sup>, Eva Hřibová<sup>8</sup>, Jaroslav Doležel<sup>8</sup>, Steven Tanksley<sup>9</sup> and Rony Swennen<sup>10</sup>, (1)International Institute of Tropical Agriculture, Arusha, Tanzania, United Republic of, (2)International Institute of Tropical Agriculture, Kampala, Uganda, (3)Kansas State University, (4)National Agricultural Research Organisation (NARO)-Uganda, Kampala, Uganda, (5)Tropical Pesticides Research Institute, ARUSHA, Tanzania, United Republic of, (6)Gnomixx, Melle, Belgium, (7)School of Agriculture and Food Sciences, The University of Queensland, St Lucia, Australia, (8)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (9)Nature Source Improved Plants, Ithaca, NY, (10)KU Leuven, Leuven, Belgium

Over the past five years, the Breeding Better Bananas project has made significant progress in genomic selection and linkage mapping of key resistance and horticultural traits in banana including the identification of genomic regions (QTL) associated with fruit filling, banana weevil, and fusarium resistance. Mapping of diploid banana populations have revealed QTL associated with fusarium race 1 resistance in the mapping population ‘Paliama’ x ‘Borneo’ as well as structural variations at a Fusarium wilt tropical race 4 resistance locus in *Musa acuminata* ssp. *malaccensis*. In a population derived from a 4x x 2x cross, where QTL mapping is less straightforward, DArTSeq data was used to estimate the allele frequencies of triploid progeny and allowed for the mapping of QTL for weevil resistance. In the first papers published in genomic selection in banana, a moderate to high prediction accuracy (0.47 to 0.75) of bunch weight and its component traits was observed in a population derived from East African Highland bananas, suggesting that genomic selection has the potential to improve banana breeding efficiency. Association genetics revealed the main QTL controlling bunch weight and its component traits in the population to be localized on chromosome 3. In addition to the mapping work, whole genome assemblies of a triploid plantain (AAB) and a diploid ‘Mchare’ (AA) clones were produced after illumina shotgun sequencing and scaffolding using the Dovetail Genomics technology and validated using Bionano optical maps. The plantain reference genome consists of A- and B- subgenome assemblies. In order to identify and characterize chromosome translocations, oligo painting FISH was used to investigate the structure of mitotic chromosomes of plantains (AAB) and Tanzanian cooking banana ‘Mchare’ (AA). Each of these outcomes is summarized, and as the project enters its next phase, the upcoming work of phase 2 is previewed.

## **W097: Banana Genomics**

### **Two Large Reciprocal Translocations and their Impact on Chromosome Segregation Characterized in *Musa acuminata burmannica***

Marion Dupouy<sup>1</sup>, Franc-Christophe Baurens<sup>2</sup>, Catherine Hervouet<sup>2</sup>, Frederic Salmon<sup>3</sup>, Mathieu Rouard<sup>4</sup>, Nabila Yahiaoui<sup>2</sup>, Arnaud Lemainque<sup>5</sup>, Guillaume Martin<sup>2</sup> and **Angélique D'Hont**<sup>2</sup>, (1)CIRAD, France, (2)CIRAD, Montpellier, France, (3)CIRAD, Capesterre, France, (4)Bioversity International, Montpellier, France, (5)Genoscope, Institut de biologie François Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France

Banana cultivars are derived from hybridizations between *Musa acuminata* subspecies that diverged following geographical isolation in distinct South-East Asian continental regions and islands. Observation of chromosome

pairing irregularities in meiosis of inter-sub-specific hybrids suggested the presence of large chromosomal structural variations. Genetic linkage analysis in a self-progeny of in the *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' accession and mate-pair sequencing were used to search for chromosomal rearrangements in comparison to the standard structure of the *M. acuminata* genome reference sequence (accession DH Pahang, ssp. *malaccensis*). Two large reciprocal translocations were characterized in 'Calcutta 4' accession. One involved a 240 kb distal region of chromosome 2 and a 7.2 Mb distal region of chromosome 8. The other involved a 20.8 Mb distal region of chromosome 1 and a 11.6 Mb distal region of chromosome 9. Signature segment junctions of these translocations were searched in whole-genome sequencing data from 123 wild and cultivated *Musa* accessions. Both translocations were found only in wild accessions belonging to the *burmannica* genetic group suggesting that they originated in this genetic group. Only two of the 87 cultivars analyzed displayed the 2/8 translocation, while none displayed the 1/9 translocation. Chromosomes segregation was studied in a diploid population involving a structurally heterozygous parent for these translocations. In the case of the 1/9 translocation, the two chromosome structures were found to be mutually exclusive in gametes and the rearranged structures (1T9 and 9T1) were preferentially transmitted to the progeny. In the case of translocation 2/8, the two chromosome structures were also generally found to be mutually exclusive, although a few individuals with chromosomes 8 and 2T8 were observed, and the rearranged structures (2T8 and 8T2) were preferentially transmitted to the progeny. These results should help genetic analysis interpretation and breeding programs involving this disease resistance-rich *burmannica* group.

## **W098: Banana Genomics**

### ***In-vitro* Activation of Retro-Transposable Elements as an Effective Mode of mutagenesis in *Musa***

**Eli Khayat, Rahan Meristem (1998) LTD, Goleta, CA**

The majority of triploid banana and plantains are sterile and contain a stagnant genome, while microbial pathogens evolve rapidly. Recently we have shown that by demethylation of specific loci in the chromatin it is possible to diversify banana genomes and modify important characteristics. Using the demethylation compounds 5-Aza-2'-deoxycytidine it is possible to control DNA methylation of plant meristems. This induces activation of retro-transposable elements (RE) and generation of new genotypes. Given the new insertions of the RE the mutations remain stable for many generations. Employing this technique, we have mutated GAL cultivar and tested the plants for resistance/tolerance to TR4. From a population of 9640 in vitro-mutated plants that have been inoculated with TR4, we selected 514 lines that were asymptomatic. These were evaluated in a field trial in an infected area in the Philippines. An analysis of the mutated genotypes demonstrated sensitivity of particular chromatin regions to the demethylation. We have targeted and selected genotypes with TR4 resistance, altered plant stature and early flowering We analyzed the polymorphic regions following demethylation. An entire genome sequence analysis comparison between the resistant genotype and its precursor mother clone revealed changes in various parts of the genome including a QTL of clustered "R-genes" but not in the coding sequences.

## **W099: Banana Genomics**

### **The Complex Story of Intergenic Recombination in ABB Allotriploid Bananas**

Alberto Cenci<sup>1</sup>, Julie Sardos<sup>1</sup>, Guillaume Martin<sup>2</sup>, Catherine Breton<sup>1</sup>, Nicolas Roux<sup>1</sup>, Rony Swennen<sup>3</sup>, Sebastien Carpentier<sup>4</sup> and **Mathieu Rouard<sup>1</sup>**, (1)Bioversity International, Montpellier, France, (2)CIRAD, Montpellier, France, (3)KU Leuven, Leuven, Belgium, (4)Bioversity International, Leuven, Belgium

Bananas (*Musa* spp.) are a major staple food for hundreds of millions of people in developing countries. The cultivated varieties are seedless and parthenocarpic clones of which the ancestral origin remains to be clarified. The most important cultivars are triploids with an AAA, AAB, or ABB genome constitution, with A and B genomes provided by *M. acuminata* and *M. balbisiana*, respectively. Previous studies suggested that inter-genome recombinations were relatively common in banana cultivars and that triploids were more likely to have passed through an intermediate hybrid. In this study, we investigated the chromosome structure within the ABB group, composed of starchy cooking bananas that play an important role in food security. Using SNP markers called from RAD-Seq data, we studied the chromosome structure of 36 ABB genotypes spanning defined taxonomic subgroups. To complement our understanding, we search for similar events within nine AB hybrid genotypes. Recurrent Homoeologous Exchanges (HEs), i.e. chromatin exchanges between A and B subgenomes were unraveled with at least 9 founding events at the origin of the ABB bananas prior to the clonal diversification. The discovery of this

nine founding events allows discussing the possible routes that led to the creation of the different subgroups and formulate new hypotheses. Based on our observations, we suggest different routes that gave rise to the current diversity in the ABB cultivars. Routes involving primary AB hybrids, routes leading to shared HEs and routes leading to a B excess ratio. Genetic fluxes took place between *M. acuminata* and *M. balbisiana*, particularly in India, where these unbalanced AB hybrids and ABB allotriploid originated and where cultivated *M. balbisiana* are abundant. The result of this study clarifies the classification of ABB cultivars and leading possibly to the revision of the classification of this subgroup. This is an important step to unravel the origin of polyploid bananas, and contributes to possible scenarios on the origin. ABB bananas are hypothesized to be more drought tolerant. Knowing the origin of our current cultivars and so their potential parents will help breeders to make the right choices for future crosses. The *M. balbisiana* genome is a good source to create new cultivars able to answer the numerous challenges of banana breeding.

## **W100: Banana Genomics**

### **New Insight for Banana Resistance to *Xanthomonas vasicola* pv. *musacearum***

**Valentine Nakato**, International Institute of Tropical Agriculture,, Kampala, Uganda

Xanthomonas wilt caused by *Xanthomonas vasicola* pv. *musacearum* (Xvm) is one of the most devastating disease of banana in East and Central Africa. All banana varieties, except the wild *Musa balbisiana*, were reported susceptible to Xvm. However, *M. balbisiana* is not ideal for breeding since it has the B genome (BB subgroup), while most edible banana have the A genome. To further confirm whether all banana varieties were susceptible to Xvm, we evaluated a broader range of banana varieties, that represented the entire genetic diversity of banana. 72 banana accessions were artificially inoculated with a virulent Xvm isolate. *M. balbisiana* was confirmed as resistant. We further identified Monyet, Zebrina, Saba, Buitenzorg, Tani, IC2, Pelipita, Kikundi, Cameroun, P. Raja and Maia Oa as potential sources of resistance to Xvm. This finding unlocks doors for using genetic resistance to manage Xvm and contribute to food security for millions of smallholder banana farmers in the region who are dependent on the crop for livelihoods. We are in the process of elucidating the genetics of resistance and development of markers to expedite transfer of resistance to susceptible but preferred east African highland bananas.

## **W101: BER Plant Genomic Science**

### **Introduction and Overview of KBase and Joint Genome Institute Plant Program**

**Robert W. Cottingham**, Oak Ridge National Laboratory, Oak Ridge, TN and Jeremy Schmutz, HudsonAlpha Institute for Biotechnology, Huntsville, AL

The Department of Energy Biological and Environmental Research (DOE-BER, [www.energy.gov/science/ber](http://www.energy.gov/science/ber)) program supports research and facilities to achieve a predictive understanding of complex biological, earth, and environmental systems with the aim of advancing the nation's energy and infrastructure security. The program seeks to discover the underlying biology of plants and microbes as they respond to and modify their environments. This knowledge enables the reengineering of microbes and plants for energy and other applications. BER research also advances understanding of the dynamic processes needed to model the Earth system, including atmospheric, land masses, ocean, sea ice, and subsurface processes.

BER funds both a large scale user facility for plant genomics at the DOE Joint Genome Institute ([www.JGI.doe.gov](http://www.JGI.doe.gov)), and an open and collaborative computational resource for predictive systems biology of microbes, plants and their communities called the DOE Systems Biology Knowledgebase ([www.KBase.us](http://www.KBase.us)). Both endeavor to help scientists conduct experiments and analyses in areas such as improving biofuel development, understanding plant model systems, advancing plant comparative science and investigating global carbon cycling. In this annual workshop, speakers present current and ongoing developments in their research enabled by JGI and KBase toward increasingly large-scale and integrative biology relevant to DOE-BER mission. We will also give a brief introduction describing how to apply for access to the JGI Community Science Program, and how you can use KBase to accelerate your plant genomics research.

JGI enables scientific advances accomplished in collaborative projects through the Community Sequencing Program and the DOE BioEnergy Research Centers. The JGI Plant Program is dedicated to applying advances in genomic technologies for understanding fundamental plant biology through comparative genomics and targeted experiments. Our major goal, in collaboration with plant scientists, is to apply this understanding from genomics to accelerate the improvement and domestication of biofuel crops. The JGI Plant program has produced many of the high-quality reference plant genomes available today and we continue to curate and make available comparative data and analysis via [phytozome.jgi.doe.gov](http://phytozome.jgi.doe.gov). The JGI Plant Flagship genomes are continually improved for accuracy and completeness of the genome sequence and the reference annotation.

KBase hosts a suite of resources, Apps, and workflows designed specifically to explore plant genomics data and model plant-microbe interactions. KBase Apps are available to run quality control on reads, annotate plant transcripts, group orthologous proteins from multiple genomes, analyze differential expression, build gene trees, and more. Data is available from both JGI's Phytozome and MycoCosm that can easily be used for analyses within a KBase Narrative to build metabolic models of your organism or with data from PlantSEED. Create your own workflows by combining Apps or follow tutorials to identify gene families, model plant- microbe metabolic interactions, and analyze RNA-seq data. To learn the RNA-seq Pipeline follow the Narrative Tutorial <http://kbase.us/expression-analysis/>. KBase provides reference guided RNA-seq pipelines for microbial, fungal, and plant genomes obtained from the Illumina platform. Visit [kbase.us](http://kbase.us) to learn how KBase can enhance your plant science research.

## **W102: BER Plant Genomic Science**

### **Exascale Networks for Arabidopsis in Kbase**

**Daniel Jacobson**, Oak Ridge National Laboratory, Oak Ridge, TN

## **W103: BER Plant Genomic Science**

### **Understanding the Regulation of Crassulacean Acid Metabolism through a New Genomic Model in *Yucca* (Asparagaceae)**

**Karolina Heyduk**<sup>1</sup>, Kerrie W. Barry<sup>2</sup>, Chris Daum<sup>3</sup>, Anna Lipzen<sup>3</sup>, Jane Grimwood<sup>4</sup>, Jerry Jenkins<sup>4</sup>, Christopher B. Plott<sup>4</sup>, Richard Field<sup>5</sup>, Jeremy Schmutz<sup>4</sup> and Jim Leebens-Mack<sup>5</sup>, (1)Yale University, New Haven, CT, (2)DOE Joint Genome Institute, Berkeley, CA, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (5)University of Georgia, Athens, GA

Crassulacean acid metabolism (CAM) is a modification to the core photosynthetic pathway of plants that increases photosynthetic capacity under water limited conditions. CAM plants include many species that experience frequent or prolonged water stress – for example, cacti and agave species use CAM to survive in their harsh desert environments. For the last several years, CAM genomics has burgeoned, with 6 published genomes and numerous transcriptomic studies. Yet progress in understanding the regulatory mechanisms of CAM have been hampered by a lack of comparative systems, where closely related C<sub>3</sub> and CAM species can be examined. Furthermore, the CAM phenotype is a continuum, where plants can use varying degrees of CAM and C<sub>3</sub> photosynthesis and can modulate that ratio in response to environmental stressors. To advance the understanding of regulatory mechanisms in CAM, we have generated draft genomes of two closely related *Yucca* (Asparagaceae) species, *Y. aloifolia* (CAM) and *Y. filamentosa* (C<sub>3</sub>). These two species have hybridized in nature to form a third species, *Y. gloriosa*; assessments of physiology and anatomy from the hybrid reveal that not only is it photosynthetically intermediate between the parental species, but genotypes of *Y. gloriosa* also vary in their ability to up-regulate the CAM pathway in response to drought stress. Using preliminary drafts of the genomes – including a chromosome-level assembly of *Y. aloifolia* – along with resequencing data from all three species and extensive RNA-seq from *Y. gloriosa*, we examine the connection between parental genomic contribution and gene expression and their combined effect on the regulation of CAM in *Y. gloriosa*.

## **W104: BER Plant Genomic Science**

### **Comparative Genomics Analysis of Drought Response between CAM and C<sub>3</sub> Photosynthesis Plants**

Rongbin Hu<sup>1</sup>, Jin Zhang<sup>1</sup>, Haiwei Lu<sup>1</sup>, Sara S. Jawdy<sup>1</sup>, Anna Lipzen<sup>2</sup>, Mei Wang<sup>3</sup>, Vivian Ng<sup>3</sup>, Avinash Sreedasyam<sup>4</sup>, Jeremy Schmutz<sup>2</sup>, Degao Liu<sup>1</sup>, Priya Ranjan<sup>1</sup>, Jay Chen<sup>1</sup>, Timothy J. Tschaplinski<sup>1</sup>, Gerald Tuskan<sup>5</sup> and **Xiaohan Yang<sup>5</sup>**, (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)DOE Joint Genome Institute, (4)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (5)Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Plants exhibiting crassulacean acid metabolism (CAM) – a specialized mode of photosynthesis – close stomata during the day and take up CO<sub>2</sub> at night, resulting in higher levels of water-use efficiency and drought resistance relative to C<sub>3</sub> and C<sub>4</sub> photosynthesis species. It is widely accepted that obligate CAM is under ontogenetic controls. However, it is largely unknown whether CAM-related genes are regulated by drought stress in obligate CAM species. Also, it is not clear whether obligate CAM and C<sub>3</sub> species share commonality in gene expression in response to drought stress. To address these questions, we performed transcriptome-sequencing of mature leaf samples collected at dawn (i.e., 2 h before the beginning of light period) and dusk (i.e., 2 h before the beginning of dark period) from *Kalanchoe fedtschenkoi* (an obligate CAM photosynthesis species) plants, which were grown in 12 h light/12 h dark photoperiod and subjected to three treatments (i.e., moderate water deficit, severe water deficit, well-watered control). We identified up-/down-regulated *K. fedtschenkoi* genes in response to drought treatments. Also, we performed weighted gene co-expression network analysis of the drought responsive genes in *K. fedtschenkoi*. Furthermore, we performed comparative analysis of transcript profiles of drought-responsive genes between *K. fedtschenkoi* and two C<sub>3</sub> photosynthesis species (*Arabidopsis thaliana* and *Populus deltoides*). This research provides new insight into the conservation and divergence of molecular mechanisms underlying drought response between CAM and C<sub>3</sub> photosynthesis species.

## **W105: BER Plant Genomic Science**

### **Genetics of Genotype-by-Environment Interactions in the Switchgrass Diversity Panel**

**Alice H. MacQueen**, University of Texas at Austin, Austin, TX

As sessile organisms, plants cannot move to escape unpredictable and changing environments. Which environments impact plants the most? How do plant genetic responses to the environment vary, and how do these responses evolve? One common hypothesis is that adaptation to specific environments, or local adaptation, occurs via tradeoffs involved in specialization: alleles with antagonistic pleiotropy increase fitness in specific environments, but have negative, pleiotropic effects in alternate environments. A contrasting hypothesis at the level of the allele is conditional neutrality, where alleles can increase fitness in specific environments without costs in alternative environments. As climates shift and climate variability increases, access to conditionally neutral alleles that improve fitness in specific stressful environments will be essential for improving crop species. Genomics-enabled research is now providing the statistical power to discover and characterize allelic variation in genes involved in adaptation.

Switchgrass (*Panicum virgatum*) is an outcrossing, polyploid C<sub>4</sub> perennial grass that has been championed as a promising biofuel feedstock. It is a common member of most native North American prairie communities and exhibits extensive phenotypic variability and adaptation across its range, particularly in response to latitude and precipitation gradients. Here, I report on the development of a diversity panel for switchgrass. This diversity panel includes over 700 sequenced genotypes sampled from the majority of the range of switchgrass across the eastern United States. Clones of the sequenced individuals were planted at ten field sites covering 17° of latitude (1800 km) in the central United States. This allowed us to evaluate the contributions of individual loci to traits and fitness over a wide range of climatic conditions. Here, I present genetic analyses of phenology data from the 2019 growing season. In particular, I present preliminary results from genome-wide association studies aimed at detecting gene-by-environment interactions for phenology traits across the species' latitudinal range.

## **W106: BER Plant Genomic Science**

### **Fungal Modeling in KBase: Automated Reconstruction, Manual Curation, Evaluation and Comparison of Diverse Genome-Scale Fungal Metabolic Models**

**Janaka N. Edirisinghe**, Argonne National Laboratory, Lemont, IL

Fungal genome-scale metabolic models are an efficient way of predicting phenotypes across various environmental conditions. These metabolic models are a key tool in understanding fungal-bacterial and plant-fungal community behavior.

However, automating the construction of high-quality fungal models has been a challenge. Here we introduce a methodology to construct genome-scale fungal models in an automated fashion based on a highly curated set of reactions that are derived from 14 published fungal metabolic models. As the basis for the method, we produced a fungal model template that encompasses the biochemistry data from the published fungal models and the structural annotations from the associated fungal genomes. The underlying biochemistry data that is being used for model generation were improved through extensive manual curation via Escher map based curation environment (<https://escher.github.io/#/>) in order to reduce the redundancy and to improve the quality. This was done through community involvement of fungal experts.

Our approach uses structural annotations of any user-submitted fungal genome and computes a set of orthologous proteins against the curated fungal template in order to assert the presence or absence of specific biochemical reactions and pathways. These orthologous families were then curated and mapped to biochemistry by the expert curators. Then the related biochemistry data is propagated to construct a new draft metabolic model. Once the draft metabolic models are derived, additional reactions been added based on available functional annotations. This method is deployed in the Department of Energy Systems Biology Knowledgebase (KBase) (<https://narrative.kbase.us/>) as an app called "Build Fungal Model" ([https://narrative.kbase.us/#catalog/apps/kb\\_fungalmodeling/built\\_fungal\\_model/release](https://narrative.kbase.us/#catalog/apps/kb_fungalmodeling/built_fungal_model/release)). This method able to produce a draft fungal metabolic model in an around one hour. We applied this new app to bulk construct draft fungal models for more than 170 fungal genomes imported from Joint Genome Institute (JGI) MycoCosm resource (<https://mycocosm.jgi.doe.gov/mycocosm/home>). We compared these models side-by-side, exploring how each genome overlaps with our curated model template and plotting model variance along the phylogenetic tree of fungal genomes. All draft fungal models are available for download via a KBase Narrative.

## **W107: BER Plant Genomic Science**

### **Plant Genomic References of the Future**

**Jerry Jenkins**<sup>1</sup>, Jane Grimwood<sup>1</sup>, Shengqiang Shu<sup>2</sup>, Christopher B. Plott<sup>1</sup>, David Flowers<sup>3</sup>, Yuko Yoshinaga<sup>2</sup>, Therese Mitros<sup>4</sup>, Kerrie W. Barry<sup>5</sup>, Chris Daum<sup>6</sup>, Kankshita Swaminathan<sup>7</sup>, Daniel Rokhsar<sup>2</sup> and Jeremy Schmutz<sup>1</sup>, (1)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)HudsonAlpha Institute, (4)UC Berkeley, Berkeley, CA, (5)DOE Joint Genome Institute, Berkeley, CA, (6)Joint Genome Institute, Walnut Creek, CA, (7)University of Illinois Urbana-Champaign, Urbana, IL

In the last few years Single Molecule, Real-Time Long Read Pacific Biosciences (PacBio) Sequencing has enabled the generation of high-quality de novo plant assemblies with contig N50s in the megabase range. Combined with Hi-C (chromosome conformation capture) we can construct accurate, whole chromosome assemblies, even without extensive mapping information. This increased contiguity enables the accurate localization of genes and regulatory elements in complex regions. The introduction of the PacBio Sequel2 has greatly increased our ability to generate plant de novo genome sequences and it also has enabled us to generate more accurate annotations with significant amounts of FLcDNA expression coverage. For our Department of Energy Joint Genome Institute plant genomes, we combine high coverage PacBio with Hi-C and PacBio and Illumina expression data to produce the most accurate annotated references available for plant genomics. For large scale production or larger genomes, long

read PacBio assemblies come at a high computational cost, with mixing of alternative haplotypes, and a high rate of indel errors in the resulting consensus. With the Sequel2, PacBio has recently developed a new data type known as CCS/"Hi-Fi" reads, where the subreads of a sequence are used to generate a 10-15 Kb read that has an accuracy of up to 99.9%. These CCS reads have a reduced computational burden and enable accurate sorting of haplotypes as each read represents a single haplotype. We will present ongoing work to produce more accurate and complete reference genomes for complex plants and a new sequencing and assembly strategy for generating CCS-based genome sequences. We will highlight initial results using CCS reads to assemble inbred genomes (*Phaseolus vulgaris* and *Sorghum bicolor*), as well as outbred *Populus trichocarpa* and *Miscanthus sinensis*.

### **W108: Beyond Drosophila: Genomic advances in non-model Diptera**

#### **Conditional Transgenic Male-Only System Development in the Sheep Blowfly, *Lucilia cuprina***

**Megan E. Williamson**, North Carolina State University, Raleigh, NC, Ying Yan, Fraunhofer IME-BR, Germany and Max J. Scott, North Carolina State University

The Australian sheep blowfly, *Lucilia cuprina*, is a major agricultural pest in Australia and New Zealand. Genetic approaches, such as the sterile insect technique (SIT), have been considered for use for control. For over 60 years, SIT has been used successfully to control the New World screwworm, a blowfly that is a close relative of *L. cuprina*. However, the high cost of SIT has prevented use of this technology for control of *L. cuprina* and also screwworm in South America. In SIT programs, females cost money to rear and can hinder the effectiveness of the program by competing with wildtype females to mate with sterile males. For this reason, our lab has developed transgenic sexing systems that produce only males by killing the females early in development when fed a diet lacking tetracycline. A disadvantage of these strains is that only half the eggs laid produce males. Here we report on a novel conditional transformation system that utilizes RNA interference to transform XX females into XX males. Phenotypes observed range from no transformation to full transformation of XX progeny. Conditional female transformation strains could also be made using CRISPR/Cas9 technology. We will report on our preliminary work using CRISPR/Cas9 to knockout genes required for body color and female development. This work shows promise to produce more efficient conditional male-only strains, allowing for more cost-effective mass rearing and improved control programs.

### **W109: Beyond Drosophila: Genomic advances in non-model Diptera**

#### **Host Shifting and Host Sharing in a Genus of Specialist Flies Diversifying alongside their Sunflower Hosts**

**Alaine C. Hippee**, University of Iowa, Iowa City, IA

Among host-specific insects, it is unusual for congeners to feed on the same tissue of the same hosts, likely because of strong selection against niche overlap. How multiple species come to use, and coexist, on the same tissue of the same host may reveal important insights into ecological factors underlying the origins and maintenance of diversity. Sunflower maggot flies in the genus *Strauzia* are endemic to North America and their larvae feed on the pith of plants in the sunflower family. Although *Strauzia* tend to be host specialists, they also appear to overlap in their host use on two different sunflower species. To resolve the origins of host sharing among these specialist flies, here we use reduced representation genomic libraries to infer the first multi-locus phylogeny for genus *Strauzia*. Our results show that *Helianthus tuberosus* and *H. grosseserratus* each harbor three different fly species, and that the flies using each respective host are not one another's closest relatives (i.e., host sharing is most likely the result of host shifts, not speciation *in situ*), highlighting the value of genomic data in resolving phylogenetic questions in non-model organisms. We also find evidence for incomplete lineage sorting across our dataset, suggesting that many *Strauzia* species may be of recent origin. Because many of the plant hosts of *Strauzia* (including *H. tuberosus* and *H. grosseserratus*) belong to a young (1-2 MYA) clade of perennial sunflowers noted for their frequent introgression and hybrid speciation events, this may be a case where rapid and recent adaptive introgression and speciation in a group of plants has led to rapid and recent host-shift speciation among their phytophagous fly associates, including the convergence of multiple species on the same host plants. Future work will include demographic

modelling of *Strauzia* divergence and host shifting events to explore the relationship between sunflower and *Strauzia* evolutionary history.

### **W110: Beyond Drosophila: Genomic advances in non-model Diptera**

#### **New Approaches to Generate Single Insect Haplotype Assemblies**

**Haynes Heaton**<sup>1</sup>, Christine Lambert<sup>2</sup>, Primo Baybayan<sup>3</sup>, Brendan Galvin<sup>3</sup>, Sarah B Kingan<sup>3</sup>, Jonas Korlach<sup>3</sup>, Mathilde Gendrin<sup>4</sup>, Jeremy Herren<sup>5</sup>, Daniel Neafsey<sup>6</sup>, Richard Durbin<sup>7</sup> and Mara Lawniczak<sup>1</sup>, (1)Wellcome Sanger Institute, Hinxton, United Kingdom, (2)Pacific Biosciences, (3)Pacific Biosciences, Menlo Park, CA, (4)Institut Pasteur, French Guiana, French Guiana, (5)icipe, Nairobi, Kenya, Kenya, (6)Harvard, (7)University of Cambridge, Cambridge, United Kingdom

High heterozygosity levels cause problems for building genome reference assemblies when multiple individuals are pooled, but when single individuals are sequenced, high heterozygosity can actually be advantageous to help build assemblies. Here we discuss progress in long-read and long-range sequencing for single small insects, including a reliable approach for multiplexing smaller genomes (< 400Mb) on single PacBio 8M SMRT Cells and using very low input tissue for Hi-C, allowing multiple sequencing technologies for single specimens. In order to generate unbiased haplotype assemblies, we have developed methods that treat haplotype phasing as a first-class object. We use *k*-mer methods along with long range genetic information from linked reads, long reads, and Hi-C to *de novo* phase heterozygous sequences prior to assembly graph construction. We demonstrate this *de novo* phasing on *Anopheles arabiensis* in which we create megabase scale linear phasing consistency graphs. This phasing can then be used to split haplotypes prior to assembly which has been shown to improve contiguity and accuracy of the resulting assemblies.

### **W111: Beyond Drosophila: Genomic advances in non-model Diptera**

#### **Precision Guided Sterile Insect Technique (pgSIT): A Novel Environmentally Friendly Technology for Insect Population Suppression**

**Nikolay Kandul**, University of California, San Diego, La Jolla, CA

Sterile insect technique (SIT) is an environmentally safe and proven technology to suppress wild insect populations. In a classic SIT, sterile males are generated with radiation, sex-sorted to remove females, and released in large numbers into a wild population, where they compete with wild males for female mates. Mating with sterile males results in no progeny and consequently can be used to suppress populations; however, the low fitness of sterile males and inability to adequately sex-sort most insects has limited the application and efficacy of SIT. To overcome these limitations, we developed a novel CRISPR-mediated SIT termed “precision guided SIT” (pgSIT) system. This system is comprised of two separate homozygous lines each carrying Cas9 or double guide RNA (dgRNA). When these lines are crossed in the laboratory, they produce F1 progeny with precise knock outs of two specific genes targets, in this case, one required for female survival and the other for male fertility, resulting in emergence of 100% sterile males from genetically identical F1 eggs. pgSIT mechanistically relies on a dominant genetic technology that enables simultaneous sexing and sterilization, permitting the release of eggs into the environment. To date, we have engineered multiple pgSIT systems in the model organism, *Drosophila melanogaster*, the crop pest, *Drosophila suzukii*, and the disease vector, *Aedes aegypti*, which consistently produce 100% sterile males. Importantly, unlike irradiated sterile males made with a classic SIT, pgSIT generated males in these species are fit and competitive compared to wild type males. Modelling has also predicted that pgSIT will induce greater population suppression over time than can be achieved by current approaches. Taken together, pgSIT may transform our ability to control insect agricultural pests and disease vectors.

### **W112: Beyond Drosophila: Genomic advances in non-model Diptera**

#### **Genomic Insights into Beneficial Virus Evolution within Fruit Fly Parasites**

**Kelsey A. Coffman**, University of Georgia, Athens, GA

Fruit flies (Family Tephritidae) serve as hosts for numerous species of insects called parasitoid wasps. As eggs, parasitoid wasps are laid within other insects and consume the host insect as part of wasp development. Host survival therefore depends on the ability of the host immune system to kill an invading wasp, while wasp survival depends on evasion of such immune responses. As a result, wasps and their hosts have coevolved strategies to gain an advantage in this evolutionary arms race. Although normally considered to be strict pathogens, some viruses have established persistent infections within parasitoid wasp lineages and bolster wasp survival via host manipulation during parasitism. Heritable associations between viruses and parasitoid wasps have evolved independently multiple times, including two recently characterized examples in parasitoids that attack agriculturally important fruit fly pests. The wasp species *Diachasmimorpha longicaudata* and *Fopius arisanus* maintain heritable associations with a poxvirus and an endogenous nudivirus, respectively. Comparative and functional genomic data within these two systems has revealed strong convergent evolution between beneficial virus acquisitions that evolved from divergent pathogenic ancestors. Our results have shed light on both shared features and key distinctions that dictate how these types of insect-microbe associations form and persist.

### **W113: Beyond Drosophila: Genomic advances in non-model Diptera**

#### **Chromosome Level Assemblies of Arthropod Pests and Vectors for Genetic Control**

**Alistair Darby**, University of Liverpool, Liverpool, United Kingdom

Insects have relatively small genomes but have high levels of heterozygosity and large regions of low complexity DNA repeats.

We have used PacBio and Hi-C to produce four chromosome level assemblies: the Potato Aphid (*Macrosiphum euphorbiae*), Diamond back moth (*Plutella xylostella*), *Aedes aegypti* mosquito and the Mediterranean fruit fly (*Ceratitis capitata*).

In all cases (except the aphid) we have sequenced the heterogametic sex to provide high resolution sex chromosome information. The chromosome level assemblies provide information on genome synteny within orders and direct evidence on chromosome fusions, fissions and large scale variation. These genomes have also allowed us to investigate the sources of heterozygosity and identify complex haplotypes differences within genomes. In some cases chromosome variations can facilitate the mapping of key genes that have potential use in genetic sexing strain and pest control.

### **W114: Big Data: Manage your data before your data kills you**

#### **Librarians are here to Help but We Can't Read Minds**

**Megan O'Donnell**, Iowa State University, Ames, IA

Academic libraries have invested in staff and infrastructure, such as repositories, to support data management and data sharing but researchers are often unaware that these services exist. A brief overview of which "data services" are available at most academic research libraries will be followed by a dive into the top five things to know about publishing data with your academic library. Come to learn how your library can publish data, stay for practical advice on how to get it done.

### **W115: Big Data: Manage your data before your data kills you**

#### **Managing Multi-Site Academic Studies: The Trials and Tribulations and the Techniques to Enable Studies and Investigators to be Successful**

**Jill Wait**, BioMarin Pharmaceutical Inc., Novato, CA

BioMarin's mission to develop therapies for rare diseases and the nonclinical studies required for submissions to FDA and international regulatory agencies often require BioMarin to work with academic institutions that have the relevant animal disease models. Unfortunately, these institutions are rarely Good Laboratory Practice (GLP)-compliant facilities with robust quality management systems to ensure

the reliability, reproducibility, quality, and integrity of the study. Furthermore, these studies often use specialized assays and techniques, which require the involvement of multiple sites to achieve all of the study objectives. Multi-site studies can be complicated to manage and are more likely to generate an increase in protocol deviations and operational miscommunications. Over the years, BioMarin has developed processes and techniques to successfully prepare, execute, and monitor multi-site academic studies to ensure compliance with the study protocol, to mitigate the gaps compared to the GLP regulations, and to ensure that a quality, scientifically-robust report is generated. These processes and techniques will be described.

### **W116: Big Data: Manage your data before your data kills you**

#### **Data Resources and Implementation for the Genomes to Fields (G2F) Initiative**

**Bridget McFarland**, University of Wisconsin-Madison, Madison, WI and Genomes To Fields Collaborators

Data resources and infrastructure are needed to support the efficient and sustainable production of food, feed, fuel and fiber for an increasing world population in the context of variable environmental conditions. The Genomes To Fields (G2F) initiative is a multi-institutional effort that seeks to respond to this challenge by developing a flexible and distributed infrastructure that can adjust to address emerging problems. As part of this effort, G2F has generated large-scale phenotypic, genotypic and environmental datasets using existing, publicly available maize inbred lines and hybrids evaluated through a network of public collaborators that are part of the G2F's Genotype by Environment (G × E) project. With the help of approximately 40 PIs across the United States and Canada, this project has released 2014, 2015, 2016 and 2017 datasets to the public. The processes of standardizing data collection, quality control and distribution in collaborations with CyVerse and G.E.M.S. will be shared to illustrate how this effort has established best practices for data resources and infrastructure generated through this collaborative effort. Through the formation of the G2F collaboration, datasets and technologies are made available to increase the range of possible research on crop improvement.

### **W117: Big Data: Manage your data before your data kills you**

#### **Managing UAS Imagery for Developmentally-Driven Decision Making**

**Katy Martin Rainey**<sup>1</sup>, Keith Cherkauer<sup>1</sup>, Fabiana Moreira<sup>1</sup>, Stuart D Smith<sup>2</sup>, Bilal Jamal Abughali<sup>2</sup>, Anthony Hearst<sup>3</sup> and E. Vincent Seal<sup>2</sup>, (1)Purdue University, West Lafayette, IN, (2)Purdue University, (3)Progeny Drone Inc., Lafayette, IN

Unmanned aerial systems (UAS) and image analysis allow plant breeders to measure phenotypic variability on the same genotype multiple times during a season, producing variables defined as longitudinal traits. From agronomic research using UAS, multiple data types are generated at different scales and over time, for both observations and metadata. The images themselves should not really be considered the data because often extensive post-processing, including modeling, is required to obtain validated plot-level metrics. Ideally these data are managed, and outputs generated, so that decision-making happens during the season. This provides breeders with new approaches like using yield predictions for early-season selection and subsampling. So that the engineering does not get in the way of the genetics, the desired data quality should be defined relative to scale, size, source, and georeferencing. Our approach is to use robust and inexpensive RGB data collected over many environments and many sampling dates. Certain data management protocols and tools allow crop scientists and agronomists to progress with applications of these new and valuable data. Emerging technologies will follow in the wake of that progress.

### **W118: Big Data: Manage your data before your data kills you**

#### **Big Data, Big Experiments, and Big Problems: Strategies for Experimental Design, Execution, and Analysis**

**Susan Vanderplas**, University of Nebraska Lincoln, Lincoln, NE

If data creates problems, then big data creates even bigger problems. In this talk, I discuss many different ways you can ensure a study fails, and how those decisions are magnified in "big data" settings. If you've ever thought pilot studies are a waste of time, or that planning your data analysis before data collection begins is for losers, or if you think of your instrumentation as a magic box that produces data, this talk is for you. Using several recent experiments which generated TB of data, I talk about the analysis of misshapen, noisy, and problematic data, and how to avoid these pitfalls in during experimental design and data collection.

### **W119: Bioactive, Nutraceutical and Therapeutic Molecules in Plants**

#### **Lunasin: An Anti-Cancer Soybean Peptide**

**Benito de Lumen**, Nutritional Sciences & Toxicology/University of California, Berkeley, Berkeley,, CA

### **W120: Bioactive, Nutraceutical and Therapeutic Molecules in Plants**

#### **Genetic Mapping of Seed Isoflavone Content in Soybean**

**Abdelmajid Kassem**, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC

Phytoestrogens (or isoflavones) are secondary metabolites abundant in soybean seeds and have tremendous health and nutrition benefits in humans and animals. The most important and widely studied phytoestrogens are glycitein, daidzein, and genistein. Here, we summarize 6+ years of research efforts in constructing genetic linkage maps based on various molecular markers and recombinant inbred line (RIL) populations and QTL mapping of seed phytoestrogen content. Efforts are under way to identify candidate genes involved in these seed phytoestrogens' biosynthetic pathways. These advances will help in developing cultivars and lines with high amounts of phytoestrogens.

### **W121: Bioactive, Nutraceutical and Therapeutic Molecules in Plants**

#### **The Tuareg Pharmacopeia: Basil; Aloe and Willow Teas; Work on Drug Resistant Cancers**

**David Lightfoot**, Southern Illinois University, Carbondale, IL

The original aim of the analysis was based on willow leaves. Aspirin was less effective than willow tea leaves? Basil tea worked the best.

Ethanol extracts were less effective.

Surprisingly, of just 13 natural products of plants used by the Tuaregs we have predicted the most likely artificial mixtures of 2-3 most effective natural products on drug resistant leukemia cells from over 364 possible mixtures. □The natural products selected included resveratrol, honokiol, chrysin, limonene, cholecalciferol, cerulenin, aloe emodin, and salicin. They had over 600 potential protein targets. Target profiling used the Ontomine™ set of tools for literature searches of potential binding proteins, binding constant predictions, binding site predictions, and pathway network pattern analysis. □The analyses indicated that 6 of the 13 natural products predicted binding proteins which were important targets for established cancer treatments. Improvements in effectiveness were predicted for artificial combinations of 2 or 3 natural products. □That effect might be attributed to drug synergism rather than increased numbers of binding proteins bound (dose effects). Among natural products, the combinations of Aloe emodin with mevinolin and honokiol

were predicted to be the most effective combination for AML-related predicted binding proteins.  
□Therefore, plant extracts may in future provide more effective medicines than the single purified natural products of modern medicine, in some cases. However, drink herbal teas whenever you can...

### **W122: Bioactive, Nutraceutical and Therapeutic Molecules in Plants**

#### **Improvement of Bioactive Ingredients in Potato: Starch and Antioxidant Compounds as Breeding Targets**

**Benoit Bizimungu**, Agriculture and Agri-Food Canada, Fredericton, NB, Canada

Potato is the most important non-cereal food crop worldwide. Its global distribution, nutritional value, taste and cooking versatility have contributed to its current position as a major staple food. Breeding research aims to add value to existing varieties to satisfy complex demands of the potato supply chain from 'farm to fork'. In order to meet consumer's demands for healthy, functional foods and food ingredients, potato geneticists and breeders are looking for ways to optimize the nutritional composition of potato. This presentation focusses on starch and antioxidant compounds as breeding targets to boost the nutritional profile of cultivated potato. Given the popularity of potato in human diets, genetic improvement of bioactive ingredients can have a positive impact on consumer's health and wellness. We will discuss breeding progress and relevant factors affecting the quality of potato-based food products including the complex interactions of genotype, production environment, and post-harvest processing. Recent advances in DNA-sequencing technology offer the prospect of more efficiently harnessing genetic diversity in potato genetic resources to meet changing consumers and production needs.

### **W123: Bioactive, Nutraceutical and Therapeutic Molecules in Plants**

#### **Explore Genetic Diversity of Soybean Fatty Acid Biosynthetic Pathway by Seed Transcriptome Sequencing**

**Yong-Qiang Charles An**, USDA-ARS Plant Genetics Research Unit, Donald Danforth Plant Science Center, ST. Louis, MO

Soybean structural genome diversity studies have received great attention since availability of next generation sequencing technologies. Over 1,500 diverse soybean genomes have been re-sequenced worldwide. However, the transcriptome diversity among different soybean genotypes have been less explored and exploited. We sequenced seed transcriptomes of 197 diverse soybean genotypes. Having developed and applied a variety of bioinformatics pipelines and data-mining strategies, we generated a large collection of transcript sequence and accumulation polymorphisms that potentially lead to functional variation of the genes among those genotypes. We demonstrated that transcriptome sequencing could offer a highly effective alternative approach for genotyping and gene function discovery. As a proof-of-concept study, we have applied transcriptome sequencing approaches to independently identify previously known and unknown transcript sequence and accumulation polymorphism, splicing variation and DNA deletions of fatty acid biosynthetic pathway genes, which are associated with changes in fatty acid profiles in soybean seeds.

### **W124: Bioactive, Nutraceutical and Therapeutic Molecules in Plants**

#### **TILLING by Target Exome Capture Sequencing as New and Effective Platform for Gene Functional Analysis**

**Naoufal Lakhssassi**, Southern Illinois University at Carbondale, Carbondale, IL

Soybean is an attractive crop for the bio-industry due to being a relatively cheap raw material. The content of the eighteen carbon chain fatty acids like saturated stearic acid and unsaturated fatty acids such as oleic acid and linoleic acid represent ~3%, ~20%, and ~55% in soybean seed oil. These fatty acids are precursors for many bioactive fatty acid compounds and therefore, extensive research to boost their content in soybean seeds is taking place. To understand genes and gene networks involved in the

FA biosynthesis pathway in soybean, we used Tilling by Target Exome Capture Sequencing (TbyTECS) to study the role of each enzyme in the soybean fatty acid biosynthesis pathway in order to produce a desirable germplasm. We developed a high-throughput TbyTECS technology coupled with universal bioinformatic tools to identify population-wide mutations in soybeans. Because of the robustness of single nucleotide polymorphisms (SNPs) calling, this novel technology ensures high quality yield of true mutations. Using this reverse genetic approach, we were able to isolate hundreds of mutants that increased our understanding. All *GmSACPD*, *GmFAD2*, and *GmFAD3* gene family members undergo a process of subfunctionalization event. The TbyTECS technology provides an unprecedented platform for highly effective screening of polyploid mutant populations and gene functional analysis. The obtained soybean mutants in this study can be used in subsequent soybean breeding for improved oil composition traits.

## **W125: Bioenergy Grass Biology: Genomes, microbiomes, traits and applications**

### **Exploiting the Plant Microbiome to Boost Plant Growth and Resilience**

**Kerrie Farrar**, IBERS, Aberystwyth University, Aberystwyth, United Kingdom

Plants have evolved with diverse bacterial endophytes, and their collective metagenome, within their tissues. These populations are complex and play roles in plant growth and resilience via phytohormone production, nutrient acquisition, and modulation of plant pathways such as stress responses. Nutrients and water are key resources limiting world agricultural production. Nitrogen fertilizers, provided by the energy-demanding Haber-Bosch process, have enhanced crop yields at huge environmental cost, whereas drought has traditionally been managed by irrigation, resulting in ground water depletion. Across the world, many farmers cannot afford to use fertilisers, and in developed countries, perennial energy crops such as *Miscanthus* must produce high annual biomass yields on low-quality land without environmentally costly inputs such as water, fertiliser or pesticides.

We have previously characterised the *Miscanthus* microbiome and demonstrated the importance of seed transmission (Cope Selby *et al.* 2017). More recently, we have targeted extreme environments (coastal-saline, heavy metal mine sites etc.) to isolate endophytic bacteria with plant growth potential and that confer stress resilience to the plant host. We have compared the capacity of these novel bacterial endophytes to improve plant growth under saline, and limited water and nitrogen conditions, using the model plant *Brachypodium distachyon* in the National Plant Phenomics Centre, Aberystwyth, UK. We have identified plant growth promoting (PGP) strains outperforming the well-characterised PGP strain *Azospirillum brasilense* Cd. Multi -omics analyses are underway to determine the plant and bacterial factors involved in the beneficial plant-endophyte interaction. We aim to apply novel PGP endophyte strains harbouring beneficial traits to improve plant performance, including energy crop production and phytoremediation applications.

## **W126: Bioenergy Grass Biology: Genomes, microbiomes, traits and applications**

### **Trials with Clonal *Miscanthus* Reveal Background Environmental x Management Variability in Perennial Energy Grasses**

**Emily Heaton**, Iowa State University, Ames, IA

Considerable effort has gone into improving perennial grasses in the Andropogonae tribe for biomass and bioenergy. Annoyingly for researchers, species in this tribe are obligate outcrossers and many have complex ploidy, meaning results observed in field trials likely represent a small subset of possible phenotypes because of huge genotypic variability. On top of that, little is known about the “background noise” in such field trials, especially in perennial species. How likely is an observed difference to be caused by changes in genotype vs changes in the environment or management? How will an observed phenotype change with plant age? To develop a baseline response to these questions, we used a staggered-start experimental design to study age-related productivity and phenology changes during the establishment phase (first three years) of clonal *Miscanthus* × *giganteus*. This unique experimental design removed genetic variability and repeated the planting year, thus allowing independent estimation

of growing season vs plant age effects without genetic variability. We also evaluated response to N fertilizer as plants aged. We found that two- and three-year-old (mature) plants produced 30% more stems, with 20% more leaves and nodes than one-year-old (young) plants. Faster developmental rates were usually seen in young stands, but they did not lead to more advanced developmental stages. Normalized over thermal time (growing degree days), older stands emerged ~3 months earlier than newly planted rhizomes. Nitrogen fertilization partially overrode age-related changes in emergence and senescence, and delayed flowering in mature stands, thereby extending the growing season at least 10 days. Our results support accounting for natural background changes in phenology and productivity with plant age and fertility when interpreting performance of new perennial grass phenotypes.

### **W127: Bioenergy Grass Biology: Genomes, microbiomes, traits and applications**

#### **Sorghum *Dw2* (AGCVIII kinase) Regulates Lipid Signaling, Endomembrane Function, and Stem Growth**

Joel Oliver, Texas A&M University, College Station, TX

Stems of the C4 grass sorghum are comprised of nodes and internodes that are produced and elongated sequentially during plant growth impacting sink strength, biomass yield and composition. Stems of high biomass bioenergy sorghum can accumulate up to 50 internodes that together span 4-5 meters and account for ~84% of harvested biomass. An AGCVIII kinase (*Dw2*) has been identified that regulates sorghum stem internode growth, but the underlying signaling network is unknown. Here we provide evidence that mutation of *Dw2* reduces cell proliferation in internode intercalary meristems, inhibits endocytosis, and alters the distribution of heteroxylan and mixed linkage glucan in cell walls.

Phosphoproteomic analysis showed that *Dw2* signaling modulates the phosphorylation of proteins involved in lipid signaling (PLD $\delta$ ), endomembrane trafficking, hormone, light and receptor signaling, and photosynthesis. Together, our results show that *Dw2* modulates endomembrane function and cell division during sorghum internode elongation providing insight into the regulation of monocot internode development

### **W128: Bioenergy Grass Biology: Genomes, microbiomes, traits and applications**

#### **CROPSR, a CRISPR Design Solution for Biofuel Crops**

Matthew E. Hudson, University of Illinois, Urbana, IL

CROPSR software, shortly due for public release, allows scientists to design CRISPR/Cas9 guide RNAs, and primers needed to validate editing events. The application is designed to provide unique and specific guide RNAs and validation primers even in complex, paleopolyploid crop genomes such as those of biofeedstocks. CROPSR consists of a computational “back end” that creates a database of specific guides that target each annotated gene. It also designs specific validation primers for each guide, and stores this data in a database for multiple whole genomes. A user interface allows users to see the guides and primers for their specific gene or pathway, arranged by location within the gene, promoter and other annotated features. The software is maintained via a GitHub repository, which will also be used for public distribution, and is currently being tested by select users.

### **W129: Bioenergy Grass Biology: Genomes, microbiomes, traits and applications**

#### **Comprehensive Above and below Ground 3D Phenomics for Grasses**

Mao Li, Donald Danforth Plant Science Center, St. Louis, MO

Grasses are the most economically important taxon of plants. However, they contain substantial morphological diversity and complexity, bringing challenges of comprehensive and precise quantification. For example, most features of inflorescence or root architecture are not well-measured manually or with standard 2D imaging. Therefore, the combination of new technologies and computational tools is crucial for 3D in-depth phenotyping. In this talk, I will give two examples of advanced 3D phenomics both above and below ground. 1) We used X-ray computed tomography and developed custom computational pipelines to accurately phenotype various plant structures with both general and organ-specific traits in a

3D framework. I will show one application of a sorghum panicle and root crown study from a subset of the diverse Sorghum Association Panel where genotype to phenotype relationships were evaluated. 2) We developed large mesocosm growth systems and combined with photogrammetry methods to facilitate the 3D visualization and quantification of large, freely-grown root systems in their entirety. Meanwhile, sensor arrays were buried to quantify and model environmental fluxes. We applied this approach for two genotypes of switchgrass at two different water stress conditions. These two 3D phenomics enable comprehensive characterization of above and below ground features, and are not limited to these two examples, but to many other experimentations and complex structures.

### **W130: Bioenergy Grass Biology: Genomes, microbiomes, traits and applications**

#### **Transcriptome Analysis of a Mutualistic Association between Perennial Ryegrass (*Lolium perenne*) and Its Fungal Endophyte *Epichloë festucae***

**Flavia Pilar Forte**<sup>1</sup>, Istvan Nagy<sup>1</sup>, Jan Schmid<sup>2</sup>, Paul Dijkwel<sup>2</sup>, David E. Hume<sup>3</sup>, Richard D. Johnson<sup>3</sup>, Wayne R. Simpson<sup>3</sup> and Torben Asp<sup>1</sup>, (1)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, (2)School of Fundamental Sciences, Massey University, New Zealand, (3)AgResearch, Grasslands Research Centre, New Zealand

*Epichloë festucae* var. *lolii* is a seed transmitted fungal endophyte, which forms a mutualistic symbiosis with perennial ryegrass (*Lolium perenne*) conferring plant protection against biotic and abiotic stresses. When forming synthetic symbioses the genetics of both the host plant and the fungal symbiont are critical to the success of the symbiosis. Currently the mechanisms contributing to compatible and sustainable synthetic associations, over time, are mostly unknown. We studied endophyte retention in a ryegrass-endophyte selection program. We performed transcriptome profiling by RNA-Seq, of both the host and its symbiont, comparing three generations (G2, G6, and G9) of endophyte-colonised (E+) and endophyte-free (E-) plants, grown under controlled conditions. Principal component analysis (PCA) revealed that E- hosts are genetically more diverse than E+ ones, which share a similar genetic makeup. Plant population structure, based on single-nucleotide polymorphisms (SNPs), revealed a more pronounced genetic difference in G9 E- plants compared to G2 and G6. One-hundred-ninety-three differentially expressed genes (DEGs) were identified between E+ and E- plants comprising two major sub-clusters. Protein families (Pfam database entries) designated to plant non-self-recognition and stress response (LLR, NB-ARC, and jacalin-like domain), chloroplast development (Sel1-like repeat and ankyrin repeats), and signal transduction (kinases) were overrepresented in E+. Conversely, proteins usually involved in plant defence response and plant hormone transport (ABC/ABC-2-type transporter, trypsin-like peptidase domain, PDR-ABC-transporter associated and kinases) were underrepresented in E+. Moreover, downregulation of genes involved in plant cell wall synthesis was observed. The endophyte transcriptome did not change through generations.

Our results suggest that the presence of *Epichloë festucae* var. *lolii* triggers early-stage plant defence mechanisms and pathways related to stress response. This possibly primes the host, but simultaneously negatively regulates other defence pathways, primarily related to plant hormones, which may regulate or maintain compatibility.

*“This project has received funding from the European Union’s Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement No. 676480”.*

### **W131: Bioinformatics**

#### **A Novel Approach for Assembly of High-Fidelity PacBio HiFi Reads Resolves Large Segmental Duplications and Haplotypes**

**Sergey Koren**<sup>1</sup>, Sergey Nurk<sup>1</sup>, Arang Rhie<sup>1</sup>, Brian P Walenz<sup>1</sup>, Robert Grothe<sup>2</sup> and Adam M. Phillippy<sup>1</sup>, (1)Genome Informatics Section, NHGRI, NIH, Bethesda, MD, (2)Pacific Biosciences, Menlo Park, CA

A complete and accurate genome sequence forms the basis of all downstream genomic analyses. Traditionally, assembly projects have had to choose between high-accuracy short sequences (such as

Illumina) or noisy long sequences (such as Oxford Nanopore Technologies (ONT) or Pacific Biosciences (PacBio)). A new data type introduced by PacBio, called HiFi, aims to combine the best of both worlds. HiFi reads exceed 10kb in length and are over 99% accurate. We present a novel assembly algorithm tailored for this data type. Run length encoding (RLE), which collapses homopolymers to a single base, can be used to eliminate the majority (>90%) of HiFi errors. Further overlap-based detection and correction of erroneous regions allows assembly at an extremely stringent overlap criteria, less than one error in 10 Kbp. This approach both increases assembly continuity and resolves repeats collapsed by current approaches. The assembly of the effectively haploid CHM13hTERT cell line NG50 exceeds 70 Mbp and resolves 95.6% of segmental duplication BACs at Q40+, exceeding the current reference (85.6% for GRCh38) and ultra-long ONT assemblies (Canu 93.5%, Flye 74.2%, Shasta 44.3%, Redbean 17.6%). On diploid human genomes, our assemblies exceed 50mb NG50s, capture more complete haplotypes, and improve SNP recall over state of the art methods. An initial implementation is available within Canu v1.9 and we expect continuing improvements in assembly quality as the data and algorithms mature.

### **W132: Bioinformatics**

#### **Long-Read Assembly: Are Metagenomes Easier than Diploid Genomes?**

**Mikhail Kolmogorov**, UC San Diego, La Jolla, CA, **Mikhail Rayko**, Saint Petersburg State University, Saint-Petersburg, Russian Federation, **Jeffrey Yuan**, University of California San Diego, La Jolla, CA, **Evgeny Pevnikov**, ITMO university, Russian Federation and **Pavel Pevzner**, Computer Science Dept., UCSD, La Jolla, CA

Long-read sequencing technologies have substantially improved the assemblies of many isolate bacterial genomes as compared to the fragmented assemblies produced from short-read technologies. However, assembling complex metagenomic datasets remains a challenge even for state-of-the-art long-read assemblers. To address this gap, we present the metaFlye assembler and demonstrate that it generates contiguous and accurate assemblies for metagenomic datasets. We benchmark metaFlye on mock dataset as well as real metagenomic sequencing of cow and sheep microbiomes, and demonstrate that long-read metagenomics can reconstruct many complete bacteria from complex samples. We also show that variations between closely-related strains and species present an additional challenge in metagenome assembly, not unlike assembling a complex diploid genome

### **W133: Bioinformatics**

#### **Speeding up Genome Assembly with Sparse and Hierarchical Minimizer Indices**

**Jason Chin**, DNAnexus, Mountain View, CA and **Asif Khalak**, Foundation of Biological Data Science

The advance of long-read sequencing technologies has significantly improved the visibility for large plant and animal genomes through genome assemblies. The long reads generated by sequencing platforms such as Oxford Nanopore Technologies and Pacific Biosciences allow for resolving longer repeats even with a higher error rate at ~10%. There have been recent algorithms introduced to make such assemblies which need extensive computing resources possible with a computing cluster or equivalent cloud resources. Meanwhile, recent advances in SMRT sequencing technology have enabled both long and highly accurate reads (>10kb read length and < 1% error rate) through consensus algorithms from single-molecule reads.

We propose a new algorithmic approach, using a Sparse and Hierarchical MiniMizER (SHIMMER) index, for indexing reads. The SHIMMER index alleviates the computation burden for assembling longer and more accurate reads. The Peregrine assembler, utilizing the SHIMMER index, can assemble 30x human genome (average length ~15kb, error rate ~ 0.5%) within an unprecedented speed (30 CPU hours/two wall-clock hours) on a single computation node. It is more than ten times faster and uses much less computational resources than any other currently available assemblers for human genome assemblies from the same data.

Most animal, insect, and plant genomes contain abundant repetitive sequences, far beyond the level observed in the human genome. It presents unique and stringent challenges to genome assembly. The assembler parameters that are optimized for human genome assembly will need adjustments to adapt to highly repetitive genomes. We evaluate the Peregrine Genome Assembler on some recently released plant genome datasets to understand how to optimize the parameters for the best performance on assembling plant and animal genomes other than the human genome with the Peregrine Genome Assembler.

#### **W134: Bioinformatics**

##### **Beyond Binning: Combining Proximity-Ligation and Metagenomic Deconvolution Algorithms Significantly Improves the Quality and Quantity of Genomes Extracted from Complex Microbiome Samples**

Ivan Liachko, Phase Genomics, Seattle, WA

#### **W135: Bioinformatics**

##### **Ubiquitous Mutational Footprints of AID/APOBEC Deaminases in Human Cancer Genomes: Application of Nucleotide Weight Matrices**

Igor Rogozin, National Institutes of Health, Bethesda, MD

Cancer genomes accumulate nucleotide sequence variations that number in the tens of thousands per genome. A prominent fraction of these mutations is thought to arise as a consequence of the off-target activity of DNA/RNA editing cytosine deaminases. These enzymes, collectively called activation induced deaminase (AID)/APOBECs, deaminate cytosines located within defined DNA sequence contexts. The resulting changes of the original C:G pair in these contexts (mutational signatures) provide indirect evidence for the participation of specific cytosine deaminases in a given cancer type. The conventional method used for the analysis of mutable motifs is the consensus approach. Here, for the first time, we have adopted the frequently used weight matrix (sequence profile) approach for the analysis of mutagenesis and provide evidence for this method being a more precise descriptor of mutations than the sequence consensus approach. We confirm that while mutational footprints of APOBEC1, APOBEC3A, APOBEC3B, and APOBEC3G are prominent in many cancers, mutable motifs characteristic of the action of the humoral immune response somatic hypermutation enzyme, AID, are the most widespread feature of somatic mutation spectra attributable to deaminases in cancer genomes. Overall, the weight matrix approach reveals that somatic mutations are significantly associated with at least one AID/APOBEC mutable motif in all studied cancers.

#### **W136: Bioinformatics**

##### **Eukaryotic Gene Prediction by Genemark-EP+ with Support from Homologous Proteins**

Alexandre Lomsadze, Tomas Bruna and Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA

One of long-standing and difficult to reach goals of computational genomics is development of an accurate and fast algorithm for gene prediction in eukaryotic genomes. We present a new automatic gene finding tool, GeneMark-EP+, that integrates footprints (hints) of spliced aligned to genome homologous proteins into iterative unsupervised process of model training and gene prediction. A novel specialized pipeline, ProtHint, generates high specificity hints for positions of intron borders as well as translation starts and stops by mapping homologous proteins from multiple species. For example, for *Arabidopsis thaliana*, with proteins of species outside *A. thaliana* taxonomical order, ProtHint retrieves 70% of all introns with 99% specificity. GeneMark-EP+ uses the hints to improve estimation of model parameters as well as to directly adjust gene prediction with guidance of the most reliable hints. We tested GeneMark-EP+ on genomes of both model and non-model organisms. The results demonstrated improvements in prediction of exon-intron structures, particularly in large eukaryotic genomes, even when only proteins from remote species (outside the species phylum) are available.

Project location: [github.com/gatech-genemark/GeneMark-EP-plus](https://github.com/gatech-genemark/GeneMark-EP-plus)

**W137: Biotech 101 – a.k.a. Navigating the U.S. Regulatory System for Products of Biotechnology**

**Neil Hoffman: USDA**

Neil Hoffman, U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Biotechnology Regulatory Services, Riverdale, MD

**W138: Biotech 101 – a.k.a. Navigating the U.S. Regulatory System for Products of Biotechnology**

**Mike Mendelsohn: EPA**

Mike Mendelsohn, U.S. Environmental Protection Agency, Washington, DC

**W139: Biotech 101 – a.k.a. Navigating the U.S. Regulatory System for Products of Biotechnology**

**FDA Oversight of Food from New Plant Varieties**

**Patrick Cournoyer**, U.S. Food and Drug Administration, College Park, MD

FDA regulates food from plants, including plants developed with biotechnology. For over 20 years, FDA has offered the Plant Biotechnology Consultation Program to help developers of new plant varieties ensure the food safety and regulatory compliance of food from new plant varieties before they enter the food supply. In 2019, FDA completed its evaluation of a genome edited high-oleic soybean, marking the first completed consultation on food from a genome edited new plant variety. The agency intends to develop guidance explaining how the current regulatory policy for foods from new plant varieties applies to foods from genome edited plants. In addition, FDA leads the Biotechnology Education and Outreach Initiative, an interagency effort for consumer outreach and education that will release science-based outreach materials on agricultural biotechnology.

**W140: Brachypodium Genomics**

**Extending the Genotype in Brachypodium by including DNA Methylation Reveals a Joint Contribution with Genetics on Adaptive Traits**

**Justin O. Borevitz**, The Australian National University, Canberra, ACT, Australia and Steve Eichten, ANU, Australia

Epigenomic changes have been considered a potential missing link underlying phenotypic variation in quantitative traits but is potentially confounded with the underlying DNA sequence variation. Although the concept of epigenetic inheritance has been discussed in depth, there have been few studies attempting to directly dissect the amount of epigenomic variation within inbred natural populations while also accounting for genetic diversity. By using known genetic relationships between Brachypodium lines, multiple sets of nearly identical accession families were selected for phenotypic studies and DNA methylome profiling to investigate the dual role of (epi)genetics under simulated natural seasonal climate conditions. Despite reduced genetic diversity, appreciable phenotypic variation was still observable in the measured traits (height, leaf width and length, tiller count, flowering time, ear count) between as well as within the inbred accessions. However, with reduced genetic diversity there was diminished variation in DNA methylation within families. Mixed-effects linear modelling revealed large genetic differences between families and a minor contribution of epigenomic variation on phenotypic variation in select traits. Taken together, this analysis suggests a limited but significant contribution of DNA methylation towards heritable phenotypic variation relative to genetic differences.

**W141: Brachypodium Genomics**

**From *Brachypodium* to *Ecteiocolea*: The Origin, Expansion, and Diversity of the NLR Integrated Domains in the Grasses**

**Samuel Holden**<sup>1</sup>, James Russell<sup>2</sup>, Helen J. Brabham<sup>1</sup>, Inmaculada Hernández-Pinzón<sup>1</sup>, Matthew Barrett<sup>3</sup> and Matthew J. Moscou<sup>1</sup>, (1)The Sainsbury Laboratory, Norwich, United Kingdom, (2)The Sainsbury Laboratory, United Kingdom, (3)Botanic Gardens and Parks Authority, Australia

Maize, wheat, sorghum, millet, and rice are important staple crops worldwide. Comparative genomics of these crop species and wild grass species such as *Brachypodium distachyon* (Pooideae) and *Streptochaeta angustifolia* (Anomochlooideae) found substantial conservation in gene content. These species all belong to the Poaceae (grasses), a family that emerged ~120 Mya in the Early Cretaceous. During the evolution and divergence of the Poaceae from other Poales families, several genetic innovations occurred that potentially contributed to the ecological competitiveness of the family worldwide. Known genetic innovations include cytoplasmic ADP-glucose synthesis, spikelet morphology and secondary metabolite diversity. In this work, we set out to study two potential Poaceae-specific genetic innovations in plant immunity: (1) the presence of diverse integrated domains (ID) within nucleotide-binding, leucine-rich repeat (NLR) immune receptors and (2) the Exo70FX gene family, a clade of Exo70 genes unique to the Poaceae, which contribute to the trafficking of molecules to the cell periphery (exocytosis). A major challenge when establishing the evolutionary history of genetic innovations is identifying outgroup species that are related, but distant to the species of interest. Using RNAseq analysis of over 50 species in the Poales, we identified *Ecdeiocolea monostachya* (Ecdeiocoleaceae) as a critical outgroup to the Poaceae without traces of these two genetic innovations. We sequenced the genome of *E. monostachya* in order to provide a robust genomic resource for this critical outgroup in our analyses. In previous work using diverse Poaceae species, the major integrated clade (MIC1) was found to be enriched with NLRs with diverse ID. MIC1 members include the rice resistance genes *Pi-ta* (NLR-TRX) and *RGA5* (NLR-HMA), and barley *Rpg5* (NLR-Kinase). We set out to investigate the evolutionary origin, expansion, and diversity of NLRs within MIC1. We developed a bioinformatic pipeline to use publicly available RNAseq data to assemble and annotate NLRs with ID. We annotated 269 accessions from 115 species in the Poales and identified 62,366 full length NLRs with 4,455 NLR-ID. These integrated domains are not random; the majority are domains known to be involved in immunity including transcription factors (WRKY), enzymes (thioredoxins), protein kinases, phosphatases (PP2C), heavy metal associated (HMA), and exocytosis (Exo70). Analysis of Exo70 genes across diverse Poales species using publicly available RNAseq and genomic data found that all NLR-ID Exo70 domains are derived from Exo70F and Exo70FX clades. The integration of exogenous domains in MIC1 clade members emerged after speciation of *S. angustifolia* but prior to the radiation of BOP and PACMAD grasses. We also identified Exo70FX genes in the *Ecdeiocolea monostachya* genome, indicating that the Exo70FX family originated prior to the radiation of the Poaceae. However, the family experienced extensive proliferation within the Poaceae, becoming the largest and most sequence-diverse of all Exo70 families observed to date; with extensive species-specific subfamily expansion and contraction.

## **W142: Brachypodium Genomics**

### **Genome Assembly Variation and Time-Series Association in *Brachypodium* Species**

**Jared Streich**<sup>1</sup>, Jonathan Romero<sup>2</sup>, Ashley Cliff<sup>2</sup>, David Kainer<sup>1</sup> and Daniel Jacobson<sup>1</sup>, (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, TN

*Brachypodium distachyon* and other *Brachypodium* species have been ideal model grass organisms, but currently lack a mitochondrial genome. Plant mitochondrial genomes are challenging to assemble due to a long inverted repeat. We used the closely related and fully assembled wheat mitochondrial genome *Triticum aestivum* as a backdrop for creating the first *B. sylvaticum* MTgenome using PacBio long-reads. First we assembled BWA mappable reads into contigs using the SPADes Assembler creating the first intermediate *Brachypodium* MTgenome. Contigs from this assembly were then randomly broken down into fastq files of varying length from 100-300bp and used to call SNP and indel variants on the *Triticum aestivum* genome. Variants were then imputed onto the wheat MTgenome using GATK's 'FastAlternateReferenceMaker' function. After several rounds of polishing through contig assembly, contigs to short reads, variant calling, and variant imputation, a high quality *Brachypodium sylvaticum* MTgenome assembly was created. Gene models were then run using Prokka to annotate the

MTgenome. A similar process was used to create MTgenomes for several accessions of other *Brachypodium* species including 121 *B. distachyon*, 2 *B. stacei*, and 2 *B. hybridum* using the *B. sylvaticum* MTgenome. The final assemblies were then used for trans-chromosomal linkage using Custom Correlation Coefficient (CCC), environmental association (GWATS), and demography analysis.

### **W143: *Brachypodium* Genomics**

#### **The Role of the BUZZ Kinase in Post-Initiation Root Hair Growth**

**Karen Sanguinet**<sup>1</sup>, Thiel A Lehman<sup>2</sup>, Rhoda AT Brew-Appiah<sup>2</sup>, Miguel Rosas<sup>2</sup>, Ying Wu<sup>3</sup>, Zara B York<sup>2</sup>, Ping Zheng<sup>1</sup>, Dorrie Main<sup>4</sup>, Tobias I Baskin<sup>5</sup>, Andrei Smertenko<sup>6</sup> and Eric H. Roalson<sup>7</sup>, (1)Washington State University, Pullman, WA, (2)WSU, (3)Northeast Normal University, Changchun, China, (4)Department of Horticulture, Washington State University, Pullman, WA, (5)University of Massachusetts, Amherst, MA, (6)Institute of Biological Chemistry-WSU, Pullman, WA, (7)School of Biological Sciences, Washington State University, Pullman, WA

Cyclin-dependent kinases (CDKs) belong to a large superfamily of homologous serine-threonine protein kinases that share conserved catalytic domains. CDKs in association with their activating cyclin partners are well known for their regulatory role in cell cycle progression. Although eukaryotic CDKs and cyclins are highly conserved and display a high degree of homology, the functionality of these proteins and their regulatory role in plants can differ from those in animals and yeast. We identified a recessive, root hairless mutant (*buzz*) with longer roots in the model temperate grass, *Brachypodium distachyon*, with no effects on shoot growth. Root growth and root hair development are essential for efficient water and nutrient acquisition. *BUZZ* was cloned using a next generation sequencing (NGS) approach and found to encode a putative CDK. The *buzz* allele harbors a G-to-A base pair transition resulting in an amino acid substitution of a conserved glycine to arginine in the catalytic domain, which leads to a functional kinase null as confirmed by a second T-DNA allele. *BUZZ* is expressed at relatively low levels in the root tip. Phylogenetic, RNAseq, and physiological studies were used to further characterize and provide insight into the function of the *BUZZ* kinase. Biochemical analyses of the kinase suggest that *BUZZ* likely requires its cognate cyclin for kinase activity. Moreover, *in silico* data coupled with expression analysis suggest that *BUZZ* may be interact with the RNA processing machinery. Future research is aimed at uncovering *BUZZ* interactors and substrates.

### **W144: *Brachypodium* Genomics**

#### **Developing EcoFABs as a Reproducible System to Study the Root Microbiome**

**John Vogel**<sup>1</sup>, Peter Andeer<sup>2</sup>, Kateryna Zhalnina<sup>3</sup> and Trent Northen<sup>3</sup>, (1)DOE Joint Genome Institute, Berkeley, CA, (2)Lawrence Berkeley National Laboratory, (3)Lawrence Berkeley National Laboratory, Berkeley, CA

In recent years the microbiome has been shown to play a major role in plant and animal health and productivity. However, despite its importance, we know little about how plants and microbes interact to shape microbiome composition and function. The lack of reproducible model ecosystems is a limitation to our ability to study and understand the microbiome. Currently, most studies either focus on a single isolate and a single plant growing in a petri dish or use they use 16s or ITS amplicon sequencing to survey microbial abundance in field samples. While the former approach allows researchers to control and manipulate nearly all aspects of the system, the lack of diverse microbes creates a highly artificial system that is often not relevant to agriculture. The latter approach captures much of the complexity of what goes on in the field, but it is impossible to precisely replicate and control the microbial community in the field. Our efforts are aimed at developing a middle ground where we use a small, defined microbial communities in a fully controlled ecosystem. To enable this, we developed small growth containers (EcoFABs) that allow us to non-destructively monitor root and microbial growth, sample the growth medium for metabolomics all while avoiding contamination from environmental microbes. Progress towards increasing throughput using mass production and robotic handling will be presented.

## **W145: Brassicas**

### **Genetic Architecture of Glucosinolate Variation in *Brassica napus***

**Varanya Kittipol**, University of York, York, United Kingdom

Glucosinolates (GSLs) are a group of secondary metabolites prevalent in the important oilseed rape crop (*Brassica napus* L.). The GSL hydrolysis products show diverse bioactivities and thus play significant biological and economical roles in the defence system and nutritional qualities of rapeseed protein meal. Hence, there is an increasing desire to harness the defensive properties of GSLs to improve pest resistance properties in the vegetative tissues while maintaining low GSLs in the seeds for animal feed.

This work aims to identify the genetic controls underlying natural GSL variations in the leaves and roots of *B. napus*, and also develop understanding of their connections with seed GSLs. To address these aims, Associative Transcriptomics (AT), was performed on a panel of 288 *B. napus* accessions. AT correlates GSL trait variations to the variations in either gene sequences or gene expression across these accessions to identify highly associated quantitative trait loci for GSL contents.

Single nucleotide polymorphism and gene expression marker associations identify the *MYB28/HAG1* orthologues on chromosomes A9 and C2 as key regulators for aliphatic GSLs in leaves. We show that the positive correlation observed between aliphatic GSLs in seed and leaf is due to the amount synthesised, as controlled by *Bna.HAG1.A9* and *Bna.HAG1.C2*, rather than by variation in the transport processes. In addition, AT and differential expression analysis in root tissues implicate *Bna.HAG3.A3*, an orthologue of *MYB29/HAG3*, as the main controlling factor for root aromatic GSL variations. This work improves our understanding of the genetic regulatory of GSL natural variations in *B. napus* that could lead to crop improvement.

## **W146: Brassicas**

### **Can Resynthesized Rapeseed be Genomically Stable?**

**Elizabeth Ihien**, Department of plant breeding Justus Liebig University, Giessen, Germany

Rapeseed (*Brassica napus*, AACC) is a young allotetraploid species formed by the hybridization of *Brassica rapa* (AA) and *Brassica oleracea* (CC). However, resynthesized *B. napus* lines are often unstable and infertile, unlike natural *B. napus*. Meiotic stability in natural *B. napus* may have arisen through allele inheritance from the progenitor species or via one or more de novo mutations post-polyploidisation. We tested these hypotheses by characterizing a diverse set of resynthesized *B. napus* lines for chromosome rearrangements, allele inheritance, fertility, and meiotic behaviour. SNP genotyping was performed using the Illumina Infinium *Brassica* 60K array, and allele copy number used to infer translocation events between the A and C genomes. Approximately 52% of lines (91/174) with SNP genotyping information were homozygous, as expected; cross-contamination resulting in heterozygosity was common in older lines maintained for many generations in the field. Self-pollinated seed-set (average 611, range 0 – 3876 per plant) and genome stability (number of copy number variants) were significantly affected by the interaction between both *B. rapa* and *B. oleracea* parental genotypes. Most lines (94%) showed evidence of unbalanced translocations between the A and the C genomes, where loss of one homoeologous region was not balanced by the presence of an extra copy of the other homoeologous region. Our results show that some resynthesized lines are more stable and fertile than others, and support the hypothesis that allelic variants inherited from parental genotypes affect genome stability in synthetic rapeseed.

## **W147: Brassicas**

### **Replaying the Evolutionary Tape to Investigate Subgenome Dominance in Neoallopolyploid *Brassica napus***

**Patrick Edger**, Department of Horticulture, Michigan State University, East Lansing, MI

Interspecific hybridization merges evolutionarily distinct parental genomes (subgenomes) into a single nucleus. A frequent observation is that one subgenome is "dominant" over the other subgenome(s) including higher gene expression levels. Which subgenome becomes dominantly expressed in allopolyploids remains poorly understood. Here we "replayed the evolutionary tape" with six isogenic resynthesized allopolyploid *Brassica napus* lines and investigated subgenome dominance patterns over the first ten generations. We found that the same parental subgenome was consistently more dominantly expressed in all lines and generations. This suggests that subgenome expression dominance is largely predetermined based on differences among the parental diploid genomes. These findings have major implications regarding the genotypic and phenotypic diversity observed following plant hybridization in both ecological and agricultural contexts.

#### **W148: Brassicas**

##### **New Genomic Resource in Watercress (*Nasturtium officinale*, R.) Enabling Molecular Breeding for Improved Anti-Cancer Properties, Flavour and Nutrient Density**

**Gail Taylor**<sup>1</sup>, Annabelle Damerum<sup>1</sup>, Nikol Voutsina<sup>2</sup>, Daniel Kliebenstien<sup>3</sup>, Mark Coldwell<sup>2</sup>, Graham Packham<sup>2</sup>, Mark Chapman<sup>2</sup>, Tingting Zhu<sup>3</sup> and Ming-Cheng Lou<sup>3</sup>, (1)University of California, Davis, CA, (2)University of Southampton, United Kingdom, (3)University of California, Davis

Watercress (*Nasturtium officinale*, R.) is a highly nutritious leafy green vegetable that belongs to the mustard family and is ideal for indoor vertical agricultural systems, since naturally, it grows in an aquatic or semi-aquatic habitat. Consumption is linked to reduced inflammation and chronic damage in cells and has been shown to have a role in cancer prevention *in vitro*. These benefits are derived from the high concentrations of secondary metabolites found in the crop, including glucosinolates. Despite this evidence, watercress remains largely undeveloped with limited breeding resources and no active breeding programmes worldwide.

We describe the first 259 F<sub>2</sub> progeny mapping population, the use of reduced representation Genotyping-by-sequencing (GBS) for marker discovery and the construction of the first genetic linkage map. At the same time, we generated sequence assembly from whole genome PacBio SMRT reads and used a Bionano optical map to validate the PacBio assembly, perform hybrid scaffolding and ultimately produce 16 pseudomolecules of the watercress genome. We identified novel quantitative trait loci (QTL) in two separate experiments in contrasting controlled-environment and field conditions. QTL were identified for plant morphology and nutritional composition traits, including total antioxidant capacity (AO) and glucosinolate content. The health-promoting properties of watercress were further investigated by assaying the ability of watercress sap to inhibit growth of human breast cancer cells, where we identified three QTL. This work represents a significant step forward for this highly nutrient dense, leafy crop and will facilitate breeding, targeted on improved traits for growth and anti-cancer nutrition.

#### **W149: Brassicas**

##### **Influence of Selection on Diversity in *Brassica oleracea***

**Sarah Turner**<sup>1</sup>, Makenzie E. Mabry<sup>2</sup>, J. Chris Pires<sup>2</sup>, Jeffrey Ross-Ibarra<sup>3</sup> and Timothy M. Beissinger<sup>4</sup>, (1)Center for Population Biology, University of California-Davis, Davis, CA, (2)Division of Biological Sciences, University of Missouri, Columbia, MO, (3)University of California, Davis, CA, (4)Georg-August-Universität Göttingen, Division of Plant Breeding Methodology, Germany

Detecting evidence of selection on polygenic traits is useful for understanding crop origins, patterns of adaptation, and the potential for continued improvement. The crop species *Brassica oleracea* is a striking example of the power of selection, which has acted on different plant organs to produce phenotypically diverse crops such as kale, cabbage, broccoli, cauliflower, Brussels sprouts, and kohlrabi. In the interest of investigating population structure and genetic differentiation among these crop types, whole genome resequencing was performed on 172-accession diversity panel, which included 13 morphotypes of *B. oleracea* and its wild relatives. This genotype data was combined with publicly available resequencing data for an additional 119 accessions, resulting in a total of 3,266,067 high-quality SNPs that were used

to characterize population structure and to test for evidence of polygenic selection. Phenotypic data included traits related to leaf morphology and metabolite composition, which are thought to have played an important role during the domestication of *B. oleracea* to improve yield and palatability. For each trait, we assessed evidence of selection using the G-hat method, which tests for a relationship between changes in allele frequency and effect size estimates at every genotyped locus. Using this approach, we present evidence for selection during the recent evolutionary history and diversification of *B. oleracea*.

## **W150: Brassicas**

### **Genome Sequencing of the Ethiopian Mustard (*Brassica carinata*): Investigation of the Genetic Evolution and Diversity for Agronomic Traits**

**Won Yim**, Mia L. Swain, David Curdie, John C. Cushman and Jeff F. Harper, University of Nevada - Reno, Reno, NV

This *Brassica carinata* reference genome can be used in combination with the reference genomes of its genetic progenitors as a comparative genomics tool to identify the loci of advantageous traits for the purpose of breeding crops which are better adapted to our changing climate. The produced genomic information will also further shed light on the genetic effects of climate change on *B. carinata* which have hardly been characterized. The advent of next-generation sequencing has led to the rapid publication of increasing numbers of reference genomes, but the short reads typical of next-generation sequencing aren't adequate for the *de novo* genome assembly of allopolyploid species like *B. carinata* due to their large genome size, high heterozygosity, and long spans of repetitive sequences. Recent genomic and bioinformatic tools, including long-read single-molecule sequencing have alleviated this complexity by reducing the number of unresolved repeats, better distinguishing paralogs from allelic variation, and improving assembly phasing. The TrioCanu trio binning method further exploits the benefits of long-read sequencing by partitioning the long reads from a diploid or polyploid offspring into haplotypes by reads from its genetics progenitors. These haplotypes are then able to be assembled independently without the interference of inter-haplotype variation. By resolving allelic variation before assembly, a set of complete, linear haplotypes for a genome are produced rather than a collapsed and artificial consensus representation of that genome. *Brassica carinata*, or Ethiopian Mustard, has traditionally been grown in Africa as a leafy vegetable, but has more recently become a commercially important oilseed crop grown as biodiesel feedstock. The seed has a high oil content, but because of its high erucic acid content, it is best utilized in industrial processes and is a promising alternative to petroleum-based lubricants, fuel, and plastics. *B. carinata* shares one genetic progenitor with *B. napus*, from which canola oil is derived from, but has more shatter-resistant siliques and can be grown on marginal lands in semi-arid regions and as a rotational crop with wheat or other cereal crops. In the past few years, massive amounts of *Brassica* genomic data have been generated from several large-scale initiatives revealing significant collinearity of the *Brassica* genomes, yet there are currently few genomic resources for *B. carinata*. This, along with the high genetic variation of *Brassica* species, abundance of morphotypes within the species, and their adaptations to various agronomic conditions, makes *Brassica* an excellent platform in comparative genomics and potential genetic donors for crop improvement.

## **W151: BreedBase and SGN: Genomics and Breeder Data Management and Analyses Tools.**

### **Introduction**

**Lukas A. Mueller**, Boyce Thompson Institute, Ithaca, NY

## **W152: BreedBase and SGN: Genomics and Breeder Data Management and Analyses Tools.**

### **Improving Quality in Africa RTB Crop Breeding Programs: A Database Management System Approach**

**Agbona Afolabi**<sup>1</sup>, Peteti Prasad<sup>2</sup>, Guillaume Bauchet<sup>3</sup>, Naama Menda<sup>4</sup>, Isaak Y. Teclé<sup>5</sup>, Olufemi Aina<sup>6</sup>, Elizabeth Yaa Parkes<sup>1</sup>, Lukas Mueller<sup>3</sup>, Ismail Rabbi<sup>1</sup>, Chiedozie Egesi<sup>6</sup> and Peter Kulakow<sup>1</sup>,

(1)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, (2)ICRISAT, Patancheru,

India, (3)BTI, Ithaca, NY, (4)Boyce Thompson Institute for Plant Research, Ithaca, NY, (5)Boyce Thompson Institute, Ithaca, NY, (6)IITA, Nigeria

High data quality, for both phenotype and genotype, is essential for the success of a breeding program. Breeding program generates large phenotypic and genotypic datasets that must be carefully managed through appropriate database management system (DBMS) to generate reliable and accurate information for decision-making. A DBMS is vital in data collection, storage, retrieval, and analysis in plant breeding programs to enhance the ultimate goal of increasing genetic gain. The International Institute of Tropical Agriculture (IITA), working on the root, tuber and banana (RTB) crops like cassava (<https://cassavabase.org/>), yam (<https://yambase.org/>), banana and plantain (<https://musabase.org/>) has deployed the use of BREEDBASE (<https://breedbase.org/>). The functionalities of these web-based databases in the area of genotypic and phenotypic data management and data analysis have been instrumental in achieving breeding goals. Such capabilities include ontology driven data management, statistical analyses, interfaces with Breeding API (BrAPI), barcodebased data collection using the PhenoApps (<http://phenoapps.org/>) like Fieldbook for phenotype data collection, Coordinate for genotype tissue sample collection and tracking, and Inventory for weighing samples without the need for data transcription have been developed to be user-friendly. Their wide acceptability among breeders has resulted in improvements in precision and quality of genotyping and phenotyping data, and as a result improvements in breeding programs goals. We will demonstrate the application of BREEDBASE in the IITA breeding program workflows.

Key words: RTB crops, BREEDBASE, PhenoApps, Genotyping, Phenotyping, Quality

### **W153: BreedBase and SGN: Genomics and Breeder Data Management and Analyses Tools.**

#### **BreedBase: Genomic Selection Analysis and Decision Support Tools**

**Isaak Y. Teele**, Boyce Thompson Institute, Ithaca, NY

BreedBase ([breedbase.org](http://breedbase.org)) is an informatics solution with integrated breeder data management and analysis tools. It has interactive interfaces to visualize and explore data, and perform analyses using mixed models. It provides analysis tools for genomic prediction, GWAS, population structure (PCA), selection index, correlation, clustering, and ANOVA. The analysis tools are available on all BreedBase powered databases such as <http://cassavabase.org>, <http://musabase.org>, <http://sweetpotatobase.org>, <http://yambase.org>, and <https://wheat.triticeaetoolbox.org/>. Breedbase eases breeders' access to their data, analysis and sharing, which ensures data integrity and improves efficiency. It is an open source software and adaptable to any breeding project. In this session, we will give an overview of the analysis tools.

### **W154: BreedBase and SGN: Genomics and Breeder Data Management and Analyses Tools.**

#### **Aerial Drone Imagery Database and Prediction of End-of-Season Traits in Maize**

**Nicolas Morales**, Boyce Thompson Institute, Ithaca, NY; Cornell University, Ithaca, NY

High-throughput phenotyping requires a well-structured and easy-to-use data management and processing pipeline. Critical to this effort is integration of ground-truth measurements, such as plant height, yield, and disease severity scores, and existing genotypic resources with the extracted image features. Furthermore, large community efforts, research projects, and breeding programs collecting image data ideally should leverage a standardized and centralized web-database for sharing and combining data. To meet these needs, we present ImageBreed for storing and processing drone images, both RGB and multi-spectral, into the BreedBase family of open-source web databases. BreedBase is used by dozens of plant breeding and genetics projects, including <https://cassavabase.org> and <https://solgenomics.net>, with plans to scale out to 50 to 100 USDA-ARS breeding programs through the Breeding Insight Platform (BIP). This system allows a researcher to login to the website, upload their field experiment information, upload their ground-truth phenotypic measurements, upload their drone image(s), stitch the aerial drone images into an ortho-mosaic if required, calculate vegetative indices

(NDVI, TGI, VARI), remove background soil from the image, define and save plot-level polygon images, calculate and save plot-level zonal-statistics phenotypes, and correlate those extracted phenotypes to the ground-truth phenotypic measurements. Also presented are algorithms for predicting end-of-season traits such as grain yield or grain moisture directly from plot-level aerial images using longitudinal convolutional neural networks (LSTM CNN); these algorithms are trained on imaging events spanning several years and locations. The system is fully operational at <http://imagebreed.org>.

### **W155: BreedBase and SGN: Genomics and Breeder Data Management and Analyses Tools.**

#### **Ricebase: A Breedbase Implementation for Rice with Tools for Gene Discovery and Marker Development**

**Jeremy Edwards**, Dale Bumpers National Rice Research Center, STUTTGART, AR

Ricebase is an integrative genomic database for rice that bridges the gap between the “big data” of rice genomics and the practical needs of rice breeders and researchers. There are numerous rice genomic data sets and published genotype-phenotype associations available to use in developing of marker assisted selection (MAS) tools for breeders. However, the amount of data and the organization of the data are obstacles that restrict one from making full use of these information resources. Finding and combining information from diverse sources for simultaneous evaluation is challenging because of varying data structures and a lack of a common coordinate system to facilitate comparison. Ricebase overcomes the information overload by curating information that is particularly relevant to rice breeding from multiple sources using a well-structured database schema and a common coordinate system based on the reference genome assembly.

The Ricebase implementation of BreedBase focuses on connections between germplasm, genetic diversity and molecular markers that are relevant to rice improvement. Data types stored include accession records, pedigrees, phenotypes, genotypes, and molecular markers. Ricebase uses the genome browser JBrowse to display the locations of various genomic features on chromosome coordinates from the MSU7/IRGSP-1.0 assembled rice genome as zoom-able and scroll-able tracks. These browser tracks include the sequence itself, annotated genes, SNPs, SSRs, and QTLs. SSR marker positions are displayed on genome browser tracks along with combined SNPs from the 700K HDRA and 3,000 rice genome resequencing studies, providing continuity between past and current marker technologies.

Ricebase has recently been updated to present the results of recent biparental and genome-wide association (GWA) mapping studies as tracks identifying genomic regions associated with the traits. Currently, genome browser tracks are available for biparental QTL mapping of chalk and fissure resistance, and GWA-QTL mapping of salt tolerance, cold tolerance, grain quality, and yield components. Genome browser track features are clickable to bring up links to additional information about the feature, relationships to other features, and database records. For example, SSRs link to accessions that are genotyped with the SSR and their allele states, and QTLs link to literature citations and reported candidate genes.

The curated phenotype-genotype associations highlight recent U.S. focused rice research and facilitate interpretation of new mapping studies in the context of previous research and meta-analyses across studies. The combined capabilities of Ricebase linking phenotypes with genetic markers, genome context, and allele states across diverse rice germplasm serve as a community resource for genetic discovery and breeding in rice.

### **W156: BreedBase and SGN: Genomics and Breeder Data Management and Analyses Tools.**

#### **A Symphony of *N. sylvestris* and *N. tomentosiformis* Sub-Genomes in the Allotetraploid Genome of *N. tabacum***

**Prashant S Hosmani** and Lukas A. Mueller, Boyce Thompson Institute, Ithaca, NY

*Nicotiana tabacum* is a large allotetraploid arising from interspecific hybridization of *N. tomentosiformis* and *N. sylvestris*. Characterization of tobacco genome is limited by its complexity and high repetitive content. Here we report a near-complete chromosome scale assembly of *Nicotiana tabacum* K326 using combination of short-read and long-read sequences with scaffolding based on chromosome conformation capture method. The considerable majority of the 4 Gb *N. tabacum* genome assembly is organized into 24 pseudomolecules corresponding to the linkage groups of the genetic map reported previously. We examined organization of two sub-genomes along with the dynamics of gene expression. Although recent in evolutionary origin, *N. tabacum* has gone under chromosomal reshuffling between the two sub-genomes. We observed that the smaller sub-genome, originated from *N. tomentosiformis* has significant expression bias compared to the *N. sylvestris* sub-genome across multiple tissues. Homoeologous biased genes also show a high level of divergence within co-expression networks, signifying neofunctionalization of homoeologous pairs in the allotetraploid *N. tabacum*.

## **W157: Buffalo Genomics**

### **Workshop Introduction**

**John Williams**, Davies Research Centre, University of Adelaide, Roseworthy, Australia

This year the Buffalo Genome Workshop has 5 presentations covering: A new genome assembly of the river buffalo; the link between genetic variation and levels of beta-carotene in buffalo and cow's milk; mapping QTL onto the buffalo genome; heat tolerance in buffalo; and, for the first time, a high quality genome for the wild African buffalo.

## **W158: Buffalo Genomics**

### ***De novo* Haplotype Phased Genome Assembly and Genomic Selection of Buffaloes in India**

**Ananthasayanam Sudhakar**<sup>1</sup>, Harish Kothandaraman<sup>2</sup>, Nilesh Nayee<sup>3</sup>, Sujit Saha<sup>3</sup>, Dushyant Singh Baghel<sup>4</sup>, Swapnil Gajjar<sup>3</sup>, Benjamin D. Rosen<sup>5</sup>, Curtis P Van Tassell<sup>6</sup> and Michael C. Schatz<sup>7</sup>, (1)National Dairy Development Board, Anand, India, (2)Nucleome Informatics Private Limited, Hyderabad, India, (3)National Dairy Development Board, India, (4)Nucleome Informatics Pvt Ltd, Hyderabad, Telangana, India, (5)Animal Genomics and Improvement Laboratory, USDA, (6)Animal Genomics and Improvement Laboratory, USDA, (7)Johns Hopkins University, Baltimore, MD

Here, we report developing a haplotype phased highly contiguous near complete genome assembly of *Bubalus bubalis* by separating haplotypes prior to assembly using a father-mother-offspring trio to accurately and completely reconstruct parental haplotypes.

Blood sample from a typical true to the breed Murrah heifer with known pedigree and normal Karyotype was collected and high molecular weight DNA was isolated. Sample was also collected from both the parents. Parental DNA samples were sequenced on the Illumina platform to generate a total of 274 Gb paired-end data. The progeny DNA sample was sequenced using PacBio long reads (217.3 GB) and 10x Genomics linked reads (226.38GB) along with 802 Gb of optical mapping data and 40X Illumina paired end data

Initially the data was partitioned at a kmer of 21 and then subsequently at 18 considering a minimum kmer coverage of 10 using Meryl. The PacBio Long-read segregation, for each parental haplotype, was over 99.99%, with less than 140Mb of raw reads being present in the non-classified set, confirming the trio. Trio binning based FALCON assembly of each haplotype was scaffolded with 10x Genomics reads and super-scaffolded with BioNano Maps to build reference quality assembly of parental haplotypes. Paternal haplotype assembly was of 2.63Gb with 59 scaffolds whereas maternal haplotype assembly was 2.64Gb in size with 64 scaffolds. N50 observed for both paternal and maternal haplotype was 81.98Mb and 83.23Mb, respectively. The assemblies were submitted to National Centre for Biotechnology Information (NCBI) database - Sire Haplotype accession number VDCB00000000 and Dam Haplotype accession number VDCC00000000. BUSCO single copy conserved core gene set coverage with other eukaryotes genomes was > 91.25%, and gVolante-CEGMA completeness was

>96.14% for both haplotypes. Finally, RaGOO was used to order and build the chromosomal level assembly with 25 scaffolds and N50 of 117.48 Mb (sire haplotype) and 118.51 Mb (dam haplotype). The final haplotype resolved Murrah genome assembly achieved > 99% genome coverage against estimated buffalo genome size of 2.66 Gb.

Whole genome sequencing of 296 buffaloes of 9 riverine breeds and 1 swamp buffalo breed was performed to study the variability among Indian buffaloes. More than 14 million variants including 12 million SNPs were identified per breed after QC, against Mediterranean water buffalo genome assembly (GCA\_003121395.1).

Using these variants a custom designed microarray panel BUFFCHIP was developed for genotyping of Indian buffaloes with 60K SNPs.

## **W159: Buffalo Genomics**

### **Marker Discovery and Associations with Beta-Carotene Content in Indian Buffalo and Dairy Cattle Breeds**

**Josue Chinchilla Vargas**<sup>1</sup>, Francesca Bertolini<sup>2</sup>, Jayant Khadse<sup>3</sup>, A. B. Juneja<sup>4</sup>, Prasad Deshpande<sup>4</sup>, P. M. Kakramjar<sup>4</sup>, A. K. Karlekar<sup>4</sup>, A.B. Pande<sup>4</sup>, Rohan L. Fernando<sup>1</sup> and Max F. Rothschild<sup>1</sup>, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)National Institute of Aquatic Resources, Lyngby, Denmark, (3)BAIF Development Research Foundation, Warje, India, (4)BAIF Development Research Foundation, Pune, India

Vitamin A is essential for human health, but current intake levels in many developing countries such as India are too low due to malnutrition. According to the World Health Organization, an estimated 250 million preschool children are vitamin A deficient globally. This number excludes pregnant women and nursing mothers, who are particularly vulnerable. Efforts to improve access to vitamin A are key because supplementation can reduce mortality rates in young children in developing countries by around 23%. Three key genes, BCMO1, BCO2, and SCARB1, have been shown to be associated with the amount of  $\beta$ -carotene (BC) in milk. Whole-genome sequencing reads from the coordinates of these 3 genes in 202 non-Indian cattle (141 *Bos taurus*, 61 *Bos indicus*) and 35 non-Indian buffalo (*Bubalus bubalis*) animals from several breeds were collected from data repositories. The number of SNP detected in the coding regions of these 3 genes ranged from 16 to 26 in the 3 species, with 5 overlapping SNP between *B. taurus* and *B. indicus*. All these SNP together with 2 SNP in the upstream part of the gene but already present in dbSNP (<https://www.ncbi.nlm.nih.gov/projects/SNP/>) were used to build a custom Sequenom array. Blood for DNA and milk samples for BC were obtained from 2,291 Indian cows of 5 different breeds (Gir, Holstein cross, Jersey Cross, Tharparkar, and Sahiwal) and 2,242 Indian buffaloes (Jafarabadi, Murrah, Pandharpuri, and Surti breeds). The DNA was extracted and genotyped with the Sequenom array. For each individual breed and the combined breeds, SNP with an association that had a P-value <0.3 in the first round of linear analysis were included in a second step of regression analyses to determine allele substitution effects to increase the content of BC in milk. Additionally, an F-test for all SNP within gene was performed with the objective of determining if overall the gene had a significant effect on the content of BC in milk. The analyses were repeated using a Bayesian approach to compare and validate the previous frequentist results. Multiple significant SNP were found using both methodologies with allele substitution effects ranging from 6.21 (3.13) to 9.10 (5.43) micrograms of BC per 100 mL of milk. Total gene effects exceeded the mean BC value for all breeds with both analysis approaches. The custom panel designed for genes related to BC production demonstrated applicability in genotyping of cattle and buffalo in India and may be used for cattle or buffalo from other developing countries. Moreover, the recommendation of selection for significant specific alleles of some gene markers provides a route to effectively increase the BC content in milk in the Indian cattle and buffalo populations.

## **W160: Buffalo Genomics**

## Updating the Buffalo Array Map and Prediction of QTL Positions in the New Water Buffalo Genome

Andre Vieira do Nascimento<sup>1</sup>, Diercles Francisco Cardoso<sup>1</sup>, Muhammad Yasir Nawaz<sup>2</sup>, Andrea R. S. Romero<sup>1</sup>, Daniel J. A. Santos<sup>3</sup>, Cedric Gondro<sup>2</sup> and Humberto Tonhati<sup>1</sup>, (1)São Paulo State University, Jaboticabal, Brazil, (2)Michigan State University, East Lansing, MI, (3)University of Maryland

When the Axiom Buffalo Genotyping 90k array (Affymetrix, Santa Clara-CA) was initially released, a buffalo reference genome was not assembled and the first version of the marker map with its genomic positions had to be based on the bovine genome (assembly\_UMD\_3.1) as a reference. In 2018, an updated map with single nucleotide polymorphisms (SNPs) distributed along the 25 buffalo chromosomes was released, following the newly released assembly of the buffalo genome (UOA\_WB\_1). This was a significant update which provided marker positions on the correct species genome; but it also generated some disadvantages: a large number of SNPs with unknown positions (16,262); and it made it more difficult to link genomic regions with known quantitative trait loci (QTL) annotated to the cattle genome. The objective of this study was to 1)map buffalo markers without known position on the Affymetrix SNP map and 2)to align the QTLs reported for cattle on CattleQTLdb (assembly\_UMD\_3.1) against the buffalo genome. To build a new map, all Affymetrix SNPs probe sequences (123,040) were re-aligned against the UOA\_WB\_1 and the sequences that had a length of at least 36bp, identity greater than 90% and the lowest single e-value were used to generate a new map which was then combined with the current Affymetrix map. There were 103,264 sequences after this quality control (QC) filtering and, out of these, 9,930 SNPs did not have a position in the Affymetrix map. The intersect of SNPs with map positions in both the Affymetrix and the aligned map was 93,334. Out of these, 90,928(97.42%) had the same position, 17(0.02%) were mapped to different chromosomes, and 2,389 SNPs(2.56%) were mapped to the same chromosome but at a different base pair position (only 16 were more than 5bp apart). When there was a divergence between positions, the original one determined by Affymetrix was maintained. The number of mapped markers increased from 106,778 in the Affymetrix map to 116,708 in the new map. This increased the mapped marker density by 9,930 SNPs and reduced the average marker spacing by ~2kb. For the second part of this work, cattle QTL annotated on the UMD\_3.1 assembly and available in the CattleQTLdb\_r.39(127,191) were downloaded. QC filtering was applied to the data to remove QTL without a position, duplicated and longer than 100bp. After QC, the sequences were aligned to the buffalo genome and a second QC step was applied to remove sequences with non-unique alignments and a probability of being wrongly mapped greater than 1%. After alignment and QC, 63,959 cattle QTL had predicted positions on the buffalo genome, this is approximately 50% of the total number of QTL in the CattleQTLdb. Trait classes with the highest number of QTL were reproduction(27,601), milk(20,805) and production traits(6,854). The new map resulting from this study improved the coverage of the Affymetrix SNP map by almost 10,000 markers, and the alignment of the bovine QTL database should be useful for association studies in buffalo and is a first step towards establishing a water buffalo QTLdb in the future. Grant FAPESP\_2018/25725-1.

### W161: Buffalo Genomics

#### Identifying Hub Genes for Heat Tolerance in Water Buffalo (*Bubalus bubalis*) using Transcriptome Data

Shenhe Liu<sup>1</sup>, Tingzhu Ye<sup>1</sup>, Zipeng Li<sup>1</sup>, Aixin Liang<sup>1</sup>, Tingxian Deng<sup>2</sup> and Liguang Yang<sup>1</sup>, (1)Huazhong Agricultural University, Wuhan, China, (2)Buffalo Research Institute, Chinese Academy of Agricultural Sciences, Nanning, China

Heat stress has a detrimental effect on the physiological and production performance of buffaloes. Elucidating the underlying mechanisms of heat stress is challenging, therefore identifying candidate genes is urgent and necessary.

We evaluated the response of buffaloes (n = 30) to heat stress using the physiological parameters including lower rectal temperature (RT) and respiratory rate (RR), ELISA indexes to check heat shock protein (HSP70 and HSP90) and cortisol (COR) levels, and hematological parameters including hemoglobin (Hb), hematocrit (Hct) and red blood cells (RBCs). To better ascertain NH and NHT

individuals by using heat stress indexes, these buffaloes were carried out to perform principal component analysis (PCA) and grouped into two groups by PC1, heat tolerant (HT) group and non-heat tolerant (NHT) group. We then extract total RNA from blood and performed mRNA and microRNA (miRNA) expression profiles analysis between HT (n = 4) and NHT (n = 4) buffaloes, as well as the specific modules, significant genes, and miRNAs related to the heat tolerance identified using the weighted gene co-expression network analysis (WGCNA).

The results indicated that the buffaloes in HT had a significantly lower RT and RR and displayed higher plasma HSP70, HSP90 and COR levels than those of NHT buffaloes. Differentially expressed analysis revealed a total of 753 differentially expressed genes (DEGs) and 16 differentially expressed miRNAs (DEmiRNAs) were identified between HT and NHT. Using the WGCNA analysis, these DEGs assigned into 5 modules, 4 of which were significantly correlated with the heat stress indexes. Interestingly, 158 DEGs associated with heat tolerance in the turquoise module were identified, 35 of which were found within the protein-protein interaction network. Several hub genes (IL18RAP, IL6R, CCR1, PPBP, IL1B, and IL1R1) were identified that significantly enriched in the Cytokine-cytokine receptor interaction. The findings may help further elucidate the underlying mechanisms of heat tolerance in buffaloes.

## **W162: Buffalo Genomics**

### **The African Buffalo and its Genomic Diversity**

**Christelle Robert**, The Roslin Institute, Edinburgh, United Kingdom

The African buffalo (*Syncerus caffer*) is the most closely related African wild bovid to the domestic cow (*Bos taurus*). Selection pressure on buffalo populations is mostly environmentally derived (*i.e.* minimal human influence –there has been no domestication), and a large component of environmental selection pressure is from pathogens. Buffalo are infected by all of the most important African cattle pathogens, but generally do not suffer the serious clinical disease seen in cattle; most likely due to prolonged co-evolution with pathogens selecting for disease control. The African buffalo therefore represents a uniquely valuable model in terms of understanding bovine co-evolution with infectious disease, and how pathogens may have shaped buffalo populations. Additionally, knowledge of how buffalo are able to control infections with these pathogens and the genetic basis of such resistance could be exploitable for devising methods to control the corresponding infections in cattle. However, to date this type of investigation has been limited by the lack of a high-quality reference genome for *S. caffer* to enable the qualitative analysis of the genetic basis behind important traits. We report the first high-quality *de novo* reference genome for *S. caffer* based on 60X PacBio reads, 90X Illumina reads and Dovetail Chicago and Hi-C data for scaffolding using HiRise. The genome assembly was further improved with gap filling using PBJelly and polishing using Pilon. Additionally, we collected an unprecedented number of African buffalo DNA samples (N=239) from across Africa covering the geographical distribution (East, West and Southern Africa) and containing representatives of the four known subspecies. This includes populations of *S. c. caffer* from areas that historically have been either endemic or free of specific pathogens or pathogen vectors. Additionally, the West African subspecies (*S. c. nanus*, *S. c. brachyceros* and *S. c. aequinoctalis*) are in areas where there have never been certain pathogen species. All samples were whole-genome sequenced (30X or 15X), and BWA-GATK pipeline was used to identify variants (SNP, Indels) specific to each sub-species and geographic population to explore the African buffalo genomic diversity. These data will further enable the study of the impact of natural selection on the buffalo genome across evolutionary timescales. Additionally, the genetic analysis has the potential to identify previously cryptic species and subspecies, as recently revealed in similar studies in giraffe.

## **W163: Cacao Genomics Workshop**

### **Prospects for Pan-Genomics in *Theobroma cacao*: A High-Resolution Diversity Panel Including Tolerant and Susceptible Genotypes**

**Eric K. Wafula**<sup>1</sup>, Noah P. Winters<sup>1</sup>, Prakash Timilsena<sup>1</sup>, Paula Ralph<sup>1</sup>, Claude W. dePamphilis<sup>1</sup>, James H. Marden<sup>1</sup>, Peter L. Tiffin<sup>2</sup>, Brett Tyler<sup>3</sup>, Dapeng Zhang<sup>4</sup>, Siela N. Maximova<sup>1</sup> and Mark J. Gaultinan<sup>1</sup>, (1)The Pennsylvania State University, University Park, PA, (2)Department of Plant Biology, University

of Minnesota, Saint Paul, MN, (3)Oregon State University, Corvallis, OR, (4)Sustainable Perennial Crops Laboratory, USDA/ARS, Beltsville, MD

The genetic variations among individuals of the same species provide the genetic material which human and natural selection can act upon to facilitate functional and phenotypic diversity. Resequencing approaches used extensively in genome-wide surveys of single-nucleotide polymorphisms (SNPs) depend on a single reference genome. They have limited capacity to capture genetic diversity within a species due to large structural variations (SVs). Copy number variants (CNVs) and presence/absence variants (PAVs) are the two classes of SVs thought to drive the majority of gene content and phenotypic variation that affects agronomically important traits in crops. Sequenced genomes of multiple individuals of a single species can enable the characterization of causal genetic variants. *Theobroma cacao* is a small understory tree native to the rainforests of the Amazon basin but cultivated globally for its seeds that are a source of cacao, the main ingredient in chocolate. We have sequenced over 30 accessions of *Theobroma cacao* covering the three major diversity clades (Contamana, Iquitos, and Curray) and three homozygous groups of special interest (Criollo, Amelonado, and Nacional). The selected accessions vary with respect to resistance to *Phytophthora*, a fungus-like pathogen that is responsible for the black pod rot disease in *Theobroma cacao*, in order to provide a stratified sample for identification of shared resistance across lineages as well as lineage-specific resistance. For this purpose, we present a collection of high quality *de novo* genome assemblies and on-going efforts to annotate a full complement of genes representing all the sequenced accessions. We hope that these data will provide the research community with a resource to facilitate genomics-assisted breeding to improve cacao production.

#### **W164: Cacao Genomics Workshop**

##### **Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in *Theobroma cacao***

**Tuomas T. Hamala**, University of Minnesota, St. Paul, MN

Separating footprints of adaptation from demography is challenging. When selection has acted on a single locus with major effect, this issue can be alleviated through signatures left by selective sweeps. However, as adaptation is often driven by small allele frequency shifts at many loci, studies focusing on single genes are able to identify only a small portion of genomic variants responsible for adaptation. In face of this challenge, we utilize coexpression information to search for signals of polygenetic adaptation in *Theobroma cacao*, a tropical tree species that is the source of chocolate. Using transcriptomics and a weighted correlation network analysis, we group genes with similar expression patterns into functional modules. We then ask whether modules enriched for specific biological processes exhibit cumulative effects of differential selection in the form of high  $F_{ST}$  and  $d_{XY}$  between populations. Indeed, modules putatively involved in protein modification, flowering, and water transport show signs of polygenic adaptation even though individual genes that are members of those groups do not bear strong signatures of selection. Modeling of demography, background selection, and the effects of genomic features reveal that these patterns are unlikely to arise by chance. We also find that specific modules are enriched for signals of strong or relaxed purifying selection, with one module bearing signs of adaptive differentiation and an excess of deleterious mutations. Our results provide insight into polygenic adaptation and contribute to understanding of population structure, demographic history, and genome evolution in *T. cacao*.

#### **W165: Cacao Genomics Workshop**

##### **Gene Expression Analysis of the Response of *Theobroma cacao* to the Infection of *Phytophthora palmivora***

**Leidy Paola Delgadillo Duran**, Colombian Corporation Agriculture Research (AGROSAVIA), Mosquera, Colombia, Mauricio Soto-Suárez, Colombian Corporation Agriculture Research (AGROSAVIA), Colombia and Roxana Yockteng, Agrosavia - Corporacion Colombiana de Investigacion Agropecuaria, BOGOTA, Colombia

Black pod cocoa disease caused by *Phytophthora spp* produces annual losses of approximately 30% of the crop production, that correspond to losses of 800 million dollars worldwide. The disease produced by

the pathogen *Phytophthora palmivora*, the most cosmopolitan species, occurs in important producing areas of Colombia such as Tolima and Huila affecting the production and quality of fruits and the generation of seedlings in nurseries. The generation of cacao plants with genetic resistance to diseases is probably the most effective, ecologically and economically viable way to make cacao cultivation sustainable. Transcriptomic studies could provide useful information about the genes involved in the resistance of diseases. The main objective of the present work was to study the genes expressed during the early stages of infection (0h, 24h, 48h and 96h) of black pod in two cacao genotypes with contrasting response to *P. palmivora*, the susceptible genotype CCN-51 and the tolerant genotypes SCA-6. RNAseq libraries were prepared and sequenced with Illumina technology. Preliminary results showed a total of 13,411 genes corresponding 28,406 transcripts reported in the cacao genome. Multiple differentially expressed genes were observed; a higher proportion of genes were found between 48 and 96hours post infection. The analyses showed a presence of genes with an important role in the signal transduction related to the plant defense response. These results are promissory to develop marker genes associated to early pathogen resistance, which can be used in the selection of cacao resistant materials.

**Key words:** *Theobroma cacao*, *Phytophthora palmivora*, inoculation, transcriptomic, differential gene expression

### **W166: Cacao Genomics Workshop**

#### **Molecular Genetic Approaches to Reducing Cadmium Accumulation in Cacao**

**Jim M Dunwell**, The University of Reading, UK, READING, United Kingdom

The presence of the toxic metal cadmium (Cd<sup>2+</sup>) in certain foodstuffs is recognized as a global problem, and there is increasing legislative pressure to reduce the content of Cd in food. Cacao (*Theobroma cacao*) is one of the crops known to accumulate Cd in certain conditions. There are a range of possible genetic and agronomic methods being tested as a route to such reduction. As part of a gene-based approach, we focused on the Natural Resistance-Associated Macrophage Proteins (NRAMPs), a family of proton/metal transporter proteins. The plant NRAMP gene family are responsible for uptake of the nutritionally vital divalent cations Fe<sup>2+</sup>, Mn<sup>2+</sup>, Zn<sup>2+</sup>, as well as Cd<sup>2+</sup>. We identified the five NRAMP genes in cacao, sequenced these genes and studied their expression in various organs. We then confirmed the expression patterns in response to variation in nutrient cation availability and addition of Cd<sup>2+</sup>. Functional analysis by expression in yeast provided evidence that NRAMP5 encodes a protein capable of Cd<sup>2+</sup> transport and suggests this gene as a target for genetic selection/modification. A similar approach has been adopted to a study of Heavy Metal ATPases (HMAs).

### **W167: Cacao Genomics Workshop**

#### **Establishment of Cacao Reference Genotypes from Different Genetic Groups using SNP Markers**

**Osman Ariel Gutierrez**, Subtropical Horticultural Research Station, USDA-Agricultural Research Service, Miami, FL, Dapeng Zhang, Sustainable Perennial Crops Laboratory, USDA/ARS, Beltsville, MD, Donald Livingstone, Mars Wrigley, Davis, CA and Chris Turnbull, School of Agriculture, Policy and Development, University of Reading, United Kingdom

Cacao (*Theobroma cacao* L.) is an important perennial crop for small farmers in tropical regions of the world whose center of origin and domestication is the Amazon Basin. Mislabeling of cacao accessions has been an ongoing problem across cacao collections worldwide. This situation has affected farmers because they are not receiving the correct planting materials, and this usually had led to poor field performance. Also, plant breeders have developed segregating populations with the assumption that some important traits would be present in their populations. Unfortunately, this mislabeling has affected the advancement of their breeding program. A set of SNP markers as well as a group of reference genotypes is urgently needed to properly classify cacao accessions that are currently present in several cacao collections worldwide. The goals of this study were to: (i) select cacao reference genotypes for each genetic group using single nucleotide polymorphism markers (ii) identify a set of single nucleotide polymorphism markers capable of differentiating cacao accessions. Five hundred and sixty-two selected

accessions belonging to the 10 cacao genetic groups (Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañon, Nacional, Nanay, and Purús) were screened using 1,060 single nucleotide polymorphic markers that are distributed across the 10 cacao chromosomes. Results show single nucleotide polymorphism markers were useful in the separation of genetic groups as well in the correct identification of accessions. Selected genotypes from each genetic group with their corresponding genetic information will be deposited in public database such as the International Cocoa Germplasm Database (ICGD) where they would serve as designated reference genotypes.

## **W168: Camelids**

### **The Camelome: Combining and Comparing the Transcriptome, Proteome, and Phosphoproteome in the Dehydrated Camel Kidney**

**Benjamin T. Gillard**<sup>1</sup>, Michael Greenwood<sup>1</sup>, Panjiao Lin<sup>1</sup>, Mahmoud A. Ali<sup>2</sup>, Phil Lewis<sup>1</sup>, Kate J. Heesom<sup>1</sup>, Fernando Alvira-Iraizoz<sup>1</sup>, Pamela A. Burger<sup>3</sup>, Abdu Adem<sup>2</sup> and David Murphy<sup>1</sup>, (1)University of Bristol, Bristol, United Kingdom, (2)UAEU, Abu Dhabi, United Arab Emirates, (3)Research Institute of Wildlife Ecology, Vetmeduni, Vienna, Austria

The one-humped dromedary camel (*Camelus dromedarius*) is renowned for its ability to survive long periods without access to water due to many physiological adaptations that assist water conservation. For example, camels produce highly concentrated urine as a consequence of efficient water reabsorption in the kidney. To investigate the molecular mechanisms responsible for minimising water loss, we compared the transcriptome, proteome, and phosphoproteome profiles of control (water *ad libitum*), dehydrated (three week water restriction), and rehydrated (three week water restriction followed by three days of water *ad libitum*) groups of camels (n=5 per group). RNA and protein samples were isolated separately from the Cortex and Medulla of the kidney, followed by the generation of transcriptome profiles by RNAseq (Illumina NextSeq) and proteome/phosphoproteome profiles by TMT-MS. RNAseq results were aligned to the camel genome (Elbers et al., 2019) using STAR. Reads were counted by featurecounts then differential expression between two conditions was estimated using DESeq2. Protein hits from the TMT-MS experiment were identified using protein sequences inferred from the transcriptome then abundance ratios were calculated against a pool of samples. Changes in phosphorylation were calculated as a ratio normalised to changes in protein abundance. Amongst the changes in dehydration in the medulla, aquaporin-2 transcript abundance and phosphorylation both increased significantly ( $P < 0.05$ ; Benjamini-Hochberg procedure adjusted p-value). Analysis of the relative abundance of transcripts, protein, and phosphorylation sites between conditions increases our understanding of how osmoregulatory mechanisms aid survival in desert-adapted species at different stages of gene expression.

Elbers, J.P., Rogers, M.F., Perelman, P.L., Proskuryakova, A.A., Serdyukova, N.A., Johnson, W.E., Horin, P., Corander, J., Murphy, D., Burger, P.A., 2019. Improving Illumina assemblies with Hi-C and long reads: An example with the North African dromedary. *Mol. Ecol. Resour.* 19, 1015–1026. <https://doi.org/10.1111/1755-0998.13020>

## **W169: Camelids**

### **Transcriptome of Supraoptic Nucleus in Dehydrated Dromedary Camel**

**Panjiao Lin**<sup>1</sup>, Michael Greenwood<sup>1</sup>, Benjamin T. Gillard<sup>1</sup>, Mahmoud A. Ali<sup>2</sup>, Pamela A. Burger<sup>3</sup>, Abdu Adem<sup>2</sup> and David Murphy<sup>1</sup>, (1)University of Bristol, Bristol, United Kingdom, (2)UAEU, Abu Dhabi, United Arab Emirates, (3)Research Institute of Wildlife Ecology, Vetmeduni, Vienna, Austria

In order to withstand long-term water deprivation in arid environments, the one-humped Arabian camel (*Camelus dromedarius*) has evolved robust water conservation mechanisms. At the level of the kidney, the camel produces a low volume of highly concentrated urine, especially following dehydration, as a consequence of the highly efficient reabsorption of water. This is mediated by the actions of the antidiuretic hormone arginine vasopressin (AVP), which is produced in the hypothalamic supraoptic nucleus (SON). Our previous studies on rodents (Stewart et al., 2011) showed that the transcriptome of

the SON undergoes dramatic function-related changes following osmotic challenge, which contribute to the facilitation of AVP production and delivery, and hence survival during dry periods. To investigate the osmoregulatory circuitry of the camel brain, we have compared the transcriptomes of the SON from water replete (water *ad libitum*, n=5), dehydrated (water deprivation for three weeks, n=5) and rehydrated (water deprivation for three weeks followed by water *ad libitum* for three days, n=5) camels. RNA *in situ* hybridization assays (RNAscope, ACD) were performed on frozen sections of the camel hypothalamus to map the mRNA distribution of AVP, which was used as a preliminary mapping to facilitate the collection of SON samples. Subsequently, RNA was isolated from SON samples using a Direct-zol™ RNA MiniPrep kit (Zymo research) and then subjected to RNAseq (Illumina NextSeq). RNAseq results were mapped to the camel genome (Elbers et al., 2019) using STAR. Reads were counted by featureCounts and differentially expressed genes were identified using DESeq2. By comparing the data from our rodent datasets (Stewart et al., 2011) we have identified common elements of the response to osmotic stress, as well as pathways unique to the dromedary camel.

## **W170: Camelids**

### **Applying Shotgun Sequencing to Detect Early Dromedary-Bactrian Camel Hybrids**

**Sara Lado**, Veterinary Medicine University, Vienna, Vienna, Austria, Pamela Burger, Vetmeduni Vienna, Research Institute of Wildlife Ecology, Vienna, Austria, Jean Pierre Elbers, Research Institute of Wildlife Ecology, Vetmeduni Vienna, Vienna, Austria, Joris Peters, Institut für Paläoanatomie, Domestikationsforschung und Geschichte der Tiermedizin, Department für Veterinärwissenschaften, Tierärztliche Fakultät der LMU München, Munich, Germany, Munich, Germany and Canan Çakırlar, Institute of Archaeology, University of Groningen, Groningen, Netherlands, Groningen, Netherlands

To create stronger and more robust camels, humans deliberately started to interbreed the two domestic Old World camel species - dromedary (*Camelus dromedarius*) and Bactrian camel (*Camelus bactrianus*), already in historic times. It is not clear when and where interbreeding began, but some postulate as early as pre-Roman times, resulting in ancient hybridization events that would not occur naturally. This human-mediated crossbreeding still practiced today in regions where camel hybrids are economically important in social and cultural events, as for example in Turkey.

Our main objective was to combine zooarchaeological and genetic analysis to understand when and where camels first were interbred. For this, we extracted DNA from camel bone fragments found at archaeological sites mainly from the Iron Age. We followed strict laboratory protocols for ancient DNA extraction and analysis. We subjected DNA extracts to ancient DNA optimized library preparation and performed Illumina shotgun sequencing. We analyzed the resulting ultra-low coverage shotgun sequencing reads with the Paleomix and Zonkey pipelines with the later specifically developed for hybrid detection in archaeological contexts. In parallel, we Sanger sequenced the ancient DNA extracts using species-specific diagnostic single nucleotide polymorphism (SNP) screens and performed osteological examination on the bone fragments. Finally, archaeologically assigned sample dates were verified with direct radiocarbon dating of hybrid samples, enhancing the chronological emergence sequence and spread of hybrid camels. Our results shed light on historically and culturally significant anthropogenic hybridization in Old World camels.

## **W171: Camelids**

### **Development of a SNP Microarray for Alpacas**

Marcos Calderon Montes<sup>1</sup>, Manuel J. More Montoya<sup>1</sup>, **Gustavo A. Gutierrez Reynoso**<sup>2</sup> and F. Abel Ponce de León<sup>3</sup>, (1)UNIVERSIDAD NACIONAL AGRARIA LA MOLINA, Peru, (2)UNIVERSIDAD NACIONAL AGRARIA LA MOLINA, LIMA, Peru, (3)Department of Animal Science, University of Minnesota, Saint Paul, MN

The production of alpaca fiber contributes to the economy of rural families in the highland areas of the Andes. The implementation of genetic improvement programs based on genomic selection could accelerate the genetic gain for fiber quantity and quality. The aim of this study was the discovery of

single nucleotide polymorphisms (SNPs) and the development of a SNP microarray in alpacas. DNA samples from 150 white Huacaya alpacas originating from two Peruvian geographical Andean regions were obtained to generate ApeKI and PstI/MspI reduced representation libraries for each sample. Libraries were sequenced on a HiSeq 2500 utilizing v4 chemistry and 1x100 single end reads generating a mean read depth of ~6X per library. A bioinformatic analysis using a variant calling pipeline allowed to identify 4'283,956 variants across the VicPac3.1 alpaca reference genome (GCA\_000164845.4). A list of 228,636 SNPs was generated considering the parameters phred-scaled quality score (>10), call rate ( $\geq 0.45$ ), minor allele frequency (between 0.05 and 0.50), Illumina Design Score ( $\geq 0.6$ ), and no other SNPs located within the 81 bp SNP sequence. Of these, 80,219 SNPs located at equidistant intervals of 25 Kbp were identified and will be part of the Alpaca SNPchip. In this manner, 97% of chromosomally assigned scaffolds and 78% of unassigned scaffolds are covered with SNPs, representing a genome coverage of 92%. Finally, we are also including SNPs localized in candidate genes for fiber growth and fiber color.

Keywords: alpaca, SNP, microarray

## **W172: Camelids**

### **Molecular Studies of the Minute Chromosome Syndrome in Alpacas**

**Brian W. Davis**, Texas A&M, College Station, TX

The *Minute Chromosome Syndrome* (MCS) is a spontaneously-occurring reproductive disorder in female alpacas with a suspected genetic component. The condition is invariably associated with sterility due to underdeveloped reproductive tract, and occurs recurrently in the population. The syndrome was named MCS because affected individuals have a dramatic size difference between homologs of chr36 due to a *de novo* insertion of a nucleolar organizing region (NOR) in one homolog. NOR is not found in chr36 in normal fertile animals. Despite this known cytogenetic abnormality, the molecular causes of MCS remain unknown. Here, we sequenced whole genomes of 8 clinically and cytogenetically confirmed cases of MCS and 6 normal, fertile controls at ~30X average genome coverage. After alignment to the recent alpaca draft assembly *VicPac3.1*, and variant calling, examination of genome context for the three contigs currently associated with chr36 in *VicPac3.1* showed no difference between case and control individuals. This may be due to inadequate resolution of this region, as only on third of chr36 is represented in this assembly. As a next step, we generated Hi-C data from the alpaca reference individual to produce a near-chromosome-level improvement to the VicPac draft assembly. In particular, since chr36 is the smallest autosome in the alpaca genome, inclusion of HiC data improves the contiguity of chr36, and facilitate continuing research on MCS and alpaca genomics as a field.

## **W173: Camelids**

### **A 2.1 Mb Assembly of the Alpaca Y Chromosome**

**Matthew Jevit**<sup>1</sup>, Andrew Hillhouse<sup>1</sup>, Mark F. Richardson<sup>2</sup>, Brian W. Davis<sup>1</sup>, Rytis Juras<sup>1</sup>, Malcom Ferguson-Smith<sup>3</sup>, Ahmed Tibary<sup>4</sup>, Vladimir Trifonov<sup>5</sup> and Terje Raudsepp<sup>1</sup>, (1)Texas A&M University, College Station, TX, (2)Deakin University, Geelong, Australia, (3)University of Cambridge, Cambridge, United Kingdom, (4)Washington State University, Pullman, WA, (5)Institute of Molecular and Cellular Biology, Novosibirsk, Russian Federation

Sequencing, annotation and functional analysis of the alpaca (*Vicuna pacos*) Y chromosome aims for better understanding male development and fertility, and the history and dynamics of patriline of the species. We sequenced flow-sorted Y DNA on PacBio Sequel and Illumina MiSeq (2x300bp reads) platforms. Long reads were assembled with CANU and short reads were incorporated with Pilon. For Y sequence annotation, we concatenated Y assembly with the latest alpaca reference genome *VicPac3.1* and mapped back to it short Illumina reads, cDNA sequences from Y-testis cDNA selection, and Trinity assembled testis transcriptome. This provided information for gene models and facilitated gene discovery. Repetitive elements were annotated with RepeatMasker. Approximately, 39% of the assembled sequence was repetitive with LINEs being the most abundant element followed STRs, LTRs

and SINEs Assembly contigs were validated for male specificity by PCR revealing four large male specific contigs with a combined total size of 2,136,669 bp of continuous Y sequence. Analysis of these contigs by discontinuous BLAST showed homology to known mammalian gametologs (*TBL1Y*, *USP9Y*, *DDX3Y*, etc.), the *SRY* gene and several known multi-copy genes such as *HSFY* and *TSPY*. The 4 contigs are currently being cytogenetically mapped by oligo-FISH using bioinformatically designed MyTag probes. In search for alpaca-specific Y genes, we have identified 2 male-specific transcripts that show no significant similarity by BLAST. We are presently assembling a second version of alpaca Y by combining previously generated sequence data with additional long reads from an Oxford nanopore MinION in order to improve the assembly.

#### **W174: Cannabis Diversity: Genomics and Phenomics**

##### **Sources of Variation: Steps Towards Understanding GxE and Development of Important Hemp Chemotypes over Time**

**Marie Turner**, Sunrise Genetics Inc., Fort Collins, CO

Although cannabis has long been deprived of the advantage of modern tools for breeding, legalization has facilitated rapid advances in this arena. However, although some types of genomic and genetic information are now becoming widely available, there still remains a striking lack of quality phenotypic data, in particular, that which has historically supported breeding of all other agriculturally important crops: data from formal outdoor field trials. Despite lifting of prohibition, there remain tight legal constraints in place for hemp that require very low levels of THC. While less problematic for classical industrial-type hemp crops (grain and fiber), in crops grown for high cannabidiol (CBD) yields, the close association of the CBD/THC metabolic pathways means that residual THC production (typically less than 1%) remains a threat to compliancy. This THC production appears to present a breeding challenge as it is not due to easily identifiable THC alleles. Therefore, understanding the source of this variation and how it is impacted by both other genes and environment will be critical under the current USDA guidelines. Other cannabis secondary metabolites (terpenes) which are important in multiple markets, also demonstrate substantial variation from both G and E. Understanding how all of these molecules behave across environments and over the course of phenological development will help farmers to select appropriate varieties and optimize agronomic practice to meet particular legal and market demands.

#### **W175: Cannabis Diversity: Genomics and Phenomics**

##### **Genotypic and Phenotypic Evaluation of Hemp Cultivars for Cannabinoid Production in New York**

**Lawrence B. Smart**<sup>1</sup>, George M. Stack<sup>1</sup>, Jacob Toth<sup>1</sup>, Craig H. Carlson<sup>1</sup>, Glenn Philippe<sup>2</sup>, Jocelyn KC Rose<sup>2</sup>, Jamie L. Crawford<sup>3</sup>, Donald R. Viands<sup>3</sup>, Ali R. Cala<sup>4</sup>, Christine D. Smart<sup>4</sup> and Denis Willett<sup>5</sup>, (1)Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (2)Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (3)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (4)Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (5)Dept. of Entomology, Cornell University, Geneva, NY

The rapid expansion of cultivation of hemp (*Cannabis sativa*) for the production of essential oils rich in cannabidiol (CBD) has led to a proliferation of seed sources, including feminized, non-feminized, and clonal cuttings. Very few of these have been characterized for their growth performance across different environments, uniformity, or compliance for tetrahydrocannabinol (THC) accumulation, and there are no established cultivar descriptions associated with plant variety protection or seed certification. We evaluated 30 hemp cultivars from various sources in replicated trials on two sites for growth performance, flowering time, uniformity, floral biomass yield, pest and disease susceptibility, and cannabinoid production across a sampling time line. Genetic similarity and homozygosity of these and over 150 other hemp cultivars sampled in NY in 2018 and 2019 were assessed based on genome-wide SNP analysis from genotyping-by-sequencing. Competitive PCR assay markers designed to CBDA synthase and predictive of cannabinoid chemotype were used to assess populations of many of the

cultivars under evaluation and revealed segregation for THC/CBD chemotype in some seed lots, raising the risk of non-compliant production of THC for some cultivars. The results of these initial studies will guide future research and breeding efforts with the aim of improving the overall quality of commercially available hemp cultivars.

### **W176: Cannabis Diversity: Genomics and Phenomics**

#### **Cataloging Existing Variation and Rebuilding Better Cannabis Genomes for New Markets**

**John McKay**, Colorado State University, Fort Collins, CO

With the end of prohibition, Cannabis cultivation is coming out of the closet. 2018 saw major changes in Federal law in many large countries, allowing for large scale production and marketing of Cannabis. In the US, the 2018 Farm Bill/Hemp Farming Act removed legal barriers to the hemp supply chain for grain, fiber and flower. The new bottleneck is genetics, due to a lack of cultivars bred to optimize sustainable production and meet the demands of new and emerging markets. In addition, production agronomy has yet to be standardized for hemp, which has important implications to defining selection targets in a breeding program. I will discuss the breeding program of New West Genetics, a company that is producing novel, proprietary, certified hemp cultivars, optimized for large scale, mechanized production. The current state of production agronomy and proposed improvements will also be addressed. I will also present some public sector research on quantitative genetics in hemp, including GxE and QTL analysis of agronomic traits.

### **W177: Cannabis Diversity: Genomics and Phenomics**

#### ***Cannabis' indica/sativa* Dilemma: Clarity from Genomics and Chemistry**

**Sophie Watts**<sup>1</sup>, Michel McElroy<sup>2</sup>, Zoë Migicovsky<sup>1</sup>, Hugo Maassen<sup>3</sup>, Arno Hazekamp<sup>4</sup>, Robin van Velzen<sup>3</sup> and Sean Myles<sup>5</sup>, (1)Dalhousie University, Truro, NS, Canada, (2)Dalhousie University, Canada, (3)Bedrocan International, Veendam, Netherlands, (4)Hazekamp Herbal Consulting, Leiden, Netherlands, (5)Dalhousie University, Halifax, NS, Canada

Cannabis producers and consumers often classify varieties as either indica or sativa as a way to distinguish differences in aroma, ancestry and/or psychoactive effect. However, the extent to which these commonly used labels reflect their ancestries and represent actual differences in genetic and chemical profile remains unresolved. To investigate the relationship of ancestry labels with genetic and chemical composition, 137 cannabis samples labeled with varying proportions of indica and sativa were collected from across the Netherlands. Over 100,000 SNPs were genotyped using Genotype-by-Sequencing (GBS) and each sample was quantified for terpene and cannabinoid content using gas-chromatography. Our results indicate that the indica and sativa labels assigned to Cannabis varieties are supported neither by the genetic data nor by the chemical data. However, specific terpenes are significantly associated with Cannabis labels, which suggests that Cannabis producers may be assigning indica/sativa labels to their varieties based primarily on aroma rather than genetic ancestry.

### **W178: Cannabis Diversity: Genomics and Phenomics**

#### **Cannabis Genome Arrangement - Comparison of Multiple Heterozygous Genomes**

**Oori Weisshaus**<sup>1</sup>, Ronny Rechtman<sup>2</sup>, Tammy Biniashvili<sup>2</sup> and Doron Shem-Tov<sup>3</sup>, (1)NRGene, Ness Ziona, Israel, (2)NRGene, Israel, (3)NRGENE Ltd., Ness-Ziona, Israel

In recent years the *Cannabis sativa* genome is at the center of research efforts due to the plant's large array of secondary metabolism products and their potential to interact with mammalian systems. Removal of legislative barriers has enabled the publication of multiple genome assemblies using various technologies for sequencing and assembly, mostly using highly heterozygous plants. In this work we demonstrate a pan-genome comparison of multiple cannabis genomes of hemp and drug type cultivars. We have de-novo assembled two new heterozygous elite-line genomes with fully phased high accuracy assemblies and compared them to four public reference-level non-phased assemblies. This was carried

out in a pan-genome structure based on a common coordinate system of the CBDRx reference genome (CS10) that was found to be the most comprehensive public assembly. We have aligned and ordered multiple genomes with phased haplotypes and created uniform chromosome mapping. We have also generated a non-redundant dataset of 43,000 transcripts and mapped it to each haplotype. This enabled identification of allelic variation and novel homologues of cannabinoid biosynthesis genes, as well as an accurate comparison of copy number, present-absent and structural variations and identification of highly conserved gene region duplications. We have identified hyper variable regions and massive genome rearrangements that may hold great significance to cannabis and hemp research and breeding.

## **W179: Cannabis Diversity: Genomics and Phenomics**

### **Choosing a Reference Genome for the Cannabis Community**

**Todd Michael**, J. Craig Venter Institute, La Jolla, CA

There are eight Cannabis genomes deposited in NCBI (Nov 2019), which started with the first genomes released in 2011 based on short-read sequencing. Recently, the quality of the genomes has greatly improved with three chromosome resolved assemblies released over the last two years. In addition to these genomes, there are many reported genomes on the way. The goal of this talk will be to address what is needed for a Cannabis reference genome. Reference genomes have been critical in the early days of plant genome communities. A reference genome can focus and align research findings and provide reference points for new researchers. Key features of reference genomes will be discussed and a path forward for the Cannabis genome will be proposed.

## **W180: Cassava Genomics**

### **Mobilization of Host DNA Sequences during Cassava Mosaic Disease**

**Linda Hanley-Bowdoin**, North Carolina State University, Raleigh, NC

Cassava mosaic disease (CMD) is one of the most important diseases of cassava and a serious constraint on production across Africa and an emerging problem in Asia. CMD is caused by a complex of whitefly-transmitted, single-stranded DNA viruses designated as cassava mosaic begomoviruses (CMBs). There was a severe CMD pandemic in Sub-Saharan Africa in the 1990s and 2000s, leading to the deployment of CMD2 resistance cultivars throughout the region. In recent years, severe, atypical CMD symptoms have been observed occasionally on resistant cultivars, some of which carry the CMD2 locus, in African fields. Two DNA sequences, SEGS-1 and SEGS-2 (sequences enhancing geminivirus symptoms), are responsible in part for breaking CMD resistance and the increased severity of CMD symptoms. Sequences related to SEGS-1 and SEGS-2 occur in the genomes of all cassava cultivars examined to date, and SEGS-1 related sequences are found throughout the *Manihot* lineage. SEGS-1 and SEGS-2 are host sequences that have been recruited by CMBs to enhance CMD symptoms and break resistance. SEGS-1 is mobilized from the cassava genome and forms an episome during CMB infection, while SEGS-2 is a novel satellite that is derived from the cassava genome, packaged into virions and transmitted by whiteflies as part of the CMB disease complex. The recruitment of SEGS sequences from the cassava genome by CMBs represents a new class of virus/host interactions that can impact disease processes.

## **W181: Cassava Genomics**

### **Resolving the Cassava Haploid Genomes**

**Ben N. Mansfeld**<sup>1</sup>, Adam Boyher<sup>2</sup>, Mark Wilson<sup>2</sup>, Shujun Ou<sup>3</sup> and Rebecca Bart<sup>1</sup>, (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)Donald Danforth Plant Science Center, St Louis, MO, (3)Michigan State University / Iowa State University, East Lansing, IA

Cassava (*Manihot esculenta* Crantz, 2n=36) is an important food crop for a billion people across 105 countries where this starchy root is a critical staple. Cassava has a highly heterozygous genome, high genetic load, genotype-dependent asynchronous flowering, and is typically propagated by stem cuttings. Thus, any genetic variation between haplotypes, including large structural variations, are preserved by

such clonal propagation. Traditional genome assembly approaches generate a collapsed haplotype representation of the genome. In highly heterozygous plants, this introduces artefacts and results in an oversimplification of heterozygous regions of the genome. To independently resolve each haplotype of the cassava genome, we use a combination of Pacific Biosciences (PacBio), Illumina and Hi-C sequence reads to the cassava genotype TME7 (Oko-Iyawo). PacBio reads were assembled into contigs using the FALCON and FALCON-Unzip pipeline and phase switch errors within contigs were corrected using FALCON-Phase and Hi-C read data. The ultra-long-range genotype information from Hi-C sequencing was then used for scaffolding and within-scaffold phasing, thus correcting phase switch errors. After gap filling, the results are 36 highly contiguous chromosome sequences representing the two haploid copies of each chromosome – we report a contig N50 of 735 and 712Kb for phase 0 and 1, respectively, and a scaffold N50 of ~40Mb for both phases. Each haploid assembly is ~750MB. Both linkage- and optical maps confirmed the contiguity, order, and directionality of the assemblies. Comparison of the two phases revealed 3,753 structural variants, including insertions and deletions of up to 10kb, affecting more than 5Mb of the sequence. Annotation of haplotype specific genes and transposable elements is underway. These two haplotype assemblies will provide an excellent means to study the haplotype specific structural variation, synteny, and allele specific gene expression contributing to important agricultural traits and further our understanding of the genetics and domestication of cassava.

## **W182: Cassava Genomics**

### **Building Targeted DNA Methylation Tools for Cassava**

**Jason Gardiner**, University of California, Los Angeles, Los Angeles, CA

## **W183: Cassava Genomics**

### **Engineering Cyanide-Free Cassava using CRISPR-Cas9**

**Jessica B. Lyons**<sup>1</sup>, Michael A. Gomez<sup>2</sup>, Baljeet Gill<sup>2</sup>, Anthony T. Iavarone<sup>3</sup>, Jessica M. Ma<sup>4</sup>, Kodiak C. Berkoff<sup>4</sup>, Stacia K. Wyman<sup>2</sup>, Elaine Zhang<sup>2</sup>, Nigel J. Taylor<sup>5</sup>, Myeong-Je Cho<sup>2</sup>, Brian Staskawicz<sup>1</sup> and Daniel Rokhsar<sup>4</sup>, (1)University of California, Berkeley, Berkeley, CA, (2)Innovative Genomics Institute, (3)California Institute for Quantitative Biosciences (QB3), UC Berkeley, (4)University of California, Berkeley, (5)Donald Danforth Plant Science Center, St. Louis, MO

Cassava (*Manihot esculenta* Crantz) is a staple for over 800 million people worldwide. This starchy root crop is a major source of calories for roughly 40% of Africans and an excellent food security crop due to its tolerance for drought and marginal soils. A major challenge, however, is the presence of cyanogenic glucosides (cyanogens), which break down into toxic cyanide. Dietary cyanide exposure is associated with conditions such as neurodevelopmental deficits, the neurological disease tropical ataxic neuropathy, and the paralytic disease konzo. Post-harvest processing of cassava to remove cyanogens can be laborious and reduce nutrient content in the resulting products. The paralogous genes *CYP79D1* and *CYP79D2* catalyze the synthesis in leaves and shoots of cassava's principal cyanogens, which are then transported into the roots. By using CRISPR-Cas9 gene editing to knock out the *CYP79D* genes, singly and in combination, we expect to reduce or eliminate cyanogen levels in cassava. We have established a platform at UC Berkeley's Innovative Genomics Institute for cassava tissue culture, *Agrobacterium*-mediated transformation, and CRISPR-Cas9 gene editing. This precision breeding method facilitates the modification of a single gene or trait in cassava, avoiding time-consuming conventional breeding methods and preserving the complement of preferred traits in a given clonally propagated variety. As measured by mass spectrometry, bi-allelic knockout of both *CYP79D* genes in a model cultivar reduces cyanogen levels below detectability, and single-gene knockouts have intermediate cyanogen levels. We are now applying this editing approach to farmer-preferred cassava varieties. We are also pioneering gene editing methods which, by avoiding the incorporation of foreign DNA into a plant's genome, are expected to obviate GMO concerns and drawn-out regulatory hurdles, easing adoption of improved cassava varieties.

## **W184: Cassava Genomics**

## **Population Structure and Linkage Disequilibrium in Brazilian Cassava: Insights into Crop Diversification and Domestication**

**Guillaume J. Bauchet**<sup>1</sup>, Alex C Ogonna<sup>1</sup>, Luciano Braatz de Andrade<sup>2</sup>, Eder Oliveira<sup>2</sup> and Lukas A. Mueller<sup>1</sup>, (1)Boyce Thompson Institute, Ithaca, NY, (2)Embrapa, Brazil

Cassava (*Manihot esculenta*) is a major staple root crop of the tropics and originates from South America with Brazil as a center of genetic diversity for the *Manihot* clade.

In this study, we provide an assessment of a Brazilian cassava germplasm collection using genotyping by sequencing. Assessment of population stratification allowed identification and characterization of 10 distinct genetic groups. Estimates of chromosomal linkage disequilibrium highlighted the heterogeneity among chromosomes and groups. Geographical structure plays a role important domestication traits.

Current findings provide support to breeding applications such as genetic mapping and molecular breeding approaches including genomic selection.

### **W185: Cassava Genomics**

#### **Genetic Gain for Cassava Brown Streak Disease in White and Yellow Fleshed Cassava Breeding Populations in East Africa**

**Alfred Ozimati**, NaCRRI, Uganda, Kampala, Uganda, Robert Kawuki, National Crops Resources Research Institute (NaCRRI), Uganda, Kampala, Uganda, Williams Esuma, National Crops Resources Research Institute, Uganda, Kampala, Uganda, Marnin Wolfe, Department of Plant Breeding Cornell University, Ithaca, NY and Jean-Luc Jannink, USDA-ARS, Ithaca, NY

Cassava (*Manihot esculenta* Crantz) is a major crop in tropical and sub-tropical parts of the world because of its food and non-food uses. However, its production and productivity in Eastern, Central and Southern Africa is greatly affected by cassava brown streak disease (CBSD), which can cause up to 100% yield loss in highly susceptible varieties. The long breeding cycle of cassava (8-10 years) limits faster variety development and replacement in CBSD worst-hit areas. Using DarT-seq genotyping technology, the National cassava breeding program in Uganda genotyped its yellow and white-fleshed breeding populations for downstream application of genomic selection (GS) with the aim of shortening the cassava breeding cycle. Accordingly, this study assessed genetic gains attained for CBSD resistance in Uganda as well as the impact of accelerated breeding on overall genetic variation and inbreeding rate. We used phenotypic and the genotypic data for white-fleshed clones categorized under; foundation population (C<sub>0</sub>), cycle one (C<sub>1</sub>) and cycle two (C<sub>2</sub>) with population sizes of 395, 813 and 1,491 respectively. For the yellow-fleshed population, a total of 297 clones that formed the initial training population (TP) and 486 cycle one (C<sub>1</sub>) clones were used for the analysis. For white-fleshed clones, the population centered average genomic estimated breeding values (GEBVs) were 0.099 (C<sub>0</sub>), -0.036 (C<sub>1</sub>) and -0.012 (C<sub>2</sub>) for foliar CBSD scored at six MAP, while for CBSD root necrosis scored at harvest, the average GEBVs were 0.219 in C<sub>0</sub>, -0.004 in C<sub>1</sub> and -0.055 in C<sub>2</sub> clones. On the other hand, the yellow-fleshed clones had GEBVs of 0.0129 in C<sub>0</sub> and -0.007 in C<sub>1</sub>, while the values were 0.0133 and -0.008 in C<sub>0</sub> and C<sub>1</sub>, respectively. Overall, we observed lower average GEBVs for advanced cycles (C<sub>1</sub> and/or C<sub>2</sub>) than for founding populations, indicating overall genetic progress in breeding for CBSD resistance in Uganda. Equally motivating was maintained genetic diversity with no obvious inbreeding based on means of diagonal elements of marker kinship matrices for both yellow and white-fleshed as advances were made in breeding cycles, suggesting that accelerated breeding through GS did not compromise overall diversity. These findings provide basis for integration of GS in cassava breeding programs in an endeavor to deploy genetic gains in farmer's fields.

### **W186: Cattle/Sheep/Goat 1**

#### **Comparative Epigenomics and Genotype-Phenotype Association Analyses Revealed Conserved Genetic Architecture underlying Complex Traits between Cattle and Human**

**George E. Liu**, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

Lack of comprehensive functional annotations across a wide range of tissues and cell types severely hinders the biological interpretations of phenotypic variation, adaptive evolution and domestication in livestock. We here cross-mapped 8 histone marks of 1,300 samples from human to cattle, covering 178 unique tissues/cell types. By uniformly analyzing 723 RNA-seq and 40 whole genome bisulfite sequencing (WGBS) datasets in cattle, we validated that cross-mapped histone marks captured tissue-specific expression and methylation, reflecting tissue-relevant biology. Through integrating cross-mapped tissue-specific histone marks with large-scale genome-wide association study (GWAS) and selection signature results, we for the first time detected relevant tissues and cell types for 45 economically important traits and artificial selection in cattle. For instance, immune tissues are significantly associated with health and reproduction traits, multiple tissues for milk production and body conformation traits (reflecting their highly polygenic architecture), and thyroid for the different selection between beef and dairy cattle. Similarly, we detected relevant tissues for 58 complex traits and diseases in humans, and observed that immune and fertility traits in humans significantly correlated with those in cattle in terms of relevant tissues, which facilitated the identification of causal genes for such traits. For instance, PIK3CG, a gene highly specifically expressed in mononuclear cells, was significantly associated with both age-at-menopause in human and daughter-still-birth in cattle. ICAM, a T-cell specific gene, was significantly associated with both allergic diseases in human and metritis in cattle. Collectively, our results highlighted that comparative epigenomics in conjunction with GWAS and selection signature analyses could provide biological insights into the phenotypic variation and adaptive evolution in the target species. Cattle may serve as a model for human complex traits, by providing additional information beyond laboratory model organisms, particularly when more novel phenotypes become available in the near future.

### **W187: Cattle/Sheep/Goat 1**

#### **Using Long Read Sequencing to Derive an Epigenetic Aging Clock for Cattle**

Loan Nguyen, Queensland Alliance for Agriculture and Food Innovation - University of Queensland, Brisbane, QLD, Australia, **Elizabeth Ross**, Queensland Alliance For Agriculture and Food Innovation - University of Queensland, Brisbane, QLD, Australia and Ben Hayes, Queensland Alliance for Agriculture and Food Innovation - University of Queensland, Brisbane, Australia

Age estimation methods based on DNA methylation data have been widely studied in human, dogs, wolves and mice. However, DNA methylation patterns in livestock species remain vastly unexplored. Thus, we present an epigenetic aging clock for cattle based on tail hair methylation profiles. Initially we compared four different methylation detection technologies (reduce representation bisulfite sequencing, whole genome bisulfite sequencing, human methylation EPIC array and Oxford Nanopore sequencing). We confirmed that all methods were able to detect DNA methylation profiles, and those profiles differ between young and old animals. Moreover, among these technologies, Oxford Nanopore sequencing showed the most promise for generating high quality methylation profiles, with up to 20.4 million methylated sites identified in some samples. The study also observed a good correlation between nanopore methylation and bisulfite calls ( $r > 0.8$ ). Phase two of this study is to develop an epigenetic clock for cattle based on nanopore sequencing of 100 cattle with ages ranging from 5 days to 17 years. The ability to accurately predict age from a genomic sample using DNA methylation patterns could become a powerful estimator to support genetic improvement programs.

### **W188: Cattle/Sheep/Goat 1**

#### **Identification of 5-Hydroxymethylcytosine Markers in the Cattle Brain**

**Emily Stassen**, University of Vermont, Burlington, VT

Epigenetic variation plays an important role in phenotypic variability. DNA methylation is a dynamic epigenetic marker that can be added and removed from DNA nucleotides. The removal pathway begins when 5-methylcytosine (5-mC) is oxidized to 5-hydroxymethylcytosine (5-hmC). Both 5-mC and 5-hmC are stable epigenetic markers that influence transcription. 5-mC is thought to inhibit transcription,

whereas 5-hmC is thought to promote transcription. It is important to distinguish between these stable epigenetic markers to gain a clear understanding of how epigenetics is influencing transcription. However, traditional whole genome bisulfite sequencing methods do not differentiate between these markers. Because of this, 5-hmC is often misrepresented as 5-mC. The grouping of these two distinct markers during data collection could lead to misinformation on the impact of DNA methylation markers on transcription and phenotype. Differentiating between markers is especially significant in the brain, where there is an increase in the presence of 5-hmC. We have previously studied the relationship between DNA methylation and docility using standard whole genome bisulfite sequencing on select brain tissues. To differentiate between 5-hmC and 5-mC, we have carried out reduced representation oxidative bisulfite sequencing on brain tissues from the limbic system of 8 Red Angus x Simmental steers. Fastq sequencing files were trimmed with Trim Galore. Subsequently, Bismark was used for alignment to bovine reference index ARS-UCD 1.2, and MethPipe was used to call hydroxymethylation. These data show the presence of 5-hmC in the bovine brain, which emphasizes the need to differentiate between 5-hmC and 5-mC.

### **W189: Cattle/Sheep/Goat 1**

#### **The Ovine FAANG Project: A High-Resolution Atlas of Transcription Start Sites in the New Rambouillet Ovine Reference Genome**

**Emily L. Clark**<sup>1</sup>, Mazdak Salavati<sup>1</sup>, Alexandra J. Caulton<sup>2</sup>, Richard Clark<sup>3</sup>, Iveta Gazova<sup>4</sup>, Timothy P.L. Smith<sup>5</sup>, Kim C. Worley<sup>6</sup>, Noelle Cockett<sup>7</sup>, Alan L. Archibald<sup>1</sup>, Shannon M. Clarke<sup>2</sup>, Brenda M. Murdoch<sup>8</sup> and The Ovine FAANG Group, (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand, (3)Clinical Research Facility, University of Edinburgh, United Kingdom, (4)Institute for Genetics & Molecular Medicine, University of Edinburgh, United Kingdom, (5)USDA, ARS, USMARC, Clay Center, NE, (6)Baylor College of Medicine, Houston, TX, (7)Utah State University, Logan, UT, (8)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID

The overall aim of the Ovine FAANG project is to provide a comprehensive annotation of the new highly contiguous reference genome for sheep Oar\_rambouillet\_v1.0. To contribute to this aim we have generated a high-resolution atlas of transcription start sites (TSS) for sheep. Mapping of TSS is a key first step in understanding transcript regulation and diversity. Using 56 tissue samples collected from the reference ewe Benz2616 we have performed a global analysis of TSS using Cap Analysis Gene Expression (CAGE) sequencing. CAGE measures RNA expression by 5' cap-trapping and has been specifically designed to allow the characterization of TSS within promoters to the single-nucleotide resolution. We have adapted an analysis pipeline which uses TagDust2 for clean-up and trimming, Bowtie2 for mapping, CAGEfightR for clustering and IGV for visualization. Mapping of CAGE tags indicated that, for numerous transcripts, TSS vary across tissues. After filtering, of TSS clusters with <10 TPM, 29% of TSS overlapped with annotated 5' ends of transcripts, and a further 26% mapped to intronic regions, most likely as a consequence of missing annotation information. As such this comprehensive global annotation of TSS in sheep will significantly enhance the annotation of gene models in the new ovine reference assembly. The CAGE dataset is also being used to validate and enhance gene expression data from Illumina RNASeq, miRNA-Seq and IsoSeq libraries and analysed in parallel with methylation information. This will provide one of the highest resolution annotations of transcript regulation and diversity in a livestock species to date.

### **W190: Cattle/Sheep/Goat 1**

#### **Integrated -Omics Approaches for Meat Quality Improvement**

**Raluca Mateescu**<sup>1</sup>, Joel David Leal<sup>1</sup>, Fernanda Rezende<sup>2</sup>, Francisco Peñagaricano<sup>3</sup>, Eduardo E Rodriguez<sup>1</sup>, Chad Carr<sup>2</sup> and Mauricio Elzo<sup>1</sup>, (1)Department of Animal Sciences, University of Florida, Gainesville, FL, (2)University of Florida, (3)University of Florida, Gainesville, FL

Transcription has a substantial genetic control and genetic dissection of gene expression could help us understand the genetic architecture of complex phenotypes such as meat quality in cattle. A total of 80

steers were selected for phenotyping, genotyping and RNA-seq evaluation. A panel of traits related to meat quality along with information on 112,042 SNPs and expression data on 8,588 autosomal genes and 87,770 exons from 8,467 genes were included in an expression and splicing quantitative trait loci (QTL) mapping (eQTL and sQTL, respectively). Expression of 1,352 genes was previously identified as associated with meat quality traits using a gene, exon and isoform differential expression analysis. The identified QTLs were classified as cis or trans using 1 Mb as maximum distance between the associated SNP and the gene. Polymorphisms associated with expression of at least 20 genes, and splicing of at least 20 exons were considered QTL hot spots. A total of 8,377 eQTLs were identified, including 75.6% trans, 10.4% cis, 12.5% DE trans and 1.5% DE cis; 11,929 sQTLs were uncovered: 66.1% trans, 16.9% DE trans, 14% cis and 3% DE cis. Twenty seven expression master regulators and 13 splicing master regulators were identified and were classified as membrane associated or cytoskeletal proteins, transcription factors or DNA methylases. These genes could control expression of other genes through cell signaling or by a direct transcriptional activation/repression mechanism. The *ZNF804A*, *ALAD*, *OR13F1* and *ENSBTAG0000000336* genes were identified as both expression and splicing master regulators. In the present analysis, we show that eQTL and sQTL mapping makes possible positional identification of gene and isoform expression regulators. Additionally, this mapping provides new insight into the regulatory network architecture in *longissimus dorsi* muscle in beef cattle.

## **W191: Cattle/Sheep/Goat 1**

### **Exploring the Genetic Resistance to *Haemonchus contortus* Infection in Dohne Merino Sheep using RNA-Seq**

**Farai C. Muchadeyi**<sup>1</sup>, Tondani Madeleine Ramantswana<sup>2</sup>, Petunia D. Malatji<sup>3</sup>, Pranisha Soma<sup>4</sup> and Rian Pierneef<sup>1</sup>, (1)ARC-Biotechnology Platform, Pretoria, South Africa, (2)ARC-Biotechnology Platform, South Africa, (3)University of South Africa, Florida, South Africa, (4)Agricultural Research Council - API, Pretoria, South Africa

#### **Abstract**

The South African Dohne Merino is a dual-purpose sheep breed which was bred with the specific aim of maximizing wool and mutton production with an ability to thrive under harsh conditions. Gastrointestinal nematode (GIN) parasitic infections are a major concern in South Africa, affecting the production and also causing serious economic losses. Breeding for improved resistance to GINs is one strategy that will minimize the use of anthelmintics and ultimately reduce the production cost associated with animal health management. A good understanding of the genetic and molecular mechanisms associated with resistance to parasites will facilitate such breed improvement strategies. The aim of the study was to use RNA-Seq and differential gene expression profiling of the gastrointestinal tract of Dohne Merino sheep that were either resistant or susceptible to *H. contortus* to identify genes and pathways that play a role in conferring resistance to this pathogen in sheep. Mature Dohne Merino sheep were sampled from the Wauldby farm in the Stutterheim District in the Eastern Cape Province of South Africa. The farm has been phenotyping for Faecal Egg Count (FEC), FAMACHA and Body Condition Score (BCS) from which Estimated Breeding Values (EBVs) were estimated for resistance to *H. contortus*. From this farm, six animals phenotyped for FEC, FAMACHA and BCS as resistant (n=3) and susceptible (n=3) were selected for the study. The animals were sacrificed and the gastrointestinal tract phenotyped for worm and egg count. Total RNA was isolated from the gastrointestinal tract of 3 resistance and 3 susceptible tissue samples. The Illumina HiSeq2500 was used to generate an average of 66,912,43 of 125bp paired-end reads per sample. Quality-controlled reads were aligned to the reference genome (*Ovis aries* 4.0) using HISAT2. DESeq2 was used for differential gene expression analysis. Functional annotation of genes was based on the gene ontology categories using WEGO and KEGG for pathways analysis. Differentially expressed genes at FDR < 0.05 were identified, of which 127 genes were highly expressed in susceptible sheep and 79 genes in resistance sheep. Functional annotation and enrichment analysis showed that several genes were involved in immune response process, responses to stimuli, biological regulation, regulation of biological processes and other cellular processes which can be associated with the resistance of sheep to *H. contortus*. The p53 related proteins (*PPM1D* and *SIVA1*) which are responsible for cellular repair were upregulated in the resistant animals. The most important down

regulated genes in susceptible animals were associated with multidrug resistance-associated proteins that are under *ABCC* subfamily of *ABC transporter pathway* and *cAMP* signaling pathways. These have been shown to play a role in effluxion of anthelmintics across membrane. Overall, the approach enabled us to uncover genes and pathways that are implicated in resistance to *H. contortus*. This study forms the basis of selection of animals with the right genotypes for breeding of sheep resistant to infections by *H. contortus* nematodes.

**Key words:** RNASeq; Differential gene expressions; sheep; parasite resistance; *H. contortus*

## **W192: Cattle/Sheep/Goat 1**

### **Identification and Characterization of microRNA Genetic Variants in Dairy Cattle, from their Detection to the Analysis of their Biological Impacts**

Céline Bourdon<sup>1</sup>, Mekki Boussaha<sup>1</sup>, Marie-Pierre Sanchez<sup>1</sup>, Thierry Tribout<sup>1</sup>, Rachel Lefebvre<sup>1</sup>, Sandrine Le Guillou<sup>1</sup>, Gwenola Tosser<sup>2</sup> and Fabienne Le Provost<sup>1</sup>, (1)INRA UMR1313 Animal Genetics and Integrative Biology, Jouy en Josas, France, (2)INRA GenPhySE, Castanet-Tolosan cedex, France

Genomic selection is now widespread in bovine species, leading to the selection of animals through their Estimated Breeding Value (EBV), taking into account relevant traits. The addition of causal mutations to anonymous genetic markers could improve EBV accuracy. To participate in this improvement in a dairy context, genetic variants of microRNAs expressed in mammary gland and localized in dairy QTL were studied in bovine.

Starting from millions of genetic variants from whole genome sequencing data, we selected those i) in a genomic region significant for dairy traits and ii) in a microRNA expressed in mammary gland. Three of them were validated thanks to GWAS data, with a validated link between genotype and phenotype. Biological impacts of the validated variants were analyzed according to their expected effect. The expression level of the microRNA was studied if its biogenesis was thought to be impacted, and the expression levels of targeted mRNAs was studied if the impact was expected on the microRNA/mRNA recognition. Notably, modifications of targeted mRNAs expression levels were observed, emphasizing the impact of a single nucleotide change in the mRNAs recognition.

These steps lead to an integrated pipeline for the analysis of microRNA genetic variants. Thanks to its validation through the achieved results, the developed approach will be applied to ovine and caprine datasets.

## **W193: Cattle/Sheep/Goat 1**

### **Harnessing Endogenous Repair Mechanisms for Targeted Gene Knock-in during Pre-Implantation Development of Bovine Embryos**

Joseph R Owen, UC Davis, Davis, CA

The use of gene-editing tools, such as the CRISPR/Cas9 system, could be used to introduce useful genetic traits into cattle breeding programs. However, current methods for harnessing these tools for whole gene insertions have been limited to alterations in cell culture followed by somatic cell nuclear transfer (SCNT) cloning. This method is problematic as SCNT cloning is inefficient, and edits are made in existing germplasm, rather than the next generation. One alternative is CRISPR-mediated modification of *in vitro* fertilized embryos by direct cytoplasmic injection. We have developed a method, which utilizes the homology mediated end-joining (HMEJ)-approach to facilitate homology directed repair (HDR) via direct cytoplasmic injection of *in vitro* produced bovine embryos. Four *in vitro* transcribed guide-RNAs (gRNA) targeting the X-chromosome were tested by directed cytoplasmic injection, resulting in mutation rates of 37.5%, 28.6%, 81.8% and 40.0%, respectively. Using the gRNA with the highest mutation rate, we developed two donor vectors to compare knock-in efficiency using the HMEJ-approach compared to the homologous recombination (HR)-method. Using the HMEJ-approach, we showed a significantly higher rate of gene knock-in as compared to the HR-method (37.1% vs. 13.6%;  $p < 0.05$ ). Along with the

increased rate of gene insertion, more than a third of the knock-in embryos (38.8%) were non-mosaic. This method harnesses the ability for HDR by direct injection of oocytes, utilizing the HMEJ-method for targeted gene knock-in of embryos. This approach could enable precise germline editing of the next generation of elite bovine embryos.

#### **W194: Cattle/Sheep/Goat 1**

##### **Estimating Partial Lactation Yield Heritabilities and Genetic Correlations in a Dairy Goat Herd using a Restriction Enzyme-Reduced Representational Sequencing Derived Relationship Matrix**

**Christian J Posbergh**<sup>1</sup>, Mary Wheeler<sup>2</sup>, Ken G. Dodds<sup>3</sup>, Rudiger Brauning<sup>3</sup>, Heather J Huson<sup>1</sup>, Hayley J. Baird<sup>3</sup> and John C. McEwan<sup>3</sup>, (1)Cornell University, Ithaca, NY, (2)AgResearch, Ruakura Research Centre, Hamilton, New Zealand, (3)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand

Restriction enzyme-reduced representational sequencing (RE-RRS) is an attractive option for introducing genomic selection within herds that have little pedigree information. RE-RRS was implemented in a large dairy goat herd in Australia to increase selection for milk volume and cheese yield. The herd has 4 separate mating dates at approximately 3-month intervals, but the current evaluation system requires waiting a full lactation (290 days) before phenotypic information on a doe's own milk yield performance is incorporated. This in turn limits the accuracy of the evaluations on the current kids and requires keeping kids longer before a selection decision is made. Daily milk yields are recorded on farm and could be used to provide some information prior to the completion of a full-length lactation. The objective of this study was to determine if partial lactation yields could be utilized for the selection of increased 290-day milk yield. Using daily milk yields we calculated partial lactation yields for the first 30, 50, 70, 90, 110, 130, and 150 days of lactation. Univariate analysis using the full genomic relationship matrix estimated heritabilities for partial lactation yields ranging from 0.13-0.21, increasing with length of measurement. Running a bivariate analysis, with a reduced genomic relationship matrix, genetic correlations between partial lactation yields and 290-day yield were estimated between 0.86-0.96, again increasing as the partial lactation yield days increased. This study shows that partial lactation yields in Australian dairy goats can be utilized for earlier and more accurate selection of higher milking does.

#### **W195: Cattle/Sheep/Goat 1**

##### **Profiling Host Transcriptome during Bovine Leukemia Viral Pathogenesis**

**Katy Kesler**<sup>1,2</sup>, Tawnie Dejong<sup>1</sup>, Tasia M. Taxis<sup>1</sup>, Kelly Sporer<sup>2</sup> and Casey Droscha<sup>2</sup>, (1)Michigan State University, East Lansing, MI, (2)CentralStar Cooperative

Prevalence of bovine leukemia virus (BLV) has now surpassed 40% in US dairy cattle. Although BLV is often asymptomatic, progression results in persistent lymphocytosis leading to immunosuppression and decreased longevity. Current BLV diagnostics include detecting anti-BLV antibodies (ELISA) and determining proviral load (PVL) by quantifying viral DNA (qPCR) to screen and stratify disease status aiding in BLV management for producers. Yet, mechanisms of BLV transmission and pathogenesis remain ambiguous. The BLV provirus encodes anti-sense transcripts and microRNAs (miRNA) eliciting post-transcriptional modulation of host Bcell proliferation and maturation as a strategy for viral propagation. MicroRNAs are reliably detected in blood and have been shown to correlate with PVL ( $r=0.928$ ), but miRNA functional relevance in BLV pathogenesis remains unclear. The purpose of this study was to elucidate the mechanisms of BLV progression, by developing miRNA profiles of animals in a progressive state of BLV infection (increasing PVL). Blood samples were obtained from a local cooperative dairy herd. BLV status was analyzed using ELISA and qPCR assays. Sequencing of mRNA and small non-coding RNA was performed using Illumina TruSeq library preparation technology. TaqMan qPCR assays were performed for validation of profiled miRNAs. The objective of this study is to describe transcriptional changes that occur during BLV pathogenesis to elucidate mechanisms behind BLV disease progression. Increased knowledge about the dynamics of BLV disease will enable improvement in diagnostic and prevention strategies to reduce BLV nationwide and increase the sustainability and profitability of the dairy industry.

## W196: Cattle/Sheep/Goat 1

### Differential Gene Expression and Identification of Growth-Related Genes in the Pituitary Gland of South African Goats

Keabetswe Tebogo Ncube<sup>1</sup>, Edgar F. Dzomba<sup>2</sup>, Benjamin D. Rosen<sup>3</sup>, Steven G. Schroeder<sup>3</sup>, Curtis P. VanTassell<sup>3</sup> and Farai C. Muchadeyi<sup>4</sup>, (1)Agricultural Research Council, Pretoria, South Africa, (2)University of KwaZulu-Natal, Pietermaritzburg, South Africa, (3)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (4)ARC-Biotechnology Platform, Pretoria, South Africa

Growth is an economically important trait in goat production. This study investigated differentially expressed genes and associated growth-related genes from the caprine pituitary gland transcriptome of South African indigenous goat breeds of varying growth performance. Tissue samples were collected from three village ecotype goats and three SA Boer goats all raised in similar conditions simulating intensive commercial production systems. Within breed differences investigated by comparing differential gene expression between three village goats raised under extensive conditions (on-farm in Pella village farming community) and three village goats raised under intensive commercial production system. Total RNA isolated from the pituitary gland of 36 weeks old animals ( $n=9$ ) was sequenced and an average of 25,044,183 trimmed and quality-controlled reads were mapped to the goat genome (*Capra\_hircus.ARS1.94*) using HiSat2 software. Transcript assembly and quantification from RNA sequence yielded 104 differentially expressed genes for extensive system and 62 for the intensive system at the false discovery rate (FRD) of 0.1% and a fold change of  $>2$  which were retained for downstream analysis. The study revealed growth-related genes such as the *POU3F4*, and *TSHZ1* that were highly expressed within populations raised under different production systems. Conversely, growth-related genes such as *FGFR2* and *SMPX* genes were highly expressed between breeds raised under similar production system. Ballgown analysis of genes of interest revealed a high expression of *GH1* and *IGF1* in the intensively raised compared to extensive while this gene was also high in the village goats compared to the Boer. The *POU1F1* gene was moderately expressed between and within breeds in both experiments. The differential gene expression data provided insights into genes and molecular mechanisms associated with growth and growth development in goats.

## W197: Cattle/Sheep/Goat 1

### Genes Regulating Calcium Availability and Utilization in Angus Steers may be useful in Identifying Cattle with Reduced Susceptibility to Pulmonary Hypertension in High Altitude Beef Production Systems

Natalie F. Crawford<sup>1</sup>, Kaysie J. Jennings<sup>1</sup>, Angela Cánovas<sup>2</sup>, Juan F. Medrano<sup>3</sup>, Alma Islas-Trejo<sup>4</sup>, R. Dale Brown<sup>5</sup>, Stephen J. Coleman<sup>1</sup>, Scott E. Speidel<sup>1</sup>, R. Mark Enns<sup>1</sup>, Kurt R. Stenmark<sup>5</sup> and Milt G. Thomas<sup>1</sup>, (1)Colorado State University, Fort Collins, CO, (2)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (3)Department of Animal Science, University of California, Davis, CA, (4)University of California-Davis, Davis, CA, (5)University of Colorado at Denver, Anschutz Medical Campus, Aurora, CO

Our understanding of the pathology underlying pulmonary hypertension (PH) of cattle in high altitude ( $> 1,500\text{m}$ ) production systems is limited. Chronic PH can result in abnormal cardiac function and heart failure of susceptible cattle. Therefore, RNA-Seq analysis of cardiac tissues was performed to discover differentially expressed genes between hypertensive and normotensive steers ( $n=7/\text{group}$ ). The RNA-Seq analysis revealed 650 differentially expressed genes ( $P < 0.05$ ). Functional analysis using Ingenuity Pathway Analysis software identified differentially expressed genes related to calcium utilization and availability as important within the gene subset. Quantitative RT-PCR was utilized to validate the expression of 10 candidate genes in cardiac muscle tissues from Angus steers ( $n=10/\text{treatment group}$ ). These genes were selected based on their calculated fold-change differences between the hypertensive and normotensive steers in the RNA-Seq analysis. The selected genes were *ASIC2*, *EDN1*, *FBN1*, *KCNMA1*, *NOX4*, *PLA2G4A*, *RCAN1*, *RGS4*, and *THBS4*. Expression differences ( $P < 0.0055$ ) existed between hypertensive and normotensive steers for *ASIC2*, *EDN1*, *NOX4*, *PLA2G4A*, *RCAN1*, and *THBS4* in right ventricle samples. Additionally, right papillary muscle exhibited expression differences

between hypertensive and normotensive steers for *NOX4*, *PLA2G4A*, *RCAN1*, and *THBS4* ( $P < 0.0055$ ). These results identify and validate differential expression of genes that may be of interest when evaluating the role of calcium regulation in cardiac tissues pertaining to PH status in Angus steers at high altitude.

### **W198: Cattle/Sheep/Goat 1**

#### **Genome-Wide Association Study for Hair Length in Brangus Heifers**

**Kaitlyn M Sarlo Davila**, Department of Animal Sciences, University of Florida, Gainesville, FL

Thermal stress limits beef cattle production and results in a loss of \$370 million in the U.S. annually due to reduced animal performance. A shorter hair coat is a key thermoregulative adaptation that allows cattle to lose heat more efficiently through conductive, convective, and evaporative cooling at the hair-skin interface. The objective of this study was to identify genetic variants associated with the length of the topcoat and undercoat of cattle. Hair samples were collected from the shoulder, 4 inches down from the spine from 1456 heifers in 2016 and 2017. ImageJ software was used to measure hair length. The length of the topcoat and undercoat were evaluated for each individual by averaging five long and five short hairs, respectively. DNA was extracted from blood samples and genotyped with the Bovine GGP F250 array. After quality control, 109,538 SNP were available for association analyses using the univariate procedures of GEMMA that fitted the genomic relationship matrix to account for the genetic covariance among animals. To correct for multiple tests, the Benjamini-Hochberg false discovery rate was constrained to 0.2. Four SNP in the PRLR gene were significantly associated with topcoat length. The SLICK mutation in PRLR has previously been demonstrated to significantly impact hair length in cattle. Seven SNP in the PCCA gene were significantly associated with undercoat length. PCCA belongs to the biotin transport and metabolism pathway. Biotin deficiency has been reported to cause hair loss. These genetic variants may contribute to a shorter hair coat and more thermotolerant animals.

### **W199: Cattle/Sheep/Goat 1**

#### **NIFA Update**

**Lakshmi Matukumalli**, NIFA-USDA, Kansas City, MO

### **W200: Cattle/Sheep/Goat 1**

#### **NRSP8 Bioinformatics Report**

**James E. Koltes**, Iowa State University, Ames, IA

### **W201: Cattle/Sheep/Goat 2**

#### **The Cattle Rumen Microbiome and its Effect on Methane Emissions and Feed Efficiency**

**Amanda Warr**<sup>1</sup>, Robert D. Stewart<sup>2</sup>, Marc Auffret<sup>3</sup>, Joana Lima<sup>4</sup>, Alan Walker<sup>5</sup>, Rainer Roehe<sup>3</sup> and Michael Watson<sup>2</sup>, (1)The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, (2)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (3)Scotland's Rural College, Edinburgh, United Kingdom, (4)Scotlands Rural College, United Kingdom, (5)The Rowett Institute, Aberdeen, United Kingdom

The cattle rumen microbiome is responsible for the breakdown of human-inedible plant materials, and conversion to energy and nutrients that the host animal needs to thrive.

Recently we have sequenced the whole-metagenome of rumen contents from hundreds of cattle from Scotland, assembling thousands of new bacterial and archaeal strains and species in the process. We have used a combination of technologies, including Illumina, Hi-C and Oxford Nanopore technologies. These assemblies encode millions of novel enzymes and proteins, many predicted to be important for the breakdown of animal feed.

Here we will describe this work, and some of the downstream results such as the likely impact of the rumen metagenome on methane emissions and feed conversion.

## **W202: Cattle/Sheep/Goat 2**

### **Identification of Potential Biomarkers for Disease in the Microbiome of the Peripartum Reproductive Tract and Colostrum of Multiparous Holstein Cows**

**Connor E. Owens**, Haley G. Huffard, Alexandra Irma Nin-Velez, A. Jane Duncan, Katharine F. Knowlton and Rebecca R. Cockrum, Virginia Polytechnic Institute and State University, Blacksburg, VA

The objective of this study was to further characterize the microbiome of the dam's reproductive tract and colostrum at calving and identify potential biomarkers for post-partum disease. Swabs of the posterior vagina were collected 24 h before calving (V; n = 4), colostrum was collected within 1 h after calving (C; n = 6), and placenta was collected within 6 h after calving (P; n = 5). 16S libraries of extracted DNA were created and sequenced via Illumina MiSeq V3, 300 PE. Operational Taxonomic Unit (OTU) clustering was performed in CLC Genomics Workbench (ver. 11.0.1) using 97% Greengenes reference database. PERMANOVA analyses and differential abundance analyses were performed with the main effects of sample location or calf sex. The dominant phylum in all locations was Proteobacteria (V = 58%, C = 96%, P = 48%). The vaginal microbiome was different from the microbiome of colostrum (P = 0.014) and placenta (P = 0.048), but colostrum and placenta did not differ (P = 0.266). There was no difference in the microbiomes based on calf sex (P = 0.855). Of the 47 OTUs that differed between locations, g\_Streptococcus had the greatest difference, being more abundant in the vagina than colostrum or placenta (Log2 fold-change = 15.61). The most abundant OTUs in placenta were g\_Acinetobacter, g\_Corynebacterium, g\_5-7N15, and g\_Stenotrophomonas, while in colostrum were g\_Stenotrophomonas, f\_Pseudomonadaceae, and g\_Yersinia. Further investigation of the relationship with these opportunistic pathogen OTUs in placenta and colostrum and post-partum diseases could provide biomarkers to prevent incidence of disease.

## **W203: Cattle/Sheep/Goat 2**

### **Considerations for Designing Imputation Studies Using Whole Genome Sequence Data: A Study in a Diverse New Zealand Sheep Population**

**Andrew S Hess**<sup>1</sup>, Rudiger Brauning<sup>1</sup>, Michael Lee<sup>2</sup>, Melanie k. Hess<sup>3</sup> and Shannon M. Clarke<sup>4</sup>, (1)AgResearch Limited, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)University of Otago, Dunedin, New Zealand, (3)AgResearch Limited, Invermay Agricultural Centre New Zealand, Mosgiel, New Zealand, (4)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand

Whole-genome sequencing (WGS) provides an in-depth view of the genetic variation that is present in an individual and population but is cost-prohibitive. Imputation to WGS is an invaluable tool to capture variation in silico in greater detail across the genome while minimizing costs. The ability to impute individuals of interest up to WGS is dependent on how well the animals with WGS data capture the genetic variation present in that population. The International Sheep Genomics Consortium (ISGC) has sequenced ~1000 sheep to explore the global diversity in sheep. Of these animals, ~200 sheep are from New Zealand, which were chosen to represent the genetic diversity in New Zealand, while also requiring that SNP array data (50K or 600K) was available. The goal of this study was to use the SNP array genotype data available on these sheep to assess their potential to impute sheep in New Zealand. We used imputation of 3K SNP array genotypes up to ~42K (the overlap between 50K and 600K SNP arrays) to assess expected accuracy of imputation to WGS and identify key factors influencing imputation accuracy. We also identified other individuals within the New Zealand sheep population that may be of interest to sequence to better capture the diversity in this population. This information can be used to narrow the set of individuals with SNP array genotypes (~190,000 available) to a high-confidence set for GWAS, thus reducing the likelihood of spurious associations due to imputation errors and reducing computational burden.

## W204: Cattle/Sheep/Goat 2

### Identification of Selection Signatures in 20 Diverse Goat Breeds

Jan Henkel, Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Domestication and human selection have formed diverse goat breeds with characteristic phenotypes. This process correlated with the fixation of causative genetic variants controlling breed-specific traits within regions of reduced genetic diversity, so called selection signatures or selective sweeps. We performed a screen for selection signatures in 20 genetically diverse modern goat breeds and Bezoar goats. We pooled DNA of 12 animals per breed and sequenced the obtained pools to ~30x coverage. The sequence reads were mapped and single nucleotide variants were called. Using sliding windows of 150 kb, we calculated heterozygosity scores and weighted population pairwise  $F_{ST}$  values. We identified 5,220 windows with significantly reduced heterozygosity (0.8% of all windows) and 1,474 windows with significant  $F_{ST}$  (0.2% of all windows). Adjacent or overlapping windows were further merged. This resulted in 2,239 selection signatures or 1.1% of the total genomic length for the reduced heterozygosity and 847 signatures or 0.4% of the total genomic length for the  $F_{ST}$  analysis. We are currently investigating the identified selection signatures for candidate causative variants and traits. So far, we identified six candidate causative variants for breed-specific coat color phenotypes. Interestingly, all six variants are large structural variants. We furthermore identified a candidate causative variant for large body size. An update on the ongoing efforts to identify causative variants for traits under selection will be given at the conference.

## W205: Cattle/Sheep/Goat 2

### Signatures of Selection for Resistance to *Haemonchus contortus* in Sheep and Goats

Zaira Magdalena Estrada-Reyes<sup>1</sup>, Yoko Tsukahara<sup>2</sup>, Rodrigo Amadeu<sup>3</sup>, Arthur Goetsch<sup>4</sup>, Terry Gipson<sup>4</sup>, Tilahun Sahlu<sup>4</sup>, Ryszard Puchala<sup>2</sup> and Raluca Mateescu<sup>5</sup>, (1)Fort Valley State University, Fort Valley, GA, (2)Langston University, Langston, OK, (3)University of Florida, Gainesville, FL, (4)Langston University, (5)Department of Animal Sciences, University of Florida, Gainesville, FL

Gastrointestinal nematode infection (GNI) is the most important disease affecting the small ruminant industry in U.S. The environmental conditions in the southern United States are ideal for the survival of the most pathogenic gastrointestinal nematode, *Haemonchus contortus*. Host genetic variation for resistance to *H. contortus* allows selective breeding for increased resistance of animals. This selection process increases the prevalence of particular alleles in sheep and goats and creates unique genetic patterns in the genome of these species. The aim of this study was to identify loci with divergent allelic frequencies in a candidate gene panel of 100 genes using two different approaches (frequentist and Bayesian) to estimate  $F_{ST}$  outliers in three different breeds of sheep and goats exposed to *H. contortus*. Our results for sheep populations showed SNPs under selection in *C3AR1*, *CSF3*, *SOCS2*, *NOS2*, *STAT5B*, *TGFB2* and *IL2RA* genes using frequentist and Bayesian approaches. For goats, SNPs in *CD1D*, *ITGA9*, *IL12A*, *IL13RA1*, *CD86* and *TGFB2* genes were under selection. Common signatures of selection in both species were observed in *NOS2*, *TGFB2* and *TLR4* genes. Directional selection was present in all SNPs evaluated in the present study. A total of 13 SNPs within 7 genes of our candidate gene panel related to *H. contortus* exposure were identified under selection in sheep populations. For goats, 11 SNPs

within 7 genes were identified under selection. Results from this study support the hypothesis that resistance to *H. contortus* is likely to be controlled by many loci. Shared signatures of selection related to mechanisms of immune protection against *H. contortus* infection in sheep and goats could be useful targets in breeding programs aimed to produce resistant animals with low FEC.

## **W206: Cattle/Sheep/Goat 2**

### **Copy Number Variation in African Goats**

**Wilson Nandolo**, University of Natural Resources and Life Sciences, Vienna, Austria

Goats are important animal genetic resources in Africa. They are used as moving banks, and are an effective way for improving the livelihoods of small holder farmers in the region. Improving goat production in the region is therefore highly desirable. Genetic improvement is a permanent way of achieving such improvement, and one step towards genetic improvement is understanding various aspects of the structure of genomes, one of which is copy number variations (CNV). CNV are a significant source of variation in the genome, and mapping out the CNV on the genome can aid selection. The aim of this study was to develop a fine-scaled CNV map for African goats. We used sequence data from 182 animals belonging to 34 breeds from multiple African countries. A total of 6,231 global CNV regions (CNVR) were found across all animals, representing 59.2 Mb (2.4%) of the goat genome. About 1.6% of the CNVR were present in all 34 breeds and 28.7% were present in all 5 geographical areas across Africa, where animals had been sampled. The CNVR had genes that were highly enriched in important biological functions, molecular functions, and cellular components including retrograde endocannabinoid signaling, glutamatergic synapse and circadian entrainment. This study presents the first fine CNV map of African goat based on WGS data and adds to the growing body of knowledge on the genetic characterization of goats.

## **W208: Cattle/Sheep/Goat 2**

### **Indian Cattle SNP Discovery**

Subeer S Majumdar<sup>1</sup>, Sarwar Azam<sup>2</sup>, Satya Pal Arya<sup>2</sup>, Pankaj Suman<sup>2</sup>, Dhirender Suttar<sup>2</sup>, Manas Ranjan Praharaj<sup>2</sup>, Neelima Hosamani<sup>2</sup>, Shakti Kumar<sup>2</sup>, **Benjamin D. Rosen**<sup>3</sup>, Curtis P Van Tassel<sup>4</sup> and Ravi Kumar Gandham<sup>5</sup>, (1)National Institute of animal Biotechnology, India, (2)National Institute of Animal Biotechnology, India, (3)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (4)Animal Genomics and Improvement Laboratory, USDA, (5)National Institute of Animal Biotechnology, Hyderabad, India

India has a wide variety of indigenous cattle breeds distributed throughout its agro climatic zones. These are known for their natural tolerance to tropical heat and disease resistance with sustainable milk production. To meet the growing demand of milk in India several crossbreeding programmes were taken up. Every state has its own crossbreeding policy which is agro climatic and breed specific. This resulted in breed dilution of valuable indigenous cattle in their native breeding tract leaving behind very few purebreds in comparison to graded cattle. Also, our indigenous breeds have not been exploited to their potential because of slow genetic gain through generations using conventional breeding policies. With the crossbreds not withstanding harsh climate, being susceptible to tropical diseases and requiring constant input of good managemental conditions, conserving our indigenous cattle genetic resources which are far superior in these aspects is the need of the hour. To strike a balance between the increasing demand for milk at 180 mn tonnes by 2022/ 330 mn tonnes by 2022 and change in environment due to global warming conserving our indigenous cattle breeds becomes all the more important. For conserving these breeds the first and foremost requirement is identification of purebreds. The advent of DNA sequencing and high-throughput genomic technologies together with the automated SNP genotyping resulted in a paradigm shift in identification and selection of animals vis - a - vis the

phenotype under consideration. Though there are several Bovine SNP Chips available in the market, literature shows that these chips would partially cover the *Bos indicus* genome. Further, genotyping of three Indigenous dairy breeds (Sahiwal(19), Tharparkar(17) and Gir(16)) using the available Illumina Bovine HDchip revealed only 40%–50% SNPs, informative for genetic analysis in these cattle breeds. Similarly at BAIF, genotyping of Gir showed only 50% of the SNPs in Bovine 50KChip to be polymorphic. These findings indicate that the available SNP chips truly do not represent our indigenous breeds. The availability of high-density chip with improved mapping resolution across indigenous breeds would further result in better imputation. This would further help in screening the indigenous breeds for their purity vis - a - vis admixture.

## **W209: Cattle/Sheep/Goat 2**

### **Genomic Selection in French Dairy Goats**

**Virginie Clément**<sup>1</sup>, Céline Carillier<sup>2</sup>, Marc Tessier<sup>2</sup>, Rachel Rupp<sup>2</sup>, Gwenola Tosser<sup>2</sup>, Pierre Martin<sup>3</sup>, Christèle Robert-Granié<sup>2</sup>, Isabelle Palhière<sup>2</sup> and Hélène Larroque<sup>2</sup>, (1)Institut de l'Elevage, France, (2)INRA GenPhySE, France, (3)Capgènes, France

The main French dairy goats breeds Alpine and Saanen are selected using a total merit that combines milk production traits (protein and fat yields and protein and fat contents), udder morphology and udder health. Genomic selection has been implemented, thanks to the following steps of research: i) the sequencing of goat genome in 2010, ii) the creation of the IGGC to promote and coordinate the development of a high-density SNP chip enabling the design of the Illumina 50K SNP chip in 2011, iii) the implementation of a reference male population. Specificities of French dairy goats - small reference population size and low linkage disequilibrium – required a suitable statistical model. INRA studies showed that with a Single Step GBLUP, expected gains in accuracy would be of 0.06 for milk yield and 0.14 for fat and protein contents. These results encouraged INRA to implement genomic evaluations, and first GEBVs were published in 2018.

The breeding scheme has been entirely renewed by the breeding organization Capgènes, particularly for AI bucks. The male selection intensity increased and the generation interval was reduced, thanks to the early choice of bucks among a larger number of males born from planned mating and genotyped at one month of age, and thanks to the use of young bucks for AI. Genetic evaluation programs have been adapted to produce early GEBVs as soon as genotyping is available. With these improvements, an increase of genetic progress is expected (20 to 30%) in the coming years, resulting in a rise on the net margin of the breeders. Moreover, new traits can be introduced in breeding goals while maintaining the same efficiency on historical traits. GEBVs for female AI fertility and for semen production are expected for 2020 (for internal use in a first time), and studies are ongoing on longevity, milk persistency and feed efficiency

## **W210: Cattle/Sheep/Goat 2**

### **Individual Breed Genome Assembly to Create the Cattle Pangenome**

**Timothy P.L. Smith**, USDA, ARS, USMARC, Clay Center, NE

The “gold standard” for conducting genomic research in any organism has been to first create a usable draft reference genome. However, it is increasingly recognized that a single reference genome, while very useful, is inadequate to fully describe the extent of genetic variation, even in relatively low-diversity species such as humans. This observation is the basis of the recently-announced Human Pangenome Reference Sequence Project (HPRSP), which aims to generate high-quality assemblies of the diversity of human genomes. Fortunately, advances in sequencing technologies have reduced the cost of creating assemblies to the point where even livestock species could reasonably achieve a similar goal if global resources can be marshaled to this end. The other big challenge to creating a useful Pangenome resource from these assemblies, is a method for representing genomes in such a way as to serve as a comprehensive reference map of genetic variation that includes all DNA segments existing in all members of the species. This latter challenge is a major initiative within HPRSP, and the tools developed

there should be applicable to other species. We propose an initiative within the global cattle genomic research community, to generate reference-quality genomes for as many cattle breeds as possible. We suggest that the best strategy will make use of F1 crosses of breeds using the trio binning approach, and will outline a proposed framework to achieve a cattle pangenome in a time frame to coincide with development of visualization and analysis tools in the human genome community.

## **W211: Cattle/Sheep/Goat 2**

### **Haplotype-Resolved Cattle Genomes**

**John Williams**<sup>1</sup>, Lloyd Low<sup>2</sup>, Rick Tearle<sup>3</sup>, Ruijie Liu<sup>3</sup>, Sergey Koren<sup>4</sup>, Arang Rhie<sup>5</sup>, Derek M. Bickhart<sup>6</sup>, Benjamin D. Rosen<sup>7</sup>, Zev Kronenberg<sup>8</sup>, Sarah B Kingan<sup>9</sup>, Elizabeth Tseng<sup>10</sup>, Francoise Thibaud-Nissen<sup>11</sup>, Fergal J. Martin<sup>12</sup>, Konstantinos Billis<sup>12</sup>, Jay Ghurye<sup>13</sup>, Alex R. Hastie<sup>14</sup>, Joyce Lee<sup>14</sup>, Andy Wing Chun Pang<sup>14</sup>, Michael P. Heaton<sup>15</sup>, Adam M. Phillippy<sup>4</sup>, Stefan Hiendleder<sup>3</sup> and Timothy P.L. Smith<sup>16</sup>, (1)Davies Research Centre, University of Adelaide, Roseworthy, Australia, (2)Davies Research Centre, Roseworthy, Australia, (3)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia, (4)Genome Informatics Section, NHGRI, NIH, Bethesda, MD, (5)NHGRI, NIH, Bethesda, MD, (6)Dairy Forage Research Center, USDA-ARS, Madison, WI, (7)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (8)PacBio, Menlo Park, CA, (9)Pacific Biosciences, Menlo Park, CA, (10)Pacific Biosciences, San Francisco, CA, (11)NCBI/NLM/NIH, Bethesda, MD, (12)European Molecular Biology Laboratory - EBI, Cambridge, United Kingdom, (13)Center for Bioinformatics and Computational Biology, (14)Bionano Genomics, San Diego, CA, (15)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, (16)USDA, ARS, USMARC, Clay Center, NE

Until recently genome assembly approaches have collapsed haplotypes, which introduces errors and does not allow the study of divergent, heterozygous regions. We have developed a haplotype-aware genome assembly pipeline which uses contigs built from PacBio long reads, by the trio-binning approach, then Bionano optical maps and Hi-C interaction maps to create haplotype-resolved, chromosome-level genome assemblies. This pipeline was used to generate high quality, phased, Angus (*Bos taurus*) and Brahman (*Bos indicus*) cattle genomes from a single F1 hybrid. The contiguity and accuracy of the final haplotype-resolved cattle assemblies set a new standard for diploid genomes. The per-base substitution quality values (QVs) for the UOA\_Angus\_1 and UOA\_Brahman\_1 reference assemblies were 44.63 and 46.38, respectively which indicates that the assemblies are more than 99.99% accurate at the nucleotide level. Unlike some of the recent PacBio-based assemblies, which required an additional polishing step with Illumina short reads to correct the high indel error rates, these haplotype-resolved assemblies only required correction of a very small number of coding sequences.

These new cattle assemblies enable precise identification of genetic variants, from single nucleotide polymorphisms (SNPs) to large structural variants (SVs). The analyses of copy number variants (CNVs) by alignment of short read sequences from 38 individuals from 7 breeds to the Brahman and Angus genomes revealed that six gene families with immune related functions are expanded in the indicine lineage. The haplotype resolved genomes have enabled us to create and annotate high-quality bovine X and Y chromosomes. The Brahman X chromosome comprises 146 Mb in 106 contigs with 983 genes, and the Angus Y chromosome which comprises 16 Mb in 67 contigs with 51 unique genes.

A further benefit of haplotype-resolved genomes is that they can be used to better interpret allele-specific expression in diploid transcriptome profiles. Among the PacBio error corrected Iso-Seq (CCS) reads pooled from seven tissues of the sequenced F1 Angus-Brahman hybrid fetus, 3,275,676 reads (55%) were classified as full-length non-concatamer (FLNC) reads. After processing with the isoseq3 software, 193,974 full-length, high-quality (HQ) consensus transcripts were generated. We mapped the HQ transcripts to the Brahman reference and obtained 99,329 uniquely mapped transcripts covering 20,940 non-overlapping loci representing 19,403 genes. The haplotype-resolved genomes allowed us to explore genes with allelic imbalance in expression. All tissues showed evidence of imbalance in allelic expression, which was most pronounced for liver, lung, muscle and placenta, whereas brain, heart and

kidney were less affected. Transcriptome analysis of 5 tissues revealed 54 genes were differentially expressed between males and females, the majority of which were located on the sex chromosomes, 28 genes encoded by the X and 10 by the Y chromosome.

## **W212: Cattle/Sheep/Goat 2**

### **Tools and Resources for Cattle Pangenomics**

**C. Titus Brown**, Department of Population Health and Reproduction, School of Veterinary Medicine, University of California Davis, Davis, CA

## **W213: Cattle/Sheep/Goat 2**

### **Mapping Sequencing Reads to Bovine Breed-Specific Genome Graphs**

**Danang Crysnanto**, ETH Zurich, Zurich, Switzerland

The current bovine reference sequence is a linear consensus sequence derived from paternal and maternal haplotypes of a single highly-inbred Hereford cow. The linear sequence lacks diversity because it does not include allelic variation. Lack of diversity causes reference allele bias, i.e., DNA fragments that contain reference alleles are more likely to align correctly than those containing non-reference alleles. Variation-aware genome graphs may address problems arising from the inadequate representation of current references. We used the existing bovine reference coordinate system (ARS-UCD1.2) as backbone and added sites of variation that were filtered according to dairy (Brown Swiss, Holstein) and dual-purpose (Fleckvieh, Original Braunvieh) cattle breeds to construct breed-specific genome graphs. Mapping accuracy did not differ between the linear reference sequence and genome graphs that were augmented with random variants. However, our results show that read mapping is more accurate to graph-based than linear reference genomes when the graph contains pre-selected variants. Variant prioritization is crucial to achieve high mapping accuracy at tractable computational complexity. Adding common variants improves mapping accuracy; but adding rare variants tends to compromise read mapping accuracy. Read mapping accuracy was the highest for breed-specific graphs i.e., when sequencing reads from Brown Swiss cattle were mapped to a variation-aware graph that was augmented with variants filtered according to the Brown Swiss population. We estimate that the number of erroneously mapped reads can be reduced by 1.5 million for a 35-fold coverage of the cattle genome when a variation-aware reference graph is considered. We anticipate that breed-specific genome graphs that were constructed from highly accurate and continuous breed-specific haplotype-resolved genome assemblies might further reduce mapping errors.

## **W214: Cattle/Sheep/Goat 2**

### **Genomic Resources for Agricultural Animals at NCBI**

**Terence D. Murphy**<sup>1</sup>, Françoise Thibaud-Nissen<sup>1</sup> and the NCBI Genome Annotation Team, (1)NCBI/NLM/NIH, Bethesda, MD

The National Center for Biotechnology Information (NCBI) provides a diverse set of genomic resources for a wide variety of organisms, with a focus on organisms of medical or agricultural relevance. NCBI's assembly resource provides systematic access to over 100,000 genome assemblies with stable sequence content to help promote future data exchange and compatibility. NCBI further provides robust automated gene predictions supplemented with manual expert curation for a subset of genomes, including sheep, pig and cattle, as part of NCBI's RefSeq project. The latest sheep assembly, Oar\_rambouillet\_v1.0 from the BCM-HGSC, provides the foundation for a substantially improved genome annotation that integrates diverse set of short and long-read (IsoSeq) RNA-seq data available in the public archives. Annotations from NCBI's pipeline for pig, cattle, sheep, goat, and over 550 other eukaryotes are available in NCBI's Gene resource, BLAST databases, and Genome Data Viewer (GDV). Gene and GDV also provide access to other genomic information including orthologs, RNA-seq expression data, and whole genome alignments to previous assembly versions or assemblies from different breeds, strains, or subspecies. Further information about NCBI's annotation resources and GDV

is available at: [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/) and <https://www.ncbi.nlm.nih.gov/genome/gdv/>.

## **W215: Cattle/Swine**

### **Beyond the Genome: What Having a Platinum Quality Brahman (*Bos indicus*) Genome Can Teach Us about What Makes a Useful Reference Genome**

**Elizabeth Ross**, Queensland Alliance For Agriculture and Food Innovation - University of Queensland, Brisbane, QLD, Australia

Brahman cattle are a member of the Zebu (*Bos indicus*) subspecies of cattle. Despite their known genetic distance from taurine cattle genomic research has historically utilised the taurine reference genome, as it was the only available option. With the rapid reduction in price of long read sequencing this is no longer the case. We have built a platinum quality reference genome for Brahman cattle using 195GB of PacBio data, Hi-C and Chicago scaffolding and 160GB of Illumina short reads for polishing. The resulting assembly contains all chromosomes in single scaffolds, with only 330 gaps. However it is not only the accuracy and completeness of a reference genome that makes it a valuable research tool. Additional long read Oxford nanopore data from 14 Brahman including the offspring of the reference animals allows us to investigate structural variants in Brahman cattle. Additionally, the interaction of topologically associating domains (TADs) with expression data from blood, and full length transcript Isoseq data from 11 of the reference animal's tissues allow the investigation variation not only of the genome, but of the genome function.

## **W216: Cattle/Swine**

### **Identification, Validation, and Management of Mendelian Traits in Livestock Breeding Programs**

**John B. Cole**, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, Christine F. Baes, University of Guelph, Guelph, ON, Canada and Christian Maltecca, NC State University, Raleigh, NC

The number of known Mendelian traits has increased as the cost of genotyping has decreased rapidly in all livestock species. In 2011, VanRaden and colleagues reported five previously unknown recessive defects in the Brown Swiss, Holstein, and Jersey breeds based on a deficiency of expected homozygotes and undesirable effects of those loci on fertility and stillbirth traits. As the number of genotyped animals has increased globally, exceeding 3.5 million in the US dairy population alone, new defects have been reported in many breeds. A number of loci associated with polled status and coat color also have been identified, and the same methodology could be used to track desirable milk protein or fatty acid profiles. In the US, 25 haplotypes currently are tracked, and carrier status is reported for all genotyped animals. Several haplotypes affecting fertility have been identified in commercial pig populations, including a mutation in *BMPER* associated with mummified piglets. Causal variants are known for most of these haplotypes, but identification efforts are hampered by the quality of annotation for livestock genomes and validation of causal variants often is based on statistical, rather than biological, association. The rapid growth in the number of Mendelian traits being reported has raised many questions. Three of the most common will be addressed in this presentation: 1) how best to manage haplotypes in a population, and 2) how to validate putative causal variants with lower-quality genome annotations than model species, and 3) if the rate of *de novo* mutations is increasing.

## **W217: Cattle/Swine**

### **Pig Genome Evolution: *De Novo* Assembly and Annotation of the Genome of the Visayan Warty Pig (*Sus cebifrons*)**

Langqing Liu, Ole Madsen, Qitong Huang, Martijn F.L. Derks, Hendrik-Jan Megens, Richard P.M.A Crooijmans and **Martien A.M. Groenen**, Wageningen University & Research, Animal Breeding and Genomics, Netherlands

The Visayan warty pig (*Sus cebifrons*) is a critically endangered species in the pig genus (*Sus*). We generated a de novo genome assembly of Visayan warty pig using linked-read sequencing (10x Genomics) and Hi-C-based chromatin interaction maps (Dovetail Genomics). The resulting chromosome level assembly consists of 17 chromosomes and 1,568 unplaced scaffolds. The scaffold N50 length is 141,782,568 bp while the contig N50 length is 159,621 bp. BUSCO analysis suggests the assembly is highly complete, with 95.7% of BUSCOs complete. The alignment of the *Sus cebifrons* and the *Sus scrofa* (Sscrofa11.1) assemblies reveals a high degree of collinearity. We reveal chromosome rearrangements during Suidae evolution, involving chromosome fusion and fission events involving *Sus scrofa* chromosomes 13, 14, 16 and 18. Annotation was done using Braker2 and included RNA-seq data from 6 different tissues. Identification of repetitive sequences was done using Repeatmasker and Repeatmodeler. In total, 38,300 protein-coding genes and 788.86 Mb of repetitive sequences were identified. This highly contiguous assembly provides a comparative framework to common pigs (*S. scrofa*), extending our understanding of the genome evolution. A comparison of the HiC data of *S. cebifrons* and *S. scrofa* revealed a high degree of conservation of the 3D chromatin structure in these two species that diverged approximately 4 million years ago.

## **W218: Cattle/Swine**

### **Development of a Low-Pass Sequencing Platform to Support Genomic Selection in Livestock**

**Joe Pickrell**<sup>1</sup>, Jeremiah H. Li<sup>2</sup>, Jesse L. Hoff<sup>2</sup>, Larry A. Kuehn<sup>3</sup> and Warren M. Snelling<sup>3</sup>, (1)Gencove, New York, NY, (2)Gencove Inc, New York, NY, (3)USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, Clay Center, NE

Genotyping platforms for breeding applications are required to be cost-effective and high throughput; the general approach to meeting these criteria involves identifying a small subset of variant sites to be measured on a genotyping array. Low-pass whole genome sequencing combined with genotype imputation can, in principle, allow for the measurement of the entire genome in a cost-effective manner. In order to build a platform for low-pass sequencing in cattle, we constructed a cattle haplotype reference panel by re-analysis of 946 genome sequences from 14 cattle breeds. All sequences were jointly genotyped using GATK 4, producing a total of around 60M markers after filtering. The identified sites contain >95% of the sites on the BovineSNP50 and BovineHD assays. We evaluated the accuracy of genotype imputation from low-pass sequencing data using this reference panel by downsampling sequencing data to coverage levels ranging from 4x to 0.4x. At a sequencing depth of 0.4x, concordance between imputed low-pass sequencing data and directly genotyped sites was over 99% in most B. taurus breeds and 97% in Brahman. Finally, in a sample of 100 cattle we compared the genomic relationship matrix and genomic predictions generated using imputed low-pass sequencing data and the SNP50 array, finding them to be effectively equivalent. We describe future prospects for genomic prediction using low-pass sequencing, including greater compatibility as a method across populations, and flexible integration with historic and future markers sets. We suggest that re-analysis of sequencing data in breeding animals will allow for continuous updates of the marker training set used to optimize genomic prediction accuracy, on a population- and trait-specific basis.

## **W219: Cattle/Swine**

### **Analyzing Population Signatures of Cattle Antibody Repertoires**

**Yana Safonova**<sup>1</sup>, Luke Kramer<sup>2</sup>, Sung Bong Sing<sup>3</sup>, James M. Reecy<sup>2</sup>, Pavel Pevzner<sup>4</sup> and Timothy P.L. Smith<sup>5</sup>, (1)University of California San Diego, La Jolla, CA, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)USDA, (4)Computer Science Dept., UCSD, La Jolla, CA, (5)USDA, ARS, USMARC, Clay Center, NE

*Repertoire sequencing* (or *Rep-Seq*) technologies enable high-throughput scanning of antibody repertoires and open up new horizons for analyzing properties of adaptive immune system with downstream applications in the developing antibody drugs, estimating vaccine efficacy, and analyzing genomic diversity of immunoglobulin loci. Rep-seq made possible sequencing antibody repertoires of non-human species, including agriculturally important animals like cows. Immunoglobulin heavy chain

locus of the cattle genome contains unusually long diversity (D) genes (~150 nt in cows vs ~30 nt in humans) that contribute to production of ultra-long CDR3s. Although recent studies have demonstrated the biomedical potential of antibodies with ultra-long CDR3s in treatment of HIV infection, their role in cattle antibody response remains unclear. In this work, we analyze cattle antibody repertoires taken before and after vaccination against BRDC (Bovine Respiratory Disease Complex) from 200+ individuals. We inferred alleles of variable (V) genes using Rep-seq data and found that genomic variations in V genes are strongly associated with the fraction of ultra-long CDR3s and the antibody production before and after the vaccination.

## **W220: Cattle/Swine**

### **Anthrax Toxin Receptor 1 Knockout Pigs are Protected from Senecavirus A Infection**

**Paula R. Chen**, University of Missouri, Columbia, MO

Senecavirus A (SVA) has been the cause of numerous cases of vesicular disease in swine across the world in recent years. Studies investigating the oncolytic properties of SVA in humans revealed anthrax toxin receptor 1 (ANTXR1) as its receptor. The objective of the current study was to determine if ANTXR1 functioned as the receptor for SVA in pigs by employing the CRISPR/Cas9 system to edit exon 1 and create a premature stop codon. Two founder *ANTXR1* knockout pigs and two age-matched wild type pigs were challenged with SVA. Serum, fecal swabs, and nasal swabs were collected throughout the duration of the study. Presence of viral nucleic acid was determined by PCR, and SVA antibody responses were assessed. *ANTXR1* knockout pigs exhibited distinct anatomical features, including frontal bossing and wide, short statures, which is characteristic of GAPO syndrome in humans. The knockout pigs did not develop vesicular lesions while the wild type pigs had coronary band lesions after SVA infection. Moreover, SVA nucleic acid was not detected in serum from either *ANTXR1* knockout pig, but virus was present in fecal and nasal swabs of one knockout pig. The same pig demonstrated evidence for production of SVA-specific antibodies; however, both knockout pigs did not exhibit virus neutralizing activity. Because founder pigs created by microinjection of the CRISPR/Cas9 system can have mosaic genotypes, a study on F1s is warranted. Overall, knocking out *ANTXR1* appears to confer protection against SVA infection in pigs, and modulation of this region may be needed to correct the phenotype associated with the edit.

## **W221: Cattle/Swine**

### **Analysis of Multi-Trait and Multi-Omics Data Supports Cattle as a Model Species for Large Scale Genome-Wide Studies**

**Ruidong Xiang**, University of Melbourne, Melbourne, VIC, Australia and Michael E. Goddard, The University of Melbourne, Parkville, Australia

While large genomic datasets generated from humans significantly improved the knowledge of genomics, it is desirable to have another species with very large sample size to test the generality of the findings in humans. Cattle is a possible model as there are 1.47 billion cattle worldwide and millions are being genotyped or sequenced as well as phenotyped through commercial breeding. The FAANG consortium started to provide large functional datasets on farm animals including cattle. We present an integrative analysis using 1) functional datasets including metabolic quantitative trait loci (mQTLs), expression eQTLs, and CHIP-seq peaks in over 400 cattle, 2) evolutionary datasets including selection signatures estimated using the 1000-Bull Genome database and conserved sites across 100 vertebrate species and 3) population datasets including 17.7 imputed million sequence variants in 44,000+ Australian dairy bulls and cows with 34 complex traits. With novel analytical approaches, we report consistent results with findings in humans. For example, variants within sites conserved across species had the largest contribution to many complex traits. Also, eQTLs substantially contributed to variants in complex traits. However, we also report results novel to findings in human genomics. mQTLs had stronger contribution to complex traits than all types of eQTLs. Also, variants with different linkage disequilibrium properties had the same contribution to complex traits. Variants under artificial selection had limited contributions to complex traits. Our study demonstrates that the increasing amount of

genomic and phenotypic data makes the cattle model a robust and critical resource for testing genetic hypotheses for large mammals.

## **W222: Cattle/Swine**

### **Using SNP Weights Derived from Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs**

**Brittney N. Keel**, Warren M. Snelling, Amanda K. Lindholm-Perry, William T. Oliver, Larry A. Kuehn and Gary A. Rohrer, USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE

The 'large  $p$  small  $n$ ' problem has posed a significant challenge in the analysis and interpretation of genome-wide association studies (GWAS). The use of prior information to rank genomic regions and perform SNP selection could increase the power of GWAS. In this study, we propose the use of gene expression data from RNA-Seq of multiple tissues as prior information to assign weights to SNP, select SNP based on a weight threshold, and utilize weighted hypothesis testing to conduct a GWAS. RNA-Seq libraries from hypothalamus, duodenum, ileum, and jejunum tissue of 30 pigs with divergent feed efficiency phenotypes were sequenced, and a three-way gene x individual x tissue clustering analysis was performed, using constrained tensor decomposition, to obtain a total of 10 gene expression modules. Loading values from each gene module were used to assign weights to 49,691 commercial SNP markers, and SNP were selected using these weights, resulting in 10 SNP sets ranging in size from 101 to 955 markers. Weighted GWAS for feed intake in 4,200 pigs was performed separately for each of the 10 SNP sets. A total of 36 unique significant SNP associations were identified across the ten gene modules (SNP sets). For comparison, a standard unweighted GWAS using all 49,691 SNP was performed, and only 2 SNP were significant. None of the SNP from the unweighted analysis resided in known QTL related to swine feed efficiency (feed intake, average daily gain, and feed conversion ratio) compared to 29 (80.6%) in the weighted analyses, with 9 SNP residing in feed intake QTL. These results suggest that the heritability of feed intake is driven by many SNP that individually do not attain genome-wide significance in GWAS. Hence, the proposed procedure for prioritizing SNP based on gene expression data across multiple tissues provides a promising approach for improving the power of GWAS. USDA is an equal opportunity provider and employer.

## **W223: Cattle/Swine**

### **Matching Cow's Genetics to the Environment using Genomics**

**Jared E. Decker**<sup>1</sup>, Troy N. Rowan<sup>1</sup>, Sara M Nilson<sup>1</sup>, Harly Durbin<sup>2</sup>, Camila Urbano Braz<sup>3</sup>, Robert D. Schnabel<sup>4</sup> and Christopher M. Seabury<sup>5</sup>, (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)University of Missouri Animal Sciences, Columbia, MO, (3)University of Missouri, (4)Division of Animal Sciences, Informatics Institute, Columbia, MO, (5)Texas A&M University, College Station, TX

Cattle poorly adapted to their environment result in lost revenue and jeopardize the stability of the food supply. Genomic data now allows us to rigorously analyze local adaptation and genotype-by-environment interactions. Use of environmental region-specific genomic predictions would avoid the generation of animals that will not thrive. We used selection scans for local adaptation, genotype-by-environment genome-wide association analyses, creation of hair shedding genomic predictions and ecoregion-specific genomic predictions of growth traits to characterize and predict local adaptation in beef cattle. Analyzing ~40,000 cattle from three breed associations with ~850,000 high-accuracy imputed SNPs, we used novel selection mapping methods to identify genomic loci responsible for adaptation. Among the three data sets, we identify 114 annotated genes as responding to selection to local adaptation. Functions related to the immune system, apoptosis, axon growth, and the circulatory system appear to be enriched. We also used GxE GWAS and vGWAS (differences in variance rather than difference in means) models to identify genotype-by-environment associations. For example, we identify genes that contribute to lighter birth weights in hotter climates. In cooperation with 74 producers across the United States, over 12,000 cattle were scored on a scale of 1-5 for the early-summer hair shedding phenotype in 2016, 2017, 2018, and 2019. Participating cattle were genotyped using the GGP-F250 SNP

panel, which contains ~170,000 candidate functional variants and ~30,000 imputation SNPs. Genomic breeding values were generated with a repeated records model using these phenotypes. Further, we identified loci with large allele substitution effects for hair shedding. When genotype-by-environment interactions exist, ranking animals using a regional genetic evaluation will be different from national cattle evaluations. We developed ecoregion-specific genomic predictions using a multivariate model in which phenotypes from different regions were fit as separate dependent variables. Genetic correlations between regions were moderate, indicating substantial re-ranking between environmental regions. Genomic loci with a large effect in one region may have little effect in a different region. These genomic predictions will allow rapid identification of cattle best suited to an environment.

## **W224: Challenges and Opportunities in Plant Science Data Management - an International Workshop**

### **Relieving the Curation Bottleneck: A Multi-Pronged Approach**

**Eva Huala**, Phoenix Bioinformatics, Fremont, CA

Manual curation of research literature generates high quality structured datasets that tie experimental data on gene function to gene sequences, but this effort is quite resource intensive. Extraction of all available experimental data on plant gene function from the literature through expert manual curation would require a significant investment of funding and time. More efficient and scalable ways to get this important work done are strongly needed. Three different approaches will be discussed: community curation via a simple and generic user interface, curation at the level of gene families containing many species rather than species by species and gene by gene, and new ML and NLP approaches with the potential to increase efficiency and partially automate the manual curation process. Solving the manual curation bottleneck will require a long time horizon and consistent, steady support, and addressing sustainability issues will also be a necessary part of the solution.

## **W225: Challenges and Opportunities in Plant Science Data Management - an International Workshop**

### **Larger Datasets, Faster Query Times. Improving the Performance of ExpVIP, an Expression Browser.**

**Ricardo Humberto Ramirez Gonzalez**, John Innes Centre, Norwich, United Kingdom

Since the publication of the initial Chinese Spring survey sequence in 2014, the amount and quality of genomic resources has increased significantly. Manual analyses are impractical and unfeasible. We have cooperated in the development of wheat genomic resources to streamline research relevant to wheat breeding and related technology.

To explore gene expression patterns of interest we have developed expVIP available at [www.wheat-expression.com](http://www.wheat-expression.com). It contains data collected from over 1,000 publicly available RNA-Seq samples and integrated so that they can be queried simultaneously to enable comparison across studies. The set of genes to study can be selected from a region within a QTL, a network of co-expressed genes, or based on any user defined criteria. This tool allows wheat researchers and molecular breeders to know when and where any gene of interest is expressed in wheat plants by leveraging those ~1,000 RNA-Seq datasets.

To be able to store and access this data in a user-friendly way the database, the querying mechanisms and the visualization techniques have gone through several rounds of optimizations. During the update to include the data published for “[The Chinese Spring genome of hexaploid wheat across tissues and cultivars](#)” (Ramirez-Gonzalez et al., 2018) we doubled the number of samples and quadrupled the amount of data, which slowed down the original system. This talk I will discuss how to keep a balance between having a viable system in a reasonable time while keeping it usable.

## **W226: Challenges and Opportunities in Plant Science Data Management - an International Workshop**

### **Tracking Data Linkage for Intelligent and Responsible Reuse**

**Hugh Williamson**, University of Exeter, Exeter, IA

Facilitating the movement of data across sites and communities of users continues to pose enormous logistical, scientific and ethical challenges, given the variety of conceptual backgrounds, material environments and social landscapes in which data are produced, evaluated and traded. Key to the successful and critically-informed reuse of data is understanding their histories, from initial provenance through subsequent processing and linkage, and on building this information into data infrastructures, analytical tools and systems of governance. Situated in the historical, philosophical and social studies of science, this talk draws on extensive, qualitative empirical studies of data curation and reuse across the biological and biomedical domains, carried out in the last five years with funding from the European Research Council and, currently, the Alan Turing Institute in London. These studies provide, firstly, a groundwork for conceptualising the changing epistemological status of data as they travel and, secondly, an exploration of the conditions under which data can be effectively linked, taking into account the institutional and infrastructural landscape of scientific research. Based on this work, we address several key challenges to understanding and facilitating the movement and linkage of data in plant science, and highlight the crucial role of data curators in ensuring modes of data re-use that are sustainable, reliable and trustworthy. We conclude by presenting an ongoing effort to map data linkage efforts in plant-related research, the PlantDBMap.

## **W227: Challenges and Opportunities in Plant Science Data Management - an International Workshop**

### **Breedbase: A Digital Ecosystem for Plant Breeding of Root and Tuber Crops and Beyond**

**Guillaume J. Bauchet**, Boyce Thompson Institute, Ithaca, NY

More than 300 million people below the poverty line in developing countries depend on root, tuber and banana (RTB) crops for food and income. National and International research centers (ie: NARO Uganda, TARI Tanzania, CGIAR centers (CIP, IITA)) working on the RTB crops like cassava, sweet potato, yam, banana and plantain.

Crop breeding experiments are data intensive and long lasting, yielding thousands of phenotype and genotype datasets over years. Accurate breeding decision requires effective data management for collection, storage, analysis.

Breedbase instances (ie: cassavabase.org, sweepotatobase.org) are web-based databases allowing storage and analysis of both phenotypic and genotypic data. Breedbase stores trait ontologies, crosses, pedigrees, images, genotyping and sequencing data. Decision support tools are provided for phenotyping, breeding management and trial analysis (ie: Genomic Selection).

To ensure data continuum, [Breedbase](#) supports barcode based digital data acquisition using PhenoApps (<http://phenoapps.org/>) such as Fieldbook for field data collection or Coordinate (genotype tissue sampling). Breedbase interfaces with the [Breeding API](#) (BrAPI) that defines standard data objects and methods for exchanging data which allows individual software components to talk to one another. Breedbase offers plant breeding communities a digital ecosystem to support their experimental activities.

## **W228: Challenges and Opportunities in Plant Science Data Management - an International Workshop**

### **Integrated Data and Cross-Organism Queries with InterMine**

**Yo Yehudi**<sup>1</sup>, Adrián Rodríguez Bazaga<sup>2</sup>, Rachel Lyne<sup>1</sup>, Daniela Butano<sup>1</sup>, Sergio Contrino<sup>1</sup>, Kevin Herald Reierskog<sup>2</sup>, Justin Clark-Casey<sup>1</sup>, Josh Heimbach<sup>1</sup>, Julie Sullivan<sup>1</sup>, Gos Micklem<sup>1</sup> and InterMine,

(1)Department of Genetics, University of Cambridge, Cambridge, United Kingdom, (2)Department of Genetics, University of Cambridge, United Kingdom

InterMine is an extensible data warehouse designed to effectively integrate biological data. Integrated data can be queried both via a friendly user interface, or programmatically via a REST API. The common data model shared by all InterMines allows cross-organism queries to be scripted with little or no effort using the automatically generated API client code, and scripts can comfortably be run by non-programmers.

Plant-based InterMines cover thale cress, brassicas, legumes (such as chickpea, cowpea, and peanut), oak, wheat, grape, maize, and one of the biggest InterMines, [PhytoMine](#), hosts 93 different plant genomes across 82 species. Tens of InterMines are available, covering, apart from plants, animal model and non-model organism, drug targeting, and more. Integration of data increases visibility, accessibility and usability of data. For example, [LegumeMine](#) integrates expression data from over 6 different sources, all of which can be compared through running just one query. Links to related plant InterMines allows expression data for related genes from other species to also be analyzed.

## **W229: Challenges and Opportunities in Plant Science Data Management - an International Workshop**

### **Software to Streamline Sharing of Agricultural Algorithms and Data**

**David LeBauer**<sup>1</sup>, Kristina Riemer<sup>2</sup> and Chris Schnauffer<sup>2</sup>, (1)University of Arizona, Tucson, AZ, (2)University of Arizona

A future in which data can easily be understood and re-used will enable more efficient scientific discovery. However, consistently curating data is difficult. Scientists face a steep learning curve alongside delayed rewards for the extra effort that is required to publish robust and reliable data products.

I will describe our current efforts to address these challenges with open software and training materials. The first is a class of tools designed for high throughput phenomics pipelines to support metadata ingestion and data curation. Second, tools that translate data among formats used in breeding, crop modeling, and high throughput phenomics, and earth science communities. The third effort focuses on developing tutorials and reusable software templates that embed best practices. Our overarching objective is to enable more efficient and collaborative science by enabling users to more easily share scientific information contained in software and data.

## **W230: ChIP-seq: Challenges and Ideas**

### **Searching for Causal Variants in Functional Regions of the Bovine Genome**

**Claire P. Prowse-Wilkins**, The University of Melbourne, Parkville, Australia

Finding genome variants that cause variation in economic traits would make selection of livestock more accurate. Work in humans and mice have shown that causal variants are enriched in functional regions of the genome, but it is unknown if this is the case in cattle as functional regions are largely uncharacterised. Therefore, identifying functional regions in the bovine genome is an important step in the search for causal variants.

Functional regions are thought to be marked by certain histone modifications which can be assayed in the genome using a technique called ChIP-seq. We report here the results of ChIP-seq for 4 histone modifications (H3K4Me1, H3K4Me3, H3K27ac and H3K27Me3) and one transcription factor (CTCF) in 6 tissues from 3 Holstein dairy cows. Sequence data was aligned to UMD3.1 using BWA mem with an average of 140 million mapped reads per sample. Sample enrichment was checked with deepTools and between 30 thousand to 800 thousand peaks were called with MACS2.

We used ChromHMM to annotate putative functional regions and DiffBind to identify tissue specific peaks. We found correlations between peak heights and gene expression. Finally, we showed that QTL were enriched in some of these regions. This is the first study to describe the location of histone modifications in mammary tissue of dairy cows and can be used to narrow the search space for causal variants.

### **W231: ChIP-seq: Challenges and Ideas**

#### **Exploring the Effect of Signal-to-Noise Ratio on ChIP-Seq Analysis and the Ability of Downsampling to Improve Comparisons between Samples**

**Colin Kern**, Department of Animal Science, University of California, Davis, CA

ChIP-seq is a widely used assay to identify histone modifications and the location of DNA-bound proteins across the genome, which is of particular interest in epigenetic studies. However, the standard protocols used to generate ChIP-seq libraries are highly sensitive to variation at many steps in the protocol. Variation in chromatin shearing, the binding specificity of the antibody used, differences in tissues or cell types, and many other factors can lead to significantly different signal-to-noise ratios between resulting libraries, making it difficult to separate biological from technical differences when comparing results between samples. Using ChIP-seq data generated from eight different tissues across 3 vertebrate species, we explore the effect that variation in data quality has on peak calling using the widely used Macs2 program as well as the results from ChromHMM, a chromatin state predictor that integrates data from multiple ChIP-seq libraries together. The effect that downsampling has on these analysis methods is evaluated, including a novel method of downsampling that incorporates the signal-to-noise ratio which may help to reduce the technical differences between libraries.

### **W232: ChIP-seq: Challenges and Ideas**

#### **Integrative Analyses of Multi Omics Data accelerate the understanding of the Genetic Basis underlying *in vivo* Rumen Development in Dairy Cattle**

**Yahui Gao**, Animal Genomics and Improvement Laboratory, USDA-ARS; The University of Maryland, Beltsville; College Park, MD

The rumen, which contributes directly to feeding efficiency, methane emission and productive performance, makes dairy cattle efficient and productive livestock. It is essential to improve our understanding of the genomic underpinning of rumen development both *in vitro* and *in vivo*. We have reported *in vitro* results before (Fang et al. 2019 BMC Biology 17: 68). Here, we are reported our recent *in vivo* effort. We collected a total of six rumen tissues from Holstein calves before (n=3) and after (n=3) weaning, respectively, to investigate transcriptome, DNA methylation, histone modifications, DNA accessibility and CTCF-binding sites using RNA-Seq, ChIP-Seq, ATAC-Seq and CTCF-Seq technologies. Differentially expressed genes and modified histone marks were detected under the comparison of before and after weaning. Integrative analysis of gene expression and histone mark modification was performed to detect genomic features that were involved in the rumen development. Sequence-based genome-wide association studies for different complex traits of economic importance were further integrated to assess the enrichment of association signals with the detected genomic features. This study provides significant insights into the genetic mechanisms underlying the rumen development before and after weaning, and facilitate the interpretation of biological and genetic data sets, such as GWAS data sets by predicting specific tissues related to specific phenotypes.

### **W233: Citrus Genome**

#### **Genetic Improvement of Citrus Fruits Rich in Anthocyanins and Lycopene through Modern Biotechnology Approaches**

Fabrizio Salonia<sup>1,2</sup>, Angelo Ciacciulli<sup>1</sup>, Helena D. Pappalardo<sup>1</sup>, Lara Poles<sup>1,2</sup>, Carmen Arlotta<sup>1</sup>, Paola Caruso<sup>1</sup>, Maria Patrizia Russo<sup>1</sup>, Giuseppe Russo<sup>1</sup>, Marco Caruso<sup>1</sup> and **Concetta Licciardello**<sup>3</sup>, (1)CREA

Research Centre for Olive, Citrus and Tree Fruit, Italy, (2)Department Di3A, University of Catania, Italy, (3)CREA Research Centre for Olive, Citrus and Tree Fruit, Acireale (CT), Italy

In the last years, several studies emphasized the importance of fruits and vegetables rich in healthy compounds. Anthocyanins and lycopene represent the most interesting and deeply studied pigments due to their protective effects, reducing the risk of several pathologies including cancer and cardiovascular diseases. Among citrus commercial varieties, to our knowledge, there are no references reporting the co-presence of anthocyanins and lycopene in the same fruit. There are several sweet orange (*Citrus sinensis*) varieties with high content of anthocyanins in the flesh, and only a limited number of grapefruit (*C. paradisi*), pummelo (*C. maxima*) and sweet orange varieties contain exclusively lycopene.

Several breeding programs are focused on the generation of hybrids that contain both pigments, even though conventional strategies in citrus are hampered by long juvenility, high heterozygosity, sterility, and nucellar embryony. Moreover, traditional breeding programs do not allow the introgression of single traits without compromising the genetic background that characterizes an appreciated cultivar. Modern biotechnology approaches, such as cisgenesis and genome editing, make possible to introduce (from sexually compatible species), or to edit, single genes preserving all the already selected traits.

Two different strategies should be hypothesized for conciliating the presence of anthocyanins and lycopene in a unique citrus fruit: cisgenesis for Ruby gene in a lycopene-rich cultivar, editing for a biosynthetic or regulatory gene involved in the degradation or accumulation of lycopene in blood oranges. Different approaches are currently tested at CREA to design constructs and choose marker-free vectors, to optimize regeneration protocols and reduce the juvenility phase, to reach the ambitious goal of producing new citrus varieties with enhanced nutritional properties.

## **W234: Citrus Genome**

### **Differences in the Genomes of *Phyllosticta citricarpa* and *Phyllosticta capitalensis***

**Marco Aurélio Takita**<sup>1</sup>, Carolina Munari Rodrigues<sup>1</sup>, Nicholas V. Silva<sup>2</sup>, Marcelo Ribeiro-Alves<sup>3</sup> and Marcos A. Machado<sup>1</sup>, (1)Citrus Center, IAC, Sao Paulo State Secretary of Agriculture and Supply, Cordeiropolis, SP, Brazil, (2)Centro de Citricultura Sylvio Moreira, IAC, Cordeiropolis, SP, Brazil, (3)Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

Citrus are affected by many diseases that are caused by different pathogens. Besides, Citrus also hosts many symbiotic microorganisms in a relationship that may be advantageous for both organisms. The fungi *Phyllosticta citricarpa* is a pathogen that is responsible for citrus black spot, and with *Phyllosticta capitalensis*, an endophytic species, are examples of closely related species with different behaviour in citrus. Both species are biologically associated and share a very similar morphology, and in order to identify genetic differences that could explain their lifestyles, genomes sequencing were carried.

Drafts genomes were assembled with sizes close to 33 Mb for both fungi. They carry 15,206 and 14,797 coding sequences for *P. citricarpa* and *P. capitalensis*, respectively. Enrichment analysis shows that the pathogenic species presents growth and development genes that may be necessary for its pathogenicity. On the other hand, family expansion analyses showed the plasticity of the genome of these species.

Genome evolution seems to be of real importance among the *Phyllosticta* isolates and it is leading to different biological characteristics of these species.

## **W235: Citrus Genome**

### **Diversity Study in Citrus with High-Density SNP Array Data**

**Yoko Eck**, Sergio Pietro Ferrante and Mikeal L. Roose, University of California, Riverside, CA

*Citrus* is one of the most widely cultivated and economically valued fruit tree crops in the world. Given the complex ancestry of Citrus, analysis of germplasm with low-cost, high throughput tools such as

genome-wide SNP arrays can significantly reduce the time breeders take to screen and characterize germplasm and to identify the genes responsible for traits of interest. We analyzed genetic diversity in *Citrus* with high-density SNP genotype data generated by a recently developed SNP genotyping array for *Citrus*, Axiom™ Citrus Genotyping Array (Affymetrix, Inc.) (58K autosomal and 500 Chloroplast SNPs). Concordance analysis of 925 named accessions in the UCR citrus variety collection (CVC) was used to identify clonally-derived/near-identical accessions prior to the diversity analyses. Ho, He, PIC (polymorphic information content) and percent polymorphic markers were calculated using array data with 925 unselected accessions, the reduced set of 396 accessions in which clonal groups are represented by one sample each, and 36 accessions included in the variant discovery panel used to design the array. In addition, we performed PCA, admixture, treemix, phylogeny network analyses and graphical genotyping showing ancestry specific loci in selected accessions. We used the most stringent PolyHighResolution (PHR) loci as classified by Axiom™ Analysis Suite in both analyses. Admixture analysis with 399 accessions shows clustering into citron, mandarin, pummelo, trifoliolate and kumquat/microcitrus/papeda at K=5 and papeda separates into its own cluster at K=7. Our results show high degree of reticulation events in citrus.

## **W236: Citrus Genome**

### **Citrus Genome Database Resources for Citrus Genomics, Genetics, and Breeding Research**

**Dorrie Main**, Department of Horticulture, Washington State University, Pullman, WA

The Citrus Genome Database (CGD, [www.citrusgenomedb.org](http://www.citrusgenomedb.org)) is a genomics, genetics and breeding database for basic, translational and applied citrus research. CGD combines curated data along with search interfaces and a variety of tools for data exploration and visualization. In addition to information about 73 genetic maps, 49,757 markers, and 597 QTLs, CGD also has numerous citrus and *Ca. Liberibacter* genomes, synteny analyses for all the genomes, PathwayCyc data for the citrus genomes, and 23,070 phenotype measurements from GRIN. All the data is searchable and downloadable using customizable search interfaces. Additional tools include the genome browser JBrowse, MapViewer, BLAST+, and the Breeding Information Management System (BIMS), an online system to manage and analyze private breeding data. BIMS works with Field Book, an Android app used to efficiently collect the field data. Data in our sister tree databases TreeGenes and Hardwood genomics can also be searched from CGD, with access to the Genome Database for Rosaceae being added in the near future. CGD is built using the Tripal database platform and is supported by USDA-NRSP10, NSF-PGRP, USDA-SCRI and US Land Grant Universities.

## **W237: Citrus Genome**

### **Suppression of Citrus Innate Immune Defense by Two Effectors from CLAs that Causes HLB**

**Goutam Gupta**, Biolab New Mexico Consortium, Los Alamos, NM

A systems biology shows how two CLAs effectors, P235 and Effector 3, interact with their citrus target proteins and how these inhibitory interactions suppress citrus innate immune defense and promote CLAs infection/HLB pathogenesis. First, we employed affinity purification and LC-MS/MS to identify *in vitro* the putative citrus binding proteins from infected and health Hamlin. Second, we selected the top-ranked citrus binding partners of P235 and Effector 3. For P235, they are- superoxide dismutase (SOD), aspartate protease (AP), glycosyl hydrolase 17 (GH17) and lipid transfer protein (LTP). For Effector 3, they are- kunitz trypsin inhibitor (KTI), aldehyde dehydrogenase (ALDH), elongation factor Tu (Ef-Tu) and lectin. Third, we designed a triple-split-GFP assay to validate the identified top-ranked citrus proteins, indeed, bind *in planta* to P235 and Effector 3. Fourth, we performed both *in vitro* and *in planta* experiments to demonstrate that P235 and Effector 3 can inhibit the functions of their citrus binding partners as determined by *in planta* split-GFP assay. Fifth, we carried out molecular dynamics simulations to predict the pairwise interactions of two citrus CLAs protein complexes: one with P235 and citrus LTP and the other with Effector 3 and citrus KTI. Finally, we introduced site-specific mutations at the contact interface and performed biochemical experiments to test our predictions. Our results indicate that P235 and Effector combine to keep the reactive oxygen species (ROS) at a higher level, block

bacterial clearance by bactericidal proteins, and induce premature programmed cell death (PCD), thereby suppressing citrus immune defense and supporting CLas infection/HLB pathogenesis.

### **W238: Citrus Genome**

#### **citrusgreening.org: An Open Access and Integrated Systems Biology Portal for the Huanglongbing (HLB) Disease Complex**

Mirella Flores-Gonzalez<sup>1</sup>, Prashant S Hosmani<sup>1</sup>, Noe Fernandez-Pozo<sup>2</sup>, Marina Mann<sup>3</sup>, Jodi L. Humann<sup>4</sup>, Dorrie Main<sup>5</sup>, Michelle Heck<sup>6</sup>, Susan J. Brown<sup>7</sup>, Lukas Mueller<sup>8</sup> and **Surya Saha**<sup>1</sup>, (1)Boyce Thompson Institute, Ithaca, NY, (2)University of Marburg, Germany, (3)Cornell University, Ithaca, NY, (4)Washington State University, Pullman, WA, (5)Department of Horticulture, Washington State University, Pullman, WA, (6)USDA ARS, Ithaca, NY, (7)Division of Biology - Kansas State University, Manhattan, KS, (8)BTI, Ithaca, NY

We have created an open access web portal with pathosystem-wide resources and bioinformatics tools for the host citrus, the vector Asian citrus psyllid (ACP) and multiple pathogens including *Ca. Liberibacter asiaticus*. To the best of our knowledge, this is the first example of a database to use the pathosystem as a holistic framework to understand an insect transmitted plant disease. This endeavor integrates and enables the analysis of data sets generated by the community to study the citrus greening disease complex. Users can submit relevant data sets to enable sharing and allow the community to better analyze their data within an integrated system. The portal contains a variety of tools for omics data. Metabolic pathway databases, DiaphorinaCyc and CitrusCyc provide organism specific pathways and can be used to analyze transcriptomics and proteomics results to identify pathways with differentially regulated genes. Psyllid Expression Network (PEN) contains expression profiles of ACP genes from multiple life stages, tissues, conditions and hosts. Citrus Expression Network (CEN) contains public expression data from multiple tissues and conditions for citrus from NCBI. All tools like Apollo/JBrowse, Biocyc, Blast, CEN and PEN connect to a central database containing gene models for citrus, ACP and multiple *Liberibacter* pathogens. The portal also includes electrical penetration graph (EPG) recordings of ACP feeding on citrus and metabolomics data in addition to traditional omics data types with a goal of combining and mining all information related to a pathosystem. The portal includes user-friendly manual curation tools to allow the research community to continuously improve this knowledge-base as more experimental research is published. Bulk downloads are available for all genome and annotation datasets from the FTP site (<ftp://ftp.citrusgreening.org>). The portal can be accessed at <https://citrusgreening.org/>. More information can be found in the preprint at <https://www.biorxiv.org/content/10.1101/868364v1>

### **W239: Citrus Genome**

#### **Developing Huanglongbing Resistant Transgenic Citrus using Tissue-Specific Promoters from Citrus Small Cyclic Amphipathic Peptides (SCAmpPs) Genes**

**Ed W. Stover**, U. S. Horticultural Research Laboratory, USDA-ARS, Fort Pierce, FL

The disease huanglongbing is devastating citrus in many regions, and is caused by a phloem limited bacterium, *Candidatus Liberibacter asiaticus* (CLas). Citrus phloem enriched ESTs were scanned for highly expressed genes, reasoning that their promoters must drive strong phloem-specific expression and could be used for transgenics to target CLAs. High expression phloem-specific genes were identified and found to comprise a highly conserved gene family. These genes encode small ~50 residue precursor proteins that are post-translationally processed, releasing 5-10 residue cyclic peptides, which were dubbed "Small Cyclic Amphipathic Peptides" (SCAmpPs). Using a GUS reporter gene, D35s drives similar expression in all leaf tissues tested, while the phloem SCAmpP promoter (396ss) drives up to 400X higher expression in leaf midribs compared to the lamina, with similar or greater GUS protein activity in midribs than that from D35s. Potential CLas-killing transgenes are being expressed using both D35S and 396ss. Previously studied tissue-specific promoters have provided only modest expression, and it appears that SCAmpPs promoters will provide much greater expression with strong tissue specificity. The function of SCAmpPs remains undetermined. In vitro experiments with several bacterial

species do not indicate that SCampPs are antimicrobial. A hairpin encoding conserved phloem SCampPs sequence of 116 nucleotides has been used to transform Carrizo. Many lines show substantial reduction in SCampPs expression, but no phenotypic differences are yet evident. These plants will be exposed to a variety of biotic and abiotic stresses to determine whether suppression of SCampPs affects tolerance.

## **W240: Climate Change and ICRCGC 1**

### **Harnessing Plant Genomics for Climate Resilience**

**Chittaranjan Kole**, ICAR-National Institute for Plant Biotechnology, New Delhi, India

Since the sequencing of the genomes of the model plant Arabidopsis in 2000 and the crop plant rice in 2002, genomes of over 100 higher plants have been sequenced that includes huge number of crop plants belonging to almost all crop groups including cereals, oilseeds, pulses, fiber crops, fruits, vegetables, ornamentals, plantation crops, forest trees and several industrial crops that cater to the needs of 'sustenance, accessories, luxuries and remediation (SALR)'. Genome sequences of many wild crop relatives (WCR) are also available now. Genotyping by sequencing of indigenous varieties and local landraces of several crops have also been accomplished. In parallel, organellar genomes have been sequenced in a lot of plants. At the same time, functional genomics has progressed through advances in proteomics, metabolomics and transcriptomics studies. On the other hand, phenotyping strategies have also been improvised. All these advances have facilitated precise depiction of traits and discovery of genes, particularly those related to climate change adaptation, and also detection of the molecular markers tightly linked to them. It is, however, imperative to pay attention to complimentary breeding and farming systems aiming at adaptation to climate change and mitigation of its impacts. The salient research achievements made so far in climate-smart crop genomics will be presented with some success stories, and ideas on some potential concepts and strategies of translational genomics research will be shared. Finally, importance of addressing Food, health, energy and environment (FHEE) security employing the available genetic resources and genomics tools will be highlighted.

## **W241: Climate Change and ICRCGC 1**

### **Genomics of Trait Discovery for Climate Adaptation**

**Robert J. Henry**, University of Queensland/QAAFI, Brisbane, QLD, Australia

The adaptation of agriculture to climate change can be approached using several different strategies. The traits required may be very different depending upon the strategy adopted. We can adapt agriculture by using genetic and management strategies and retain production in the current locations. Moving crops to new production environments is another option. Protected cropping is the extreme response for high value crops. This approach dramatically changes the breeding targets. Efforts to achieve sustainable intensification of horticultural tree crop production illustrates disruptive change in breeding objectives. Conventional breeding is a key way we adapt to climate change with traditional selection processes using the production environment delivering a significant level of adaptation. However rapid climate change will require more proactive intervention protecting against future changes that will impact during the production life of the genotype. Rice is a good example of a crop that is relocating in some regions because of increasingly limiting water resources. The development of more sustainable rice production in higher rainfall environments will require the accelerated breeding of locally adapted rice genotypes. Use of local wild rice populations may be a key resource for delivering this objective. Genome analysis of relevant wild populations has revealed evolutionary relations and ongoing evolution in the wild. This knowledge will facilitate efficient capture of desirable wild alleles for use in plant improvement. Similar strategies will be described with examples from other species.

## **W242: Climate Change and ICRCGC 1**

### **Strengthening Food Security through Climate-Smart Rice Hybrids**

**Jauhar Ali**, Rice Breeding Platform, International Rice Research Institute, Philippines

Food security is going to be an enormous challenge given the global climatic alterations (GCA) threatening the food production channels. All major food crops, including rice production, will be directly affected by GCA such as drought, flood, salinity, submergence, and high/low temperature. Rapid development and deployment of climate-smart rice (inbred parental lines and hybrid) varieties (CSRVs) showing resilience to biotic and abiotic stresses could offer as a mitigation option to GCA. Employing genomics assisted breeding (GAB) helped to develop CSRVs. The successful releases of 27 CSRVs and their deployment over 2.7mha in Asia and Africa have triggered the new strategy to roll out our new Climate-smart rice hybrids (CSRH). The strategy for the development of CSRH addressing the critical target market segment requirements by combining tolerances to biotic stresses like brown planthopper (BPH), green leafhopper (GLH), Tungro, blast, and bacterial leaf blight (BLB), and abiotic stresses like drought, salinity, and flooding will be discussed. Development of high yielding CSRH will be possible through exploitation of heterotic pools and integrating it well with artificial intelligence (AI) technologies. Many Asian and African countries would be largely benefitted with the adoption of high yielding CSRHs. The revitalized IRRI-Hybrid Rice Development Consortium (HRDC) would be an excellent platform to ensure CSRH's wide-scale adoption and create massive socio-economic impacts on farmer livelihoods.

### **W243: Climate Change and ICRCGC 1**

#### **Deciphering Genetic Basis of Responses to Climate Temperature Fluctuations in Rice**

**Kiyosumi Hori**<sup>1</sup>, Daisuke Saisho<sup>2</sup>, Kazufumi Nagata<sup>1</sup>, Shuichi Fukuoka<sup>1</sup>, Jun-ichi Yonemaru<sup>1</sup> and Keiichi Mochida<sup>2,3</sup>, (1)National Agriculture and Food Research Organization, Tsukuba, Japan, (2)Okayama University, Kurashiki, Japan, (3)RIKEN CSRS, Yokohama, Japan

In order to develop climate resilience crops, it is necessary to elucidate genetic basis of plant response to ambient environmental changes. Flowering time is one of crucial factors determining adaptability in plant species. A lot of flowering time genes have been isolated in rice. However, almost of the genes are associated with response to daylength and photoperiod sensitivity. Daylength is stable, while ambient temperature easily fluctuates among years for rice cultivation period. To genetically dissect differentiation of flowering time among years, we developed chromosome segment substitution lines (CSSLs) derived from crosses between *japonica* rice cultivars Koshihikari and Khao Nam Jen. Khao Nam Jen showed a large difference of flowering time (36.7 days) between the earliest and latest years from 2011 to 2019. Difference of flowering time in Koshihikari was 7.8 days between the earliest and latest years. Among the 40 lines of the CSSLs, one line having a Khao Nam Jen segment on chromosome 11 showed 25.4 days of flowering time difference, and another line having a Khao Nam Jen segment on chromosome 3 showed 14.2 days of flowering time difference. Low temperature treatments at the panicle initiation stage delayed flowering time in the two CSSLs. RNA-seq analysis revealed that several flowering time genes were decreased their expressions in the two CSSLs, as compared with Koshihikari. These results suggested that a part of previously isolated genes was involved in the control of flowering time difference under ambient temperature fluctuations and different cultivation conditions.

### **W244: Climate Change and ICRCGC 1**

#### **Efficient Characterization of Tetraploid Wheat Plant Genetic Resources for Wheat Resilience Improvement**

**Marco Maccaferri**<sup>1</sup>, Elisabetta Mazzucotelli<sup>2</sup>, Giuseppe Sciarra<sup>3</sup>, Raj Pasam<sup>4</sup>, Anna-Maria Mastrangelo<sup>5</sup>, Danara Ormanbekova<sup>1</sup>, Martina Bruschi<sup>3</sup>, Elisabetta Frascaroli<sup>1</sup>, Silvio Salvi<sup>1</sup>, Simona Corneti<sup>3</sup>, Sandra Stefanelli<sup>1</sup>, Reem Joukhadar<sup>6</sup>, Francesca Desiderio<sup>2</sup>, Steven Xu<sup>7</sup>, Matthew Hayden<sup>8</sup>, Hakan Ozkan<sup>9</sup>, Benjamin Kilian<sup>10</sup>, Aldo Ceriotti<sup>11</sup>, Curtis Pozniak<sup>12</sup>, Karim Ammar<sup>13</sup>, Hans Brown<sup>14</sup>, Luigi Cattivelli<sup>15</sup>, Filippo Maria Bassi<sup>16</sup> and Roberto Tuberosa<sup>17</sup>, (1)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy, (2)CREA-Research Centre for Genomics and Bioinformatics, Italy, (3)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Italy, (4)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Australia, (5)CREA Research centre for cereal and industrial crop, Foggia, Italy, (6)DEDJTR, AgriBio, Bundoora, Australia, (7)USDA-ARS, Edward T. Schafer Agricultural Research Center, Fargo, ND,

(8)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Vic, Australia, (9)University of Cukurova, Adana, Turkey, (10)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (11)National Research Council (CNR), Institute of Agricultural Biology and Biotechnology (IBBA), Italy, (12)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (13)CIMMYT, Mexico D.F., Mexico, (14)CIMMYT, International Maize and Wheat Improvement Center, Mexico, (15)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d' Arda, Italy, (16)ICARDA, Rabat, Morocco, (17)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, Bologna, Italy

Climate change poses major challenge for global wheat production and thus food security. Drought, heat, salinity and recurrent rust, septoria and fusarium epidemic waves are among the most common abiotic and biotic stresses limiting wheat crop productivity and sustainability worldwide. If adequately characterized, natural variants present in underutilized plant genetic resources (PGRs) are potentially valuable for providing improvements on resilience to abiotic stress and durable resistance to plant pathogens. In this context, Golden-standard reference genomes as well as characterized PGRs are instrumental. Both *Triticum aestivum* and *Triticum turgidum* ssp. *durum* share BB and AA genomes inherited from wild and domesticated emmer throughout complex steps of domestication, spread and migration events from the Fertile Crescent to diverse environments and agro-ecological conditions. With the aim of facilitating germplasm characterization and use, we assembled two comprehensive and complementary tetraploid collections: *i*) the Global Durum Wheat Panel (GDP) and the Tetraploid wheat Global Collection (TGC, Maccaferri et al. *Nature Genetics* 2019). GDP was established through a cooperative effort in the frame of the Wheat Initiative. GDP, currently maintained by ICARDA, was assembled by bringing together the durum wheat cultivated germplasm from more than 50 countries worldwide, including ca. 500 cultivars and 400 landraces, pre-breeding lines and emmer. The Tetraploid wheat Global Collection (TGC), of 1,856 single-seed descent derived-genotypes, represents 11 tetraploid BBAA wheat taxa covering the whole distribution range. The Illumina 90K wheat SNP array was used to characterize both collections. Ca. 20,000 unique, non-redundant, single Mendelian SNP markers that were both genetically and physically mapped were used to obtain a haplo-based map of germplasm. We provide a detailed dissection of the huge reservoir of genetic diversity available in these two tetraploid genetic resources and examples of successful utilization of these resource to conduct GWAS for traits related to improvement of wheat sustainability and adaptation to climate change effects.

## **W245: Climate Change and ICRCGC 1**

### **Perspectives from Integrating Genomic and High-Throughput Phenotyping Tools in Breeding for Climate-Resilience in Bread Wheat**

**Philomin Juliana**, International Maize and Wheat Improvement Center (CIMMYT), El Batan, Mexico

Genomic and high-throughput phenotyping tools offer exciting opportunities for the development of climate-resilient bread wheat varieties, in the face of climate change. Given the escalating demand for the integration of new technologies in increasing climate-resilience, we were interested in determining the value that technologies can add to the current breeding strategies adopted by the International Maize and Wheat Improvement Centre (CIMMYT), that plays a key role in developing stress-resilient global wheat germplasm. To determine the potential of genomic selection (GS) and sparse testing for grain yield (GY), we modeled more than 60,000 GY observations from CIMMYT's key testing site, Obregon, Mexico and sites in Asia, Africa, Europe and the Americas using genomic and genotype x environment (GxE) interaction models and compared the accuracies with baseline pedigree accuracies and environment and line accuracies, respectively to understand the relative advantage of using genomics for predicting GY. Considering different scenarios like (i) the prediction of GY for a subset of lines within a year (ii) sparse-testing GY across years and sites, we observed highly variable accuracies across the different cycles of evaluations. The mean genomic prediction accuracy for GY within years in the Obregon selection environments across four cycles ( $0.48 \pm 0.07$ ), was higher than the mean genomic prediction accuracy across the target international sites ( $0.09 \pm 0.18$ ), owing to the low heritabilities ( $0.34 \pm 0.26$ ) and small training population sizes in the target sites. When the same lines evaluated in another year were predicted using GY in one year, we observed a mean accuracy of  $0.28 \pm 0.18$  using the GxE model. We also evaluated the feasibility of using a representative target site for predicting GY of the

same lines in correlated sites and observed that there were significant differences in the predictability of correlated sites across years with a mean accuracy of  $0.19 \pm 0.22$ . Overall, we observed only a negligible advantage of using genomics over the baseline models, mainly because of the small family sizes in the yield testing stages and the large year variance compared to the genetic and GxE variance. We also evaluated the integration of green normalized difference vegetation index (GNDVI) from high-throughput phenotyping in predicting grain yield from the drought and heat stressed simulated environments of Obregon and observed that the maximum increase by integrating GNDVI was 25% in genomic prediction models and 40% in pedigree-based prediction models. In across-year GY predictions, genomic predictions resulted in low accuracies compared to using GNDVI in both the training and validation populations, thereby highlighting the importance of having some information on a line's performance in a particular environment (a correlated trait like GNDVI in this case) for successfully predicting GY and sparse testing should be considered cautiously. Overall, we conclude that GY is a challenging trait for GS because of its environmental sensitivity and multi-environment trials are still indispensable for minimizing risk and developing climate-resilient wheat varieties.

## **W246: Climate Change and ICRCGC 1**

### **Strategies for Improving Plant Drought Tolerance for a Hotter, Drier World**

Sung Don Lim<sup>1</sup>, Won Yim<sup>2</sup> and **John C. Cushman**<sup>2</sup>, (1)Kangwon National University, Gangwon-do, South Korea, (2)University of Nevada - Reno, Reno, NV

The climate crisis driven by increasing greenhouse gas emissions threatens to increase the frequency, duration, and geographical distribution of droughts over global landmasses during the 21<sup>st</sup> century. As crops become more sensitive to heat and drought stress, the rates of grain yield increases are projected to decline with losses worsening in the latter part of this century. In order to develop novel strategies for improving, we have investigated a specialized form of photosynthesis known as crassulacean acid metabolism (CAM) that is present in more than 6% of vascular plant species along with a suite of co-adapted traits (e.g., tissue succulence, water capture and storage strategies, thick cuticles, enhanced epicuticular wax accumulation, reduced stomatal density, increased stomatal responsiveness, and rectifier-like roots) that might serve to improve the adaptability of plants to hotter and drier climates. CAM increases water-use efficiency (WUE) and reduces water demand through the use inverted stomatal behavior coupled with a temporal CO<sub>2</sub> pump with nocturnal CO<sub>2</sub> uptake and concentration. Thus, introducing the CAM pathway into C<sub>3</sub> photosynthesis plants (CAM Biodesign) is expected to confer enhanced photosynthetic performance and WUE. Current steps achieved to date for CAM Biodesign will be summarized including subcellular localization and phenotypic analysis of overexpressing 14 individual ice plant C<sub>4</sub>-cycle genes, mesophyll-specific, circadian clock-controlled promoter mining, vector set construction for multi-gene circuit assembly, and the phenotypic effects of engineering a four-component carboxylation module in *Arabidopsis*. In addition to engineered CAM, we have explored the effects of increasing tissue succulence on plant growth, productivity, drought acclimation, and salinity stress tolerance in *Arabidopsis*. Increasing cell size resulted in a 2–3-fold increase in leaf succulence (defined as the water content of the leaf/leaf area) with a corresponding decrease in stomatal density and aperture, which resulted in a 1.5–2.8-fold increase in instantaneous WUE and a 2.1–2.3-fold increase in integrated WUE compared to control lines. This improved WUE resulted in significant improvements in aerial biomass and seed yield under both acute and chronic water-deficit stress. Enhanced tissue succulence also resulted in significant increases in aerial biomass and seed yield under both acute and chronic salinity stresses due to a reduction in the effective Na<sup>+</sup> and Cl<sup>-</sup> concentrations within leaves and reduced Na<sup>+</sup> uptake. These results indicate that new approaches to improving drought attenuation and salinity tolerance are possible through relatively small changes in leaf anatomy and architecture.

## **W247: Climate Change and ICRCGC 2**

### **Translational Genomics for Developing Climate Resilient and High Yielding Chickpeas**

**Rajeev K Varshney**, ICRISAT, Hyderabad, India

Chickpea (*Cicer arietinum* L.) is an important pulse crop grown in more than 50 countries across the globe especially in South Asia and Sub-Saharan Africa. During the last decade, genomic revolution

empowered the chickpea community with large scale genomic resources for understanding the genetics of the trait and trait improvement using modern breeding approaches. Deploying the available genomic resources, we dissected important abiotic and biotic stresses that hinder chickpea production. The “*QTL-hotspot*” on CaLG04 explaining more than 58% phenotypic variation was introgressed into different elite backgrounds in India and Africa. Two high yielding and drought tolerant varieties, Geletu (in the background of ICCV 10) and BGM 10216 (in the background of Pusa 362) were released for commercial cultivation in Ethiopia and India respectively. In addition, MABC-WR-SA-1 a high yielding and Fusarium wilt resistant variety developed using marker assisted backcrossing approach was also release for comer cultivation in India. Following deciphering of the draft genome sequence of CDC Frontier variety, we re-sequenced >3000 chickpea germplasm accessions (that include global composite collection and 195 wild species accessions from primary, secondary and tertiary gene pools) at an average ~12X coverage. Large scale resequencing provided greater insights to genome-wide variations, the haplotype diversity, mutation burden, deleterious alleles, bottlenecks and selections sweeps during domestication. Extensive multi-location phenotyping data and the genome wide SNPs enabled us to identify genome-wide associations for agronomically important and > 40 nutritional quality traits. Furthermore, we are using these datasets for genomic prediction for developing climate resilient chickpea varieties.

## **W248: Climate Change and ICRCGC 2**

### **Selecting Optimal Allelic Portfolios in Domesticated Genepools for Seasonal and Climatic Adjustment**

**Graham J. King**, Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia and Jing Wang, Huazhong Agricultural University, Wuhan, China

Following domestication of crop species, processes of primary and secondary eco-geographical radiation typically involve adaptation of landraces to local environments and end-uses. These may result in inadvertent diversity bottlenecks, and at some loci canalisation of allelic variation with associated linkage drag. As both major and niche crops become globalised, specific challenges arise in managing the portfolio of genes and alleles that contribute to adaptation where latitudinal day-length and seasonal variation in temperature represent key environmental parameters affecting flowering time and harvest. It is therefore important to understand the relative contribution and combinations by which key regulators involved in vernalisation, flowering and maturity time may have been subject to selection and contributed to different histories of secondary domesticated radiation. This talk will present a conceptual framework supported by detailed experimental evidence drawn from a range of sources, particularly including the various crops of *Brassica* species. The aim will be to highlight how understanding the full history of crop domestication, including multiple rounds of inter-genic hybridisation and introgression, can help plan and manage assembly of suitable allelic portfolios that may allow deployment in changing landscapes and climate scenarios.

## **W249: Climate Change and ICRCGC 2**

### **Climate Change in the Midwest: What Soybeans and Corn are Telling Us about Food Security**

**David Lightfoot**, Southern Illinois University, Carbondale, IL

The US Midwest helps feed the World. 25% of the soybean and corn and wheat crops are exported. Over the past 28 years the same 100 cultivars were planted in mid may and harvested later and later. Once early October, now late November. Maturity groups headed North, along with Southern diseases. Mid season rains are scarcer. Therefore, new crop plants suited to growth in semi-arid environments will be fundamental to the future of agriculture. The interactions between nitrogen supply and water availability that determine yield and quality in crops grown in semi-arid environments have to be elucidated. Tools for analyzing the metabolic changes associated with enhanced nitrogen assimilation, nutritional value, disease resistance and yield under drought have been generated, but more are needed. Nutritional value must be enhance, weight at the farm gate is not enough. Cancer incidences due to fungal toxin contamination must be reduced by 50%. See US patents 5,998,700; 6,329,573; and pending. The future challenges fall to a new generation guided by the few. It is on us.

## **W250: Climate Change and ICRCGC 2**

### **Food-Legume Genetic Resources for Sustainable Agriculture, Food Security and Human Health**

**Roberto Papa**, Marche Polytechnic University (UNIVPM), Ancona, Italy and INCREASE Consortium

INCREASE is a European project (H2020-SFS-2019) focusing on the management and use of Genetic Resources on food legumes, which are crucial for sustainability, food security and human health. To meet this challenge INCREASE will expand the utilisation of food legumes genetic resources targeting users' needs in terms of accessibility, quality and quantity of information available. INCREASE, working with four important food legumes (chickpea, common bean, lentil, lupin) with significant value for the innovation of EU agriculture and food industry, will be based on four pillars:

- i) innovative data management solutions to develop gold standards for data sharing and integration into the central infrastructure, with decentralised data input, defined methodologies and best practices for exploitation of the novel information produced as well as the development of user friendly visualization tools;
- ii) developing novel tools and principles for germplasm management, based on the development of "Intelligent Collections" as a set of nested core collections of different sizes representing the entire diversity of each crop;
- iii) adoption of cutting-edge technologies for genotyping and phenotyping combined with the potential of Artificial Intelligence focusing on traits of interest for users;
- iv) international effort with the involvement of non-European partners and international organization to expand the scope and ambition of INCREASE. We will develop a citizen-science experiment, primarily aimed at dissemination of the project to stakeholders and citizens.

Overall, INCREASE will strengthen the field of legumes genetic resources and simultaneously it will represent an important model and tool for all crop genetic resources.

## **W251: Climate Change and ICRCGC 2**

### **Adapting Lentil to Changing Environments**

**Kirstin Bett**<sup>1</sup>, Sandesh Neupane<sup>2</sup>, Derek Wright<sup>2</sup>, Taryn Heidecker<sup>1</sup>, Teketel A. Haile<sup>1</sup>, Crystal Chan<sup>3</sup>, Tania Gioia<sup>4</sup>, Diego Rubiales<sup>5</sup>, Eleonora Barilli<sup>6</sup>, Sripada Udupa<sup>7</sup>, Reena Mehra<sup>8</sup>, Ashutosh Sarker<sup>8</sup>, Mohammad Hossain<sup>9</sup>, Babul Anwar<sup>9</sup>, Rajeev Dhakal<sup>10</sup>, Clarice J Coyne<sup>11</sup>, Larissa Ramsay<sup>1</sup> and Bert Vandenberg<sup>1</sup>, (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)University of Saskatchewan, Canada, (3)Dept of Plant Sciences/University of Saskatchewan, Saskatoon, SK, Canada, (4)University of Basilicata, Italy, (5)Spanish National Research Council, Cordoba, Spain, (6)CSIC, Spain, (7)ICARDA, Rabat, Morocco, (8)ICARDA, India, (9)BARI, Bangladesh, (10)LiBird, Nepal, (11)USDA ARS, Pullman, WA

Lentil has three primary macro-environmental adaptation groups determined by photoperiod and temperature: spring-sown northern temperate, fall-sown mediterranean and post-rainy season, subtropical savannah. Genotypes from one region often fail to perform well in the others due to inappropriate response to the local day length and temperature regime during the growing season. Breeders are reluctant to use unadapted germplasm from other regions as parents in crosses because this strategy requires additional time and resources to reduce the influence of the alleles related to poor adaptation.

To better understand the genetics underlying adaptation in a field environment, we grew a panel of 324 diverse lines in 9 locations around the globe for a total of 18 site-years. Genotypes flowered quickly in Saskatchewan, and very slowly in the Mediterranean region. In South Asia, genotypes from Canada, and other regions with long days during the growing season, did not flower or did not set seed before experiencing extreme temperatures at the end of the local growing season. Interactions with both temperature and photoperiod were evident but not all genotypes responded similarly. Some were early to flower everywhere, some were late everywhere and there were those that responded differently

depending on the location. Modelling was used to predict days to flower under different photoperiod and temperature regimes allowing us to predict what could happen if lentil is moved into new environments.

We genotyped all lines, and using a GWAS approach we were able to identify several regions of the genome that were associated with flowering time at each of the locations. Some were unique to short-day environments and others to long-day. Markers for MAS were generated and are being deployed in lentil breeding programs.

## **W252: Climate Change and ICRCGC 2**

### **Genomics-Assisted Utilisation of Landraces and Wild Species for Development of Flooding and Salt-Tolerant High-Yielding Rice Cultivars**

**Nagendra K. Singh**<sup>1</sup>, Sarita Kumari<sup>1</sup>, Dhriti Satya<sup>1</sup>, Manoj M.L.<sup>2</sup>, Kabita Tripathy<sup>1</sup>, Balwant Singh<sup>1</sup>, Sangeeta Singh<sup>1</sup>, Dipika Singh<sup>2</sup>, Shefali Mishra<sup>1</sup>, S.L. Krishnamurthy<sup>3</sup> and Vandna Rai<sup>1</sup>, (1)ICAR-National Institute for Plant Biotechnology, New Delhi, India, (2)ICAR-National Institute for Plant Biotechnology, India, (3)Central Soil Salinity Research Institute, India

One of the greatest challenges for the 21<sup>st</sup> century is to produce enough food for growing human population from diminishing crop acreage, deteriorating soil health and stresses induced by global climate change. Crop wild relatives are naturally adapted to extreme soil and climatic conditions and present a rich source of genes that can be harnessed to develop climate-resilient cultivars. Therefore, exploration, evaluation and utilisation of fast depleting crop wild relatives gene pool is the need of the day. We have evaluated a large pool of wild rice (*Oryza rufipogon* Griff./ *Oryza nivara* Sharma et Shastry) germplasm collected from different agro-climatic zones of India and identified accessions that can withstand drought, flood and soil salinity stresses better than what is available in the cultivated rice germplasm. Introgression of novel QTLs/genes for drought, flood and salinity tolerance identified in these wild rice accessions is in progress using modern genomic tools. After screening of 292 wild rice accessions for salinity tolerance, two were found highly salt tolerant, 11 tolerant, 29 moderately tolerant, 70 sensitive and 180 highly salt sensitive. Screening of 202 wild rice accessions and check varieties in a rain-out shelter for drought stress tolerance at vegetative stage using four parameters viz. canopy temperature, chlorophyll content, leaf rolling and relative water content identified 35 accessions highly tolerant, 52 moderately tolerant and 115 highly susceptible to drought stress. Screening of 283 wild rice accessions for anaerobic germination identified 13 and 26 genotypes highly and moderately tolerant to anaerobic germination, respectively. Screening for submergence tolerance identified 11 accessions highly tolerant, 40 accessions moderately tolerant and 232 accessions sensitive. Crosses have been made between accessions highly tolerant to anaerobic germination, submergence, drought and salinity stresses. BC1 derived progenies from five crosses for anaerobic germination, and salinity tolerance were genotyped using a 50k SNP chip array and phenotyped to identify major QTLs for anaerobic germination and salt tolerance in a bid to simultaneously map and transfer the useful genomic regions in popular high yielding rice cultivars. In addition to genes for climate resilience efforts are also underway to combine genes for biotic stress resilience and further enhancing the yield potential of mega rice varieties without losing the quality attributes crucial for consumer acceptance.

## **W253: Climate Change and ICRCGC 3**

### **Tapping Wheat Diversity to Cope with Climate Change**

**Etienne Paux**, INRA GDEC, Clermont-Ferrand, France

Global agriculture is currently facing an unprecedented challenge: to meet the changing food demand of a growing world population under sustainable environmental and social conditions. To feed a population that is expected to reach nine billion people by 2050, an annual increase in yields of around 1.7% is necessary. Achieving this increase would be possible in a stable environment but seems more questionable in an environment subject to climate change that affects not only yields but also their stability. Developing varieties that are more tolerant or

resilient to abiotic stresses is therefore one of the main priorities of breeding programs. Since its formation in the Fertile Crescent during the Neolithic, bread wheat has undergone a complex history of spread, adaptation and selection. Today, bread wheat is grown in a wide range of environments, thanks to its large genetic diversity. Exploiting this diversity is a way to breed for varieties that can cope with climate change. However, to fully exploit genetic resources, one needs to better characterize and understand the wheat phylogeography and genetic diversity. To this aim, we genotyped a set of 4,500 accessions originating from 105 countries and comprising both landraces, traditional cultivars and elite varieties. Based on 113,457 SNPs, we identified 8,741 haplotypes and used them to characterize our panel. Our results shed light on the complex history of bread wheat and show how man has influenced the worldwide genetic diversity of this species, leaving a significant part largely underexploited.

### **W254: Climate Change and ICRCGC 3**

#### **Studies on Genetic Diversity and Key Traits in Underutilised Legumes to Address Climate Change in West Africa**

**Michael Abberton**, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

There is considerable focus on the need to develop climate resilient and nutritious food systems in West Africa particularly in the northern savannahs and Sahel. At GRC international collections of two underutilised legumes, Bambara groundnut (*Vigna subterranea*) and African yam bean (*Sphenostylis stenocarpa*) are being studied to understand the genetic relationships between accessions and to relate this diversity to phenotypic information for key traits. To this end we have deployed DArT markers and evaluation of drought tolerance, yield, N fixation, cooking time and other important traits. This work will provide the foundation for breeding programmes in these crops for West Africa

### **W255: Climate Change and ICRCGC 3**

#### ***Saccharum* × *Miscanthus* Intergeneric Hybrids (Miscanes) Show Low Temperature Stress Tolerance**

Suraj Kar, Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan, Na Zhang, Beijing Key Laboratory of Growth and Developmental Regulation for Protected Vegetable Crops, College of Horticulture, China Agricultural University, Beijing, China, Tzu-Ya Weng, Graduate School of Global Food Resources, Hokkaido University, Sapporo, Japan, Taiken Nakashima, Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan, Antonio Villanueva-Morales, Department of Statistics, Mathematics and Computing, Forest Sciences Division, Universidad Autónoma Chapingo, Texcoco, Mexico, J. Ryan Stewart, Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, Erik J. Sacks, Dept of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, Yoshifumi Terajima, Japan International Research Center for Agricultural Sciences, Ishigaki, Japan and **Toshihiko Yamada**, Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Japan; Graduate School of Global Food resources, Hokkaido University, Sapporo, Japan

Sugarcane is cultivated on 26.7 mha in tropical and sub-tropical regions and yields nearly 1.9 billion metric tons per year with a peak dry matter yield >100 tons (ha yr<sup>-1</sup>). Sugarcane is used for sugar as human consumption as well as a feedstock for biorefinery. However, the lack of environmental adaptation of sugarcane has been a persistent problem, especially owing to its susceptibility to cold. *Saccharum* can be crossed with related genera belonging to so-called “*Saccharum* complex” including *Miscanthus* which is a native C<sub>4</sub> grass of East-Asia. *Miscanthus* was reported to produce shoots at a temperature as low as 6 °C and survive after prolonged exposure to temperatures <-6.5 °C. True hybrids plants, often termed as miscane, between commercial sugarcane (*Saccharum* spp. *hybrid*) genotypes and *Miscanthus* genotypes as *M. sinensis* and *M. sacchariflorus*, have been recently obtained. The aim of the present study is to test chilling tolerance in miscane plants compared to its parents, which will confirm the introgression of chilling tolerance traits into sugarcane from *Miscanthus*.

In a greenhouse experiment on long-duration chilling stress (12–13 °C day / 7–9 °C night), seven miscane genotypes exhibited higher photosynthetic rates than their sugarcane parents after seven days of chilling, whereas after 14 days only four miscanes had significantly higher photosynthetic rates than their sugarcane parents, but notably two of these did not differ from their highly tolerant *Miscanthus* parents. The results indicated variability in chilling tolerance in miscanes, thus selection will be a key aspect of improving chilling tolerance in sugarcane. In addition, field experiment trials evaluated that several miscane genotypes have high early- and late-season photosynthesis coupled with high biomass production, likely indicating chilling tolerance. One miscane genotype was identified as a superior genotype for introgression breeding programs of sugarcane.

### **W256: Climate Change and ICRCGC 3**

#### **Development of Water Use Efficient Rice Varieties Suitable for Alternative Irrigation Systems**

**Niranjana Baisakh**<sup>1</sup>, Rodante E. Tabien<sup>2</sup>, Manoch Kongchum<sup>3</sup>, Dustin Harrell<sup>3</sup> and Adam Famoso<sup>3</sup>, (1)School of Plant, Environmental, and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA, (2)Texas A&M AgriLife Research, Beaumont, TX, (3)H. Rouse Caffey Rice Research Station, Louisiana State University Agricultural Center, Rayne, LA

Irregular rainfall pattern leading to drought is one of the many environmental consequences of climate change. This results in shortage of fresh surface water and groundwater resources worldwide, which threatens the sustainability of irrigated rice production. In the U.S., the Mississippi River Valley alluvial aquifer has shown declines in Arkansas and Mississippi. Limited irrigation water following droughts in Texas in recent years has severely affected their rice industry. Louisiana may experience short or prolonged droughts during the growing season. Fresh water availability can be affected by storm surge or drought conditions in the south Louisiana canal system that can affect the quality of surface irrigation water by making it temporarily saline or alkaline. Alternative water management strategies, such as semi(aerobic), alternate wetting and drying, furrow-irrigation etc, are being adopted as promising strategies to combat future water shortages. Furrow-irrigated rice production, also known as row rice, is becoming more popular in some southern states of U.S. This allows farmers to grow rice on fields that are traditionally not used for rice production because of the topographical difference of lands. For varieties to fit to a (semi)aerobic or row rice production system, they need to have improved water-use efficiency with the ability to tolerate high blast pressure. Development of successful aerobic rice cultivars could be achieved by combining the high-yielding traits of irrigated rice with the drought-tolerance traits of traditional upland rice cultivars. To this end, crosses involving a U.S. bred rice variety and two Indian rice cultivars were made. Following preliminary greenhouse-based screening of early-generation progenies, evaluation of advanced-generation lines has identified potential lines that can withstand periodic dry spells under field conditions at Crowley, Louisiana and under semi-aerobic conditions at Beaumont, Texas with minimal reduction in grain yield compared to non-stressed conditions. Research is being undertaken to quantify their water-use efficiency and evaluate yield and other production traits for their suitability to alternative irrigation management schemes such as row rice.

### **W257: Climate Change and ICRCGC 3**

#### **Climate-Ready Rice for South America: The Search for Stress Response Genes**

**Antonio Costa De Oliveira**, Universidade Federal de Pelotas, Pelotas-RS, Brazil

Climate changes and increasing biotic and abiotic stresses are affecting crops worldwide. Rice, one of the top three cereals, is under constant constraint by these threats. Brazil, a very important rice producer and the largest producer outside Asia, has an annual production of over 10 million tons. Many stresses affect yields and lower total production each year. Stresses such as cold, flooding, drought, iron toxicity and salinity are constantly affecting farmer's yields. Our lab has been working in developing stress resilient lines to these different stresses. The understanding of plant response mechanisms is key to the development of stress resilient crops. WRKY transcription factors (TFs) are responsible for the regulation of genes responsive to many plant growth and developmental cues, and are involved in biotic and abiotic stress responses. Recently, functional genomics studies in model plants have enabled the identification

of function and mechanism of action of several WRKY TFs in plants. Our group has been using molecular tools and mutation breeding to accelerate breeding for abiotic stress tolerance.

### **W258: Climate Change and ICRCGC 3**

#### **Detecting and Classifying Abiotic and Biotic Stress Response using Phenomic Tools**

**Cory D. Hirsch**, University of Minnesota, St. Paul, MN

The ability of plants to respond to and overcome different types of stress with minimal effects on yield and quality is critical to ensure food/product production and farming economics. Most crop losses due to stresses are mitigated most effectively, economically, and sustainably by understanding the underlying genetics and by breeding new cultivars. Development of rapid sequencing and genotyping technologies has increased the speed and lowered the cost of understanding the genetics of plant traits. To take advantage of fast and low-cost sequencing technologies researchers are developing and using low cost and high-throughput plant phenotyping methods to characterize varietal performance. Classifying maize seedlings tolerance to cold stress could mitigate yield reductions in increasing climate variability. Towards this goal we subjected a panel of maize inbred lines to cold stress in controlled environments. By utilizing image analysis multiple traits were extracted and used to classify the stress response of each genotype. Beyond RGB imaging we will also present the use of hyperspectral imaging to predict biotic stress in soybean. This work is the basis for ongoing research investigating the genetic basis and spectral response of plants in different stress environments.

### **W259: Cloud Computing for Genomics Data Analysis**

# **Cloud Computing Bridges Big Biological Data and AI to Accelerate Research and Application in Plant Genomics**

**Xinshuai Qi**, Benson Hill, St Louis, MO

In the post sequencing era, genomics data analysis and knowledge generation become more challenge than data generation itself. Cloud computing provides a “pay as you go” option for large scale biological data and has great potential to accelerate the genomics research through platform and collaboration. Here, we will provide a comparison of cloud and local computing, successful cloud computing applications in genomics research, and vision for the future direction. Invited speakers from both academic and industry will provide insights on the application and direction of cloud computing in computational biology.

### **W260: Cloud Computing for Genomics Data Analysis**

#### **National Center for Genome Analysis Support (NCGAS): Genomics and other Science in the NSF-Funded Jetstream Cloud**

**Thomas G. Doak**<sup>1</sup>, Sheri Sanders<sup>2</sup>, Carrie Ganote<sup>3</sup>, Bhavya Papudeshi<sup>3</sup>, Jeremy Fischer<sup>4</sup> and David Y Hancock<sup>4</sup>, (1)National Center for Genome Analysis Support; manager, Bloomington, IN, (2)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (3)Indiana University, Bloomington, IN, (4)Indiana University

The National Center for Genome Analysis Support (NCGAS) is an NSF-funded (NSF-1445604) center that helps all NSF-funded researchers doing genomics research. Genomics includes transcriptomics, metagenomics, genome annotation, etc. Our support includes providing access to large memory

computing, maintaining curated sets of genomics applications, providing one-on-one consultation, and creating educational opportunities. A resource that we have come to rely on for providing these services is the NSF-funded Jetstream Cloud—maintained by Indiana University (led by the Indiana University Pervasive Technology Institute (PTI) and the University of Texas at Austin's Texas Advanced Computing Center (TACC). Additionally, we leverage Globus data transfer tools. Globus at the University of Chicago is responsible for integrating Jetstream with the NSF-funded Extreme Science and Engineering Discovery Environment (XSEDE), and for integrating Globus data movement and management tools, as well as Globus-based secure user authentication.

With a focus on ease of use and broad accessibility, Jetstream is designed for those who have not previously used high performance computing and software resources—for researchers who need more than desktop-strength computing but less than full-scale High Performance Computing (HPC). Jetstream features a web-based user interface based on the popular Atmosphere cloud computing environment—developed by [CyVerse](#)—extended to support science and engineering research generally. The system is particularly geared toward 21st-century workforce development at small colleges and universities – especially historically black colleges and universities, minority serving institutions, tribal colleges, and higher education institutions in EPSCoR States.

Jetstream provides a library of virtual machines designed to do discipline-specific scientific analysis, but researchers can also develop their own VMs, with their own software sets, or sets specialized to a particular task. These VMs can be both saved and shared with collaborators. Currently there are 19 genomics VMs, including RStudio instances with bioconductor, ready-made genome browsers with JBrowse/Tripal, and metagenomic tools like QIIME2 and Anvi'o. biology and molecular biology researchers are the largest users of Jetstream. NCGAS has found VMs extremely useful in education and workshops: we develop class-specific VMs, with all the applications needed, then clone, so that each student has their own VM to work on (making courses easy to scale).

In addition to on-demand VMs, persistent science gateways can be established using template VMs NCGAS has built. These can be used to provide services to collaborators or to the world. Users can easily create Galaxy servers on Jetstream: each server comes preconfigured with hundreds of tools and commonly used reference datasets—once running, researchers can use it or customize it. Many NCGAS users establish genome browsers—specific to their organism—that are shared with small sets of collaborating researchers—but can be shared to the world.

Jetstream is accessed via an allocation process at XSEDE—a startup allocation is typically approved within a day.

## **W261: Cloud Computing for Genomics Data Analysis**

### **Reproducibility in the Clouds - Harnessing the *in nube* Paradigm**

**Jason Williams**, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

While bioinformatics (*in silico*) research encompasses how we do genomics, where we do bioinformatics is an important consideration as researchers consider how or if to adopt the cloud-based research paradigm (*in nube*). Over the past several years open science clouds (*CyVerse Atmosphere*, *JetStream*) have provided mid-scale infrastructure solutions that have accelerated the pace of research. This talk will introduce these research capabilities as well as “getting started” tips to making use of the publicly funded resources. In addition, we will provide recommendations for developing reproducible solutions (e.g. containers and workflow systems) that can help researchers take advantage of commercial cloud solutions (which are increasingly evolving to cater to data and analysis solutions for life sciences).

## **W262: Cloud Computing for Genomics Data Analysis**

### **Scalable Genome Assembly on the Cloud: From One to Seventy Thousand Species**

**Arkarachai Fungtammasan**, DNAnexus, Mountain view, CA

The assembled genome is an essential resource for plant and animal genetics studies. The high-quality reference genome serves as a strong foundation for population genetics and comparative genomics. The variety-specific or individually assembled genome allows us to detect large structural variants that could not be discovered by resequencing. However, generating high quality assembled genome is complex and require a massive amount of computing power which is not trivial to train scientist or to maintain such a computing facility.

In this presentation, we will demonstrate the two main benefits: on-demand scalability and collaborative aspects which make cloud computing is a favorable solution over local clusters. We will discuss the different challenges from a single genome to a population and to a consortium level of effort.

### **W263: Cloud Computing for Genomics Data Analysis**

#### **Industrializing Genotype Data on Public Cloud Infrastructure**

**Rob Long**, Bayer Crop Science, Creve Coeur, MO

The uses and collection methods of genotype data needed by a global seed company are many and varied. The organic evolution of systems to track these genotype datasets results in a patchwork of systems spread across various functions. The decentralized solution to genotype data results in heterogeneous application of standards, resulting in a multiplication of toil when consumers put together datasets that span the various technologies. We solved this problem by creating a single service, a cloud-based repository for genotypes, integrating all data sources, and making an API available to access the data. A service providing access to all genotype data allows sustainable development of higher-order operations on top of the foundational data layer. Such higher-order operations can read and write into the existing data layer, without requiring point-to-point integrations for each new source or sink of genotype data.

This talk will cover our approach to organizing multi-organism genotype data, making sure that streaming updates are received from the various labs and sources of data, and providing high-availability APIs that allow data scientists and application developers programmatic access. We describe an example of a higher-order process built upon our genotype API: Pedigree-based Genotype Inference. The architecture of the pedigree-based genotype inference engine is highly-distributed, allowing it to take full advantage of cloud-based commodity servers, with production surges to over 100,000 CPU cores in simultaneous use. The distributed nature of the engine allowed us to take advantage of low-cost “preemptable” compute instances. These preemptable instances require more up-front engineering time, but allow a 5x reduction in resource costs.

### **W264: Comparative Genomics**

#### **Conservation and Variation in Stress-Responsive Gene Regulatory Circuitry across Angiosperms**

**Julia Bailey-Serres**, Center for Plant Cell Biology, UC RIverside, Riverside, CA

Greater understanding the evolutionary conservation of genes and their regulation by abiotic stress in diverse species can aid the development of climate resilient crops. Flooding is a significant challenge to crop production and a factor in plant distribution in ecosystems. To broadly define gene regulation mechanisms relevant to flooding, we carried out a high-resolution analysis of dynamics in chromatin accessibility, transcription and mRNA translation in roots of four angiosperms, ranging from a dry adapted wild tomato to the wetland crop rice, in response to submergence (1). This was facilitated by the application of technologies that refine the evaluation of gene activity including INTACT (Isolation of Nuclei Tagged in specific Cell Types), TRAP (Translating Ribosome Affinity Purification) and ATAC (Assay for Transposon-Accessible Chromatin). Integration of the data identified 69 submergence upregulated gene families (SURFs) broadly conserved across angiosperms. By integrating phylogenomics and *cis*-regulatory motif occurrence in accessible chromatin regions of these gene families, we recognized evolutionary ancient regulatory networks mediated by four transcription factor families. The conservation was recognized in both syntenic and non-syntenic genes, with the highest

frequency of a demonstrated hypoxia-responsive *cis*-regulatory element in regions of accessible chromatin in SURFs of the wetland species. These data highlight variation in hypoxia-response networks is influenced by plant natural history. Further evaluation of response to water deficit and waterlogging in defined cells of roots and the shoot apex of rice identify dynamics in gene modules associated with plastic traits associated with flooding and drought resilience.

(1) Reynoso MA, Kajala K, Bajic M, West D, Pauluzzi G, Yao A, Hatch K, Zumstein K, Woodhouse M, Rodriguez-Medina J, Sinha N, Brady SM, Deal RB and Bailey-Serres J (2019) Evolutionary flexibility in flooding response circuitry in angiosperms. *Science*, 365(6459), 1291–1295. (2) Supported by NSF-PGRP (IOS-1238243; IOS-1546879, IOS-1856749).

## **W265: Comparative Genomics**

### **Temporal Expansion and Diversification of the Polyploid *B. rapa* Transcriptome**

**Kathleen Greenham**<sup>1</sup>, Ryan Sartor<sup>2</sup>, Ping Lou<sup>3</sup>, Todd C. Mockler<sup>4</sup> and C. Robertson McClung<sup>3</sup>, (1)University of Minnesota, Saint Paul, MN, (2)North Carolina State University, Raleigh, NC, (3)Dartmouth College, Hanover, NH, (4)Donald Danforth Plant Science Center, Saint Louis, MO

The genus *Brassica* supplies crops with leaf, flower and root vegetables for consumption, oil production, and fodder. *B. rapa* captures much of this diversity in one species with Chinese cabbage, pak choi, oilseed, turnip and leafy vegetable varieties. As with many crops, *B. rapa* is polyploid, diverging from *Arabidopsis thaliana* roughly 24 million years ago (MYA) and undergoing genome triplication followed by extensive gene fractionation. In *Arabidopsis*, up to 90% of the transcriptome has been found to be rhythmically expressed under diel or circadian conditions and around 30% has been shown to be circadian regulated. We wondered whether the temporal control of the transcriptome was maintained in *B. rapa* and whether the expansion of the genome resulted in an associated expansion of temporal regulation. We performed a series of time course RNA-seq experiments to profile the diel and circadian control of the transcriptome. We sought to classify gene expression patterns as well as compare patterns of retained paralogs to look for diversification in regulation. Applying our newly developed R package DiPALM (Differential Pattern Analysis by Linear Models), we uncovered expansion of the circadian transcriptome in *B. rapa* supporting an important role for the circadian network. Additionally, we find evidence for genome wide expansion of phase domains among retained paralogs with support for divergence of abiotic stress response among pairs. Divergence between paralogous transcription factors is supported by variation in gene expression patterns and presence of conserved noncoding sequences of predicted target genes.

## **W266: Comparative Genomics**

### **Using Orthologous Mutations to Understand the Role of RNA-Directed DNA Methylation during Seed Development**

**Rebecca A. Mosher**, University of Arizona, Tucson, AZ

Twenty-four nucleotide short interfering (si)RNAs direct DNA methylation to thousands of genomic loci in a process called RNA-directed DNA methylation (RdDM). These siRNA are especially abundant in plant reproductive tissues, including the developing seeds. *Arabidopsis thaliana* mutants that lack RdDM are fully fertile and show no defects in transmission of mutant alleles, however cognate mutations in *Brassica rapa* have a severe seed production defect. Genetic characterization demonstrates that loss of maternal sporophytic RdDM causes abortion of fertilized tissues, suggesting there might be siRNA communication between the maternal soma and filial tissues. We have identified a small number of siRNA loci account for over 90% of siRNA expression during *B. rapa* seed development. These loci are expressed predominantly from maternal sporophytic tissue, and also show a marked maternal bias in endosperm, suggesting that they might be transported to, and function in, filial tissues. Additionally, we have created RdDM mutants in other members of the Brassicaceae. Initial analysis indicates that inbreeding species *Capsella rubella* is similar to *Arabidopsis thaliana* and does not require RdDM for seed development. However, an RdDM mutant of the closely-related outbreeding species *Capsella grandiflora* has reduced

seed set and increased seed abortion. These data indicate that breeding strategy might be the key to understanding the role of RdDM during seed development.

### **W267: Comparative Genomics**

#### **Comparative Chloroplast, Mitochondrial and Nuclear Genomics of the Genus *Oryza***

**Li-zhi Gao**, South China Agricultural University, Guangzhou, China

### **W268: Comparative Genomics**

#### **Structural and Allelic Variation in Plant NLR Immune Receptors**

Daniil Prigozhin<sup>1</sup>, Janina Tamborski<sup>2</sup>, Erin Baggs<sup>3</sup>, Paul Bailey<sup>4</sup> and **Ksenia Krasileva**<sup>3</sup>, (1)Lawrence Berkeley National Laboratory, Berkeley, CA, (2)University of California Berkeley, Berkeley, CA, (3)Department of Plant and Microbial Biology UC Berkeley, Berkeley, CA, (4)Kew Gardens, United Kingdom

Plant immunity relies on an arsenal of receptors that can directly detect pathogen molecules or their activities in host cells. Keeping up with rapidly evolving pathogens requires the generation of diversity in plant immune receptors. NLRs are intracellular immune receptors with diverse roles and diversification mechanisms, including structural and allelic variation. We have previously demonstrated that plants deploy gene fusions to integrate exogenous domains into NLRs. The integrated domains serve as baits for the pathogen and their modification triggers immune signalling. Grasses deploy a specialist intracellular NLR immune receptor clade that undergoes continuous gene fusions incorporating exogenous protein domains. The mechanism driving this structural variation remains elusive. Our current work addresses the sources of new specificities of receptors that directly recognise pathogen-derived effectors. We show that a quarter of NLR immune receptors in the model plant *A. thaliana* evolve much faster than the rest and that these changes concentrate to the leucine-rich repeat domain. These genes show a strong overlap with hybrid incompatibility loci, suggesting that rapid generation of new immune specificities comes at the cost of autoimmunity. The most variable amino acids cluster on the concave surface of the LRR but are not adjacent in primary amino acid sequence. This allows the prediction of putative effector binding sites. Together, our analyses uncover distinct mechanisms of diversity generation among different groups of plant NLR receptors.

### **W269: Comparative Genomics**

#### **Pack-Mule Transposable Elements in Rice and Its Wild Relatives**

**Ning Jiang**, Department of Horticulture, Michigan State University, East Lansing, MI

Transposable elements comprise a large part of eukaryotic genomes and their amplification causes genome size variation as well as allelic diversity. Moreover, some transposable elements are capable of duplicating and amplifying non-transposon sequences include genes. In rice (*Oryza sativa*), there are 3000 Pack-MULEs, which refer to *Mutator*-like elements carrying gene or gene fragments. The genus *Oryza* consists of 11 distinct genome types but the conservation of Pack-MULE insertions are largely limited within AA genomes, suggesting those elements only remain recognizable for a few million years. The age of Pack-MULEs was estimated through the conservation of insertions among different species as well as the divergence between Pack-MULEs and their parental genes. Interestingly, Pack-MULEs with different age are associated with distinct expression and distribution patterns, suggesting variable selection pressure in different genomic regions.

### **W270: Components of Apomixis**

#### **Introduction to the New Era of Clonal Seeds Production**

**Emidio Albertini**, University of Perugia, Perugia, Italy

In angiosperms two pathways of reproduction through seed exist: sexual, or amphimictic, and asexual, or apomictic. The essential feature of apomixis is that an embryo within an ovule is formed autonomously. The progeny of apomictic plants is generally identical to the mother plant and over the years apomixis has been seen as either a gain- or a loss-of-function over sexuality, implying that the latter is the default condition. However, describing the origin of apomixis in a recently evolved subset of eukaryotes is possibly like speculating on the headwaters of rivers from downstream. Recently some authors considered another scenario, that eukaryotes retain in their genomes a multi-billion-year-old propensity to reproduce asexually.

Moreover, up to now scientists have studied genes controlling key apomixis steps in apomictically reproducing species and have engineered plants with candidate genes that induce apomixis-like in sexually reproducing species. Recently, artificial apomixis systems have been engineering into rice, but attempts to engineer a natural form of apomeiosis into plants failed suggesting that the genetic control of apomixis is more complicated than expected.

Other studies demonstrated that in reproductive cells some DNA methylation deregulation induces apomeiosis-like phenotypes, suggesting that specialization of a DNA methylation pathway acts upon germline or germline associated cells.

This year the different aspects of apomixis control, including the two researches on clonally reproducing rice, will be presented and discussed.

## **W271: Components of Apomixis**

### **Synthetic Apomixis in Hybrid Rice**

**Kejian Wang**, China National Rice Research Institute, Hangzhou, China

Heterosis, or hybrid vigor, is exploited by breeders to produce elite high-yielding crop lines but beneficial phenotypes are lost in subsequent generations owing to genetic segregation. Apomixis, or clonal propagation through seeds, would enable self-propagation of F<sub>1</sub> hybrids and permanently fix heterosis. Here, we established synthetic apomixis in hybrid rice using genome editing. We fixed the heterozygosity of F<sub>1</sub> hybrid rice by multiplex CRISPR-Cas9 genome editing of the *REC8*, *PAIR1* and *OSD1* meiotic genes to produce clonal diploid gametes and tetraploid seeds. Next, we showed that editing the *MATRILINEAL (MTL)* gene could induce formation of haploid seeds in hybrid rice. Finally, we combined fixation of heterozygosity and haploid induction by simultaneous editing of all four genes (*REC8*, *PAIR1*, *OSD1* and *MTL*) in hybrid rice, and obtained plants that could propagate clonally through seeds. Application of synthetic apomixis strategy may enable self-propagation of a broad range of elite F<sub>1</sub> hybrid crops.

## **W272: Components of Apomixis**

### **Synthetic Apomixis for Hybrid Crop Propagation**

**Imtiyaz Khanday**, University of California-Davis, Davis, CA

Crop yields have substantially increased by the use of hybrids that exhibit enhanced vigor over inbred varieties. However, due to genetic segregation, high yielding hybrids cannot be maintained through seed propagation, and need to be generated afresh for each season by labor-intensive methods that result in higher prices for farmers. Consequently, hybrids are underutilized in crops, including staple crops rice, wheat, barley etc. Plant species in many different lineages have evolved a process called apomixis that results in asexual reproduction through seeds. The introduction of apomixis into major crops has been a long-sought goal for agricultural biotechnology. Here, we describe tools for synthetic apomixis, by which a hybrid crop plant could be engineered to self-reproduce through seeds while maintaining parental heterozygosity. The principle is that bypassing meiosis and fertilization, hence sexual reproduction, should result in plants that produce seeds with clonal progeny. Our previous studies of initiation of embryogenesis in rice showed that expression in the egg cell of a male-expressed zygotic transcription

factor, BABY BOOM1 (BBM1), induces parthenogenesis. We then used genome editing of three meiotic genes (*OSD1*, *PAIR1* and *REC8*) to substitute mitosis for meiosis (*MiMe*) in the rice germline. When we introduced *MiMe* into plants expressing BBM1 in the egg cell, up to 30% of the progeny had the same ploidy and genetic constitution as the mother plant. The clonal nature of the progeny was confirmed by whole genome sequencing of the diploid mother plant and progeny representing two successive generations of clones. Specifically, we identified heterozygosity at 57 SNP loci in the genome of mother plant and determined that heterozygosity at all these SNPs was maintained in the clones and grand clones. We have now established that the synthetic apomictic trait is heritable through multiple generations. These results establish the feasibility of clonal propagation and the fixation of hybrid vigor in crop plants.

### **W273: Components of Apomixis**

#### **Meiosis to Apomeiosis Switching Is Metabolically Inducible in Angiosperms**

**John G. Carman**<sup>1</sup>, David A. Sherwood<sup>2</sup>, Lei Gao<sup>3</sup>, Mayelyn Mateo de Arias<sup>4</sup> and Bo J Price<sup>1</sup>, (1)Utah State University, Logan, UT, (2)Sherwood Pet Health, Logan, UT, (3)Jiujiang University, China, (4)Santo Domingo Institute of Technology, Santo Domingo, Dominican Republic

The first cytologically detectable events of gametophytic apomixis in angiosperms are sexual process termination (pre-meiosis to early post-meiosis abortion) and apomeiosis onset (unreduced gametophyte and egg formation). We profiled ca. 20,000 ovules and discovered that those from sexual *Boechera* (Brassicaceae) preferentially transcribe (compared to apomicts) genes involved with transcription, translation, gene silencing, chromatin synthesis, mitosis, meiosis, and DNA repair. In contrast, ovules from apomictic *Boechera* preferentially transcribe genes associated with signaling (hormone, nutrient, stress, calcium, phospholipid), responses (light, defense, unfolded protein), processes (reactive oxygen species attenuation, autophagy, homeostasis, detoxification), and posttranslational modifications. We then tested whether stress (osmotic and heat) shifts apomeiosis to meiosis in facultatively apomictic *Boechera*. Stress increased meiosis frequencies from 5 to 15-fold. We then compared, within taxa, pistil profiles of well-watered plants (highly apomeiotic) to those of droughted plants (highly meiotic). Profiles of the stressed, meiosis-induced pistils mimicked those of sexual *Boechera*. From these results, we identified biochemical pathways and combined them into network diagrams that reasonably explain stress-induced apomeiosis to meiosis switching. We then designed chemical treatments to metabolically modify the pathways with the goal of inducing meiosis in apomictic plants and apomeiosis in sexual plants. Treatments were identified that efficiently achieved these goals for apomictic *Boechera* and for sexual *Boechera*, *Arabidopsis* and *Vigna*. Molecular marker lines of sexual *Arabidopsis* were used to verify meiosis abortion and unreduced gametophyte initiation.

### **W274: Components of Apomixis**

#### **Evolution of Ectopic Plantlet Formation in *Kalanchoe***

**Bharti Parihar** and Ana Almeida, California State University East Bay, Hayward, CA

One of the unique features of some species within the genus *Kalanchoe* is their ability to produce plantlets at the margins of adult leaves, resulting in an impressive potential for asexual reproduction. The genus consists of approximately 140 species, distributed mainly throughout Madagascar, South and East Africa, and Southeast Asia, and is roughly divided into three taxonomic sections based on the manner through which their species reproduce. Section I includes most of the species that propagate exclusively through sexual reproduction. Section II includes species that reproduce either via seed production or vegetative propagation, while Section III comprises species that only reproduce asexually. So far, the leading hypothesis to explain the evolution of leaf-associated plantlet formation in *Kalanchoe* suggests that constitutive-plantlet forming species have coopted both organogenesis and embryogenesis developmental programs to leaf margins, especially involving ectopic expression of *SHOOT MERISTEMLESS (STM)* and *LEAFY COTYLEDON 1 (LEC1)*. However, a detailed understanding of the genetic components of leaf-associated plantlet formation is still lacking. The sequenced genome of *K. fedtschenkoi* suggests the existence of two events of whole-genome duplications after its divergence from grape, early in the history of the rosoid lineage. Here, we investigate the role of genome size in the

evolution of leaf-associated plantlet formation, based on the hypothesis that whole-genome duplication events throughout the evolution of the lineage underlie the molecular basis for asexual reproduction in the genus. If that is the case, we expect to see significantly larger genomes in Section II and III plants when compared to Section I species. We will also carry out a detailed morphological study of plantlet formation and meristem establishment on leaf margins in order to fully characterize plantlet development in the leaf margins of Section II and Section III plants. Here we present the results of these two initial explorations: genome size and plantlet development in *Kalanchoe*. Future studies will include comparative transcriptomics and functional studies in order to elucidate the regulatory gene network involved in leaf-associated plantlet formation in *Kalanchoe* Sections II and III species.

## **W275: Components of Apomixis**

### **Water Stress, Apomixis and Gene Expression in the Diplosporous Grass *Eragrostis curvula***

Juan Pablo Selva<sup>1</sup>, Diego Zappacosta<sup>1,2</sup>, José Carballo<sup>1,2</sup>, Juan Manuel Rodrigo<sup>1,2</sup>, Andres Bellido<sup>1</sup>, Cristian Gallo<sup>1</sup>, Jimena Gallardo<sup>1,2</sup> and **Viviana Echenique<sup>1,2</sup>**, (1)Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) Universidad Nacional del Sur-CONICET, Bahía Blanca, Argentina, (2)Dpto. Agronomía, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina

*Eragrostis curvula* has become a model to understand apomictic mechanisms, specially diplosporous development since our group have generated many plant and genomic resources to study this trait, like a mapping population, a reference genome, the first high density linkage map, sRNAs and transcriptomes databases. The characterization of the molecular pathways underlying apomixis is a prerequisite to transfer the trait to economically important crops. In nature it is possible to observe mainly facultative genotypes of this grass, which retain a certain percentage of sexual pistils that is increased by external factors such as stress conditions, indicating that some regulators activated by stress could be present affecting the apomixis/sex switch. A series of experiments with *E. curvula* plants of the facultative apomictic cv. Don Walter under control and water deprivation conditions were performed in order to associate the increase of sexual embryo sacs in apomictic plants with the differential expressed genes in inflorescences. Three control and three stressed plants were used for the experiments. Inflorescences were collected for cytoembryological analysis and for RNA sequencing (Illumina HiSeq1500 platform). As a result, 501 differentially expressed transcripts between control and stressed plants were found, 305 out of them were annotated and their expression mainly coincided with up or downregulated pathways previously associated with apomixis, like ubiquitin mediated proteolysis, RNA degradation and vesicle trafficking molecular pathways among the most relevant. Differentially expressed genes involved in these pathways were downregulated in stressed plants and its role in apomixis is being considered.

## **W276: Compositae**

### **CRISPR Lettuce? Manipulating Lettuce Shelf Life using Genome Editing**

**Annabelle Damerum**, University of California, Davis, CA

Lettuce is one of the most widely consumed vegetables in the US and bagged salads represent an increasing proportion of the market. However, additional processing of such convenience products can lead to a reduction in post-harvest quality, resulting in reduced product shelf life and vast amounts of waste. Previous research efforts have focused on increasing the understanding of the genetic basis of shelf life in lettuce, which proceeds via complex biochemical and physiological leaf changes. Through physiological assessment, extensive phenotyping of shelf life and leaf biophysical traits and QTL mapping studies, we have identified a rich candidate gene list. Here, we describe the use of the RNA-guided CRISPR/Cas genome editing system for the functional investigation of candidate genes for shelf life. Six cell wall modifying enzymes, xyloglucan endotransglucosylase/hydrolases (XTHs), identified within key QTL for shelf life traits were targeted for gene knockout in lettuce using a multiplex gRNA strategy. Target gRNA design attempted to achieve larger (>100 nucleotide) deletions by making multiple double-stranded breaks within the same gene, which could be observed during mutant screening stages. Additional constructs were designed to target homologous regions within genes for multiplex editing using a single gRNA. Single and double gene mutant lines were generated, with an array of mutations including small (1-2) nucleotide insertions and deletions and larger deletions (80-300

nucleotides). Preliminary investigations suggest mutants display modifications in shelf life, which may result from subtle differences in cell size and cell wall strength. This study demonstrates the successful application of CRISPR/Cas genome editing for multiplex gene targeting in lettuce.

## **W277: Compositae**

### **Phylogenomics Yields Insights into the Evolution and Biogeography of the Complex Genus *Antennaria***

**Ramhari Thapa**, Randall J. Bayer and Jennifer Mandel, University of Memphis, Memphis, TN

*Antennaria* are dioecious perennial herbs distributed mainly in the Holarctic Region with their major center of diversity in the Rocky Mountains of Western North America. The genus comprises 33 known sexual diploid/tetraploid species and at least five polyploid agamic complexes which mostly reproduce by forming asexual seeds. We performed a phylogenetic reconstruction of the 31 sexually-reproducing *Antennaria* species using a novel target enrichment method that employs custom capture probes and is designed to work across Asteraceae. Both concatenated and coalescent-based analyses of DNA sequence data from hundreds of nuclear loci recovered *Antennaria* as a monophyletic group except for the long-disputed species, *Antennaria linearifolia*, which was recovered outside of the genus. *Antennaria* was further resolved into three distinct, major lineages. Analysis of ancestral state reconstruction of 12 taxonomically important morphological characters elucidated patterns of character evolution throughout the genus. Estimations of ancestral geographic ranges and molecular dating analyses demonstrated the Rocky Mountain region, including the Vancouverian Province, as the center of origin for the genus *Antennaria*, around 5.8 MYA. Subsequent dispersals of *Antennaria* into the Arctic and Appalachian provinces, Canadian provinces, and Eurasia took place roughly 3.2 MYA, 2.4 MYA and 1.6 MYA, respectively. Biogeographical Stochastic Mapping indicated that 51.4% of biogeographical events were based on within-area speciation. The remaining 48.6% of the events were divided into two types of dispersals: i) range expansion dispersals (anagenic, 37%) and ii) founder/jump dispersals (cladogenic, 11.6%). Our results provide a framework for future evolutionary studies of *Antennaria*, including speciation, origin(s) of polyploidy, and agamospermy in the genus.

## **W278: Compositae**

### **The Cultivated Sunflower Pan-Genome Provides Evidence That Introgression from Wild Species Shaped the Gene Repertoire and Enhanced Disease Resistance**

**Sariel Hübner**, Galilee research institute (MIGAL), Israel and Loren Rieseberg, University of British Columbia, Vancouver, BC, Canada

The cultivated sunflower is known to be highly variable genetically despite having passed through domestication and improvement bottlenecks in the past 4,000 years. Sunflower is cross-compatible with numerous wild species, which are prized by breeders due to their enhanced tolerance to both biotic and abiotic stress. To determine the extent and nature of genetic diversity in cultivated sunflower, and to quantify potential contributions from introgression with crop wild relatives, we sequenced 476 sunflower genotypes, including 287 cultivars and 189 wild accessions representing 11 wild species. The outcome of these introgressions is variation among lines in the genomic composition and presence/absence of specific genes. Here we present the cultivated sunflower pan-genome which is comprised of 61,205 genes, of which ca. 27% vary across genotypes and represent so-called 'dispensable' genes. At least 10% of the cultivated sunflower pan-genome is derived through introgression from wild sunflower species, and ca. 1.5% of genes originated solely through post domestication introgression. Annotations and GO functional analyses further indicate that genes associated with biotic resistance are over-represented among introgressed regions, an observation consistent with breeding records. We further provide examples in which wild introgressions have contributed to resistance to Downey mildew.

## **W279: Compositae**

## **Characterization of Seven Polymorphic Genes Controlling Red Leaf Color in Lettuce and their Applications in Lettuce Improvement**

**Rong Tao**<sup>1,2</sup>, Jiongjiong Chen<sup>1</sup>, Richard Michelmore<sup>2</sup> and Hanhui Kuang<sup>1</sup>, (1)Huazhong Agricultural University, Wuhan, China, (2)The Genome Center, University of California, Davis, CA

Lettuce (*Lactuca sativa*) is one of the most important vegetable crops worldwide and shows dramatic variation in morphology and leaf color. Anthocyanin accumulation in lettuce was one of the first traits to be analyzed genetically in plants. Using genome-wide association studies, eQTLs, gene expression network as well as bi-parental segregation analysis, we identified seven loci controlling variation in leaf color of lettuce, and cloned five of them, which encode R2R3-MYB, R3-MYB, bHLH, WD40 and ANS, respectively. Mutations in the *bHLH* or *ANS* gene completely abolished flavonoid biosynthesis, resulting in pure green leaves. In contrast, mutations in the other three genes promoted anthocyanin accumulation, revealing of disruptive selection for leaf color in lettuce. Another two loci were fine mapped to a small region, and one of them originated from a spontaneous mutation discovered in our field, and the causal mutation is currently under investigation. Pyramiding color-promoting alleles of the seven loci generates dark red phenotype, and different combinations of alleles of these seven genes may generate different levels of anthocyanin. Green cultivars with high levels of colorless flavonoids may be produced when mutated *ans* gene is selected. Green cultivars with rich flavonoids and red cultivars with different color intensity would provide health-promoting lettuce products to consumers with diverse preferences of the color of lettuce leaves.

## **W280: Compositae**

### ***Stevia rebaudiana* : Cultivated and Landraces Phenotypic and Genotypic Diversity, Towards Yield Components Mapping**

**Zoé Le Bihan**<sup>1</sup>, Galadriel Brière<sup>2</sup>, Patrick Cosson<sup>3</sup>, Erwan Guichoux<sup>4</sup>, Cecile Hastoy<sup>5</sup>, Alexis Groppi<sup>2</sup>, Luis Ernesto Errazu<sup>6</sup>, Dominique Rolin<sup>3</sup>, Raluca Uricaru<sup>7</sup> and Valerie Schurdi-Levraud<sup>8</sup>, (1)INRA Bordeaux University, Biology and Fruit Pathology / Oviatis, Villenave d'Ornon, France, (2)Center of bioinformatic of Bordeaux (CBiB), Bordeaux, France, (3)INRA, Villenave d'Ornon cedex, France, (4)BIOGECO / Genomic center, Cestas, France, (5)Oviatis INRA, Lacaussade, France, (6)INTA-EEA, Famaillà, Argentina, (7)Informatic Laboratory of research in Bordeaux (LABRI), Talence, France, (8)INRA Bordeaux University, Villenave d'Ornon, France

Worldwide *Stevia rebaudiana* breeding improvement and genotype traceability suffer from a huge lack of information. Genomic information is particularly missing as no reference genome is available yet. Very few traceable and fixed cultivars are currently available for European producers and all have been selected outside of Europe. This absence of traceability makes difficult the comparison of cultivars' phenotypic evaluation over main agronomic traits and performed in different environments. Therefore, *Stevia* modern breeding is a real challenge to offer adapted and elite cultivars to producers.

This global objective means to be able to (1) get knowledge on available genotypic and phenotypic variability, (2) link genotypic and phenotypic variability in order to detect loci involved in main agronomic traits, such as biomass production, steviol glycosides content, favorable UGT alleles, response to pathogens and regrowth after winter time, (3) understand the genetic architecture of the traits and develop marker assisted selection for future breeding purposes.

We developed SSR markers and SNP through Reduced-Representation Library (RRL) sequencing approaches and adapted SNP calling pipeline. All these markers were used to analyze the genetic diversity of 145 worldwide cultivated and landraces genotypes. They allow us to classify and analyze *Stevia rebaudiana* genetic diversity.

Phenotypic evaluation of 15 cultivars was performed for 3 years in a field trial, and main agronomic traits were scored. This work allowed to identify cultivars of interest and selection criteria in our environmental

conditions. Contrasted cultivars were selected and crossed. A phenotypic evaluation was performed on two offspring for 2 years in a field trial.

Genetic map and QTL detection is ongoing.

Key words: *Stevia rebaudiana*; germplasm, molecular markers, SSR, SNP, phenotyping

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol. England*; 2013;22: 3151–3164. doi:10.1111/mec.12084

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One. United States*; 2011;6: e19379. doi:10.1371/journal.pone.0019379

Cosson P, Hastoy C, Ernesto Errazu L, Jorge Budeguer C, Boutié P, Rolin D, Schurdi-Levraud V. 2019. Genetic Diversity and Population Structure of the Sweet Leaf Herb, *Stevia Rebaudiana* B., Cultivated and Landraces Germplasm Assessed by EST-SSRs Genotyping and Steviol Glycosides Phenotyping . *BMC Plant Biology* 19 (1): 436. <https://doi.org/10.1186/s12870-019-2061-y>.

Hastoy C, Cosson P, Cavaignac S, Boutié P, Waffo-Teguo P, Rolin D, Schurdi-Levraud V. 2019. Deciphering performances of fifteen genotypes of *Stevia rebaudiana* in southwestern France through dry biomass and steviol glycoside evaluation. *Industrial Crops and Products* 128: 607-19. <https://doi.org/10.1016/j.indcrop.2018.09.053>.

Hastoy C, Le Bihan Z, Gaudin J, Cosson P, Rolin D, Schurdi-Levraud V. 2019. First report of *Septoria* sp. infecting *Stevia rebaudiana* in France and screening of *Stevia rebaudiana* genotypes for host resistance. *Plant Disease*. <https://doi.org/10.1094/PDIS-10-18-1747-RE>.

## **W281: Compositae**

### **The Importance of Phylogenomics in Understanding the Evolution of North American Compositae**

**Bort Edwards**, Smithsonian Institute, NMNH, Washington, DC, DC, Pamela S. Soltis, University of Florida, Gainesville, FL, Jennifer Mandel, University of Memphis, Memphis, TN and USGS Powell Center Geo-Eco-Evo Working Goup

The generation of multi-locus phylogenies for large numbers of taxa is critical to an understanding of many fundamental evolutionary processes and patterns. Yet while Big Data, in the form of global collections databases (eg. GBIF), soil maps (eg. SoilGrids250), and climate layers (eg. WorldClim) as well as machine learning are revolutionizing the scale at which ecological and character trait evolution can be studied, the lack of equivalent genomic resources in many groups hinders this research. The Compositae, comprising more than 1/10 flowering plants in North America are an exceptionally diverse and important component of the flora, and represent an unrivaled opportunity to study large-scale adaptive and evolutionary processes. However, a robust understanding of relationships within the family, even at the tribal level, is still elusive. Here we present phylogenetic diversity and ecological diversification analyses of North American Compositae with available metaphylogeny data to illustrate the scale and scope of questions that can be addressed with existing data, and call for an ongoing effort to build a cohesive phylogenetic understanding of this group.

## **W282: Computational Gene Discovery**

### **Stand-Alone PGAP: The NCBI Open-Source Pipeline for the Annotation of Prokaryotic Genomes**

**Terence D. Murphy**, Françoise Thibaud-Nissen, Douglas Slotta, Azat Badretdin, Wenjun Li, Daniel Soren and Wratko Hlavina, NCBI/NLM/NIH, Bethesda, MD

The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) combines ab initio gene prediction algorithms and homology-based methods for the annotation of bacterial and archaeal genomes. PGAP has become a reliable resource for the prokaryotic community. It is used to annotate RefSeq prokaryotes assemblies and is offered as a service to researchers submitting genome assemblies to GenBank.

We have re-factored PGAP into a stand-alone pipeline that scientists can download and execute on a personal computer, a computer grid or the cloud, to produce annotation on their own assemblies that is comparable to what NCBI would produce internally. The pipeline, available on Github (<https://github.com/ncbi/pgap/>), is written in the Common Workflow Language (CWL) and is packaged with the necessary binaries and cwltool, the CWL reference implementation. All necessary reference data, including manually curated evidence and other datasets, are bundled and distributed with the pipeline. A single Python convenience script downloads all required components and runs the pipeline. In addition to the annotation per se, the pipeline provides validation of the input data by verifying the taxonomic assignment based on comparison to type strains by Average Nucleotide Identity analysis. It also compares the size of the user-provided genomes to other genomes in GenBank for the same species and detects the presence of vector and adaptor sequences. The annotation results are produced in a variety of formats including ASN, so that the annotated genomes can be submitted to GenBank without further transformation.

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

## **W283: Computational Gene Discovery**

### ***Blastocrithidia*, a “Non-Stop” Trypanosome**

**Vyacheslav Yurchenko**<sup>1,2</sup>, Anna Nenarokova<sup>3,4</sup>, Kristína Záhonová<sup>3</sup>, Serafim Nenarokov<sup>3</sup>, Ambar Kachale<sup>3</sup>, Zdenek Paris<sup>3</sup> and Julius Lukeš<sup>3,4</sup>, (1)University of Ostrava, Ostrava, Czech Republic, (2)Martsinovskiy Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov University, Moscow, Russian Federation, (3)Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic, (4)Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Ciliates are notorious for their frequent use of alternative genetic codes. Recently, *Blastocrithidia* spp., which belong to the trypanosomatid flagellates (along with the genera *Trypanosoma* and *Leishmania*) became the first representative of this well-studied group that in its nuclear genome not only deviated from the canonical code, but even reassigned all three stop codons into sense codons. We have sequenced, assembled and analyzed the genomes and transcriptomes of two cultivable *Blastocrithidia* species and another trypanosomatid from the “*jaculum*” clade, which is the closest known relative of *Blastocrithidia* yet has a canonical genetic code. The annotation of the *Blastocrithidia* genome has been performed with a novel in-house generated software, since the available annotation programs could not deal with the ambiguous stop codons. This allowed us to dissect the general trends in the across-genome distribution and other features of the reassigned stop codons. Moreover, mass spectrometry analysis of the total *Blastocrithidia* proteome experimentally confirmed the predicted amino acids specified by the in-frame stop codons and showed that only one stop codon has a double meaning, as it is also used as a genuine stop. We have also predicted and experimentally analyzed tRNAs that are responsible for decoding of the stop codons. We are attempting to genetically modify *Blastocrithidia*, which would enable novel approaches to address questions such as: How functions its translation termination? How do in-frame stop codons influence translation? What triggered this massive reassignment?

## **W284: Computational Gene Discovery**

### **Fully Automated and Accurate Annotation of Eukaryotic Genomes with BRAKER2**

**Katharina J. Hoff**<sup>1,2</sup>, Tomas Bruna<sup>3</sup>, Alexandre Lomsadze<sup>3</sup>, Mario Stanke<sup>1,4</sup> and Mark Borodovsky<sup>3</sup>,  
(1)Institute for Mathematics and Computer Science, University of Greifswald, Greifswald, Germany,  
(2)Center for Functional Genomics of Microbes, University of Greifswald, Greifswald, Germany,  
(3)Georgia Institute of Technology, Atlanta, GA, (4)Center for Functional Genomics of Microbes,  
University of Greifswald, Germany

While the number of sequenced genomes is ever growing, a vast majority of already available eukaryotic genomes may not be utilized to its full potential since it is lacking a high quality annotation of protein coding genes.

Automation of the process of eukaryotic genome annotation is a challenging task due to diversity of input data situations.

BRAKER2 is an automated pipeline for annotation of protein coding genes in eukaryotic genomes. Common external data scenarios supported by BRAKER2 include the availability of i/ alignments of RNA-Seq short reads to the target genome, ii/ alignments of proteins of possibly distantly related species to the target genome or even iii/ absence of the evidence data. In all cases, BRAKER2 runs a self-training GeneMark-ET/-EP/-ES depending on the external data situation, trains AUGUSTUS on the genome annotation produced by GeneMark-ET/-EP/-ES and predicts genes (including alternative isoforms) with AUGUSTUS. Available extrinsic evidence is used by both tools.

To use cross-species proteins, BRAKER2 automatically calls a novel ProtHint pipeline introduced in GeneMark-EP (Bruna et al., 2020) for generating protein evidence for gene prediction with GeneMark-EP and AUGUSTUS. ProtHint enables users to map proteomes of a large number of species to the target genome. Recent improvements in genome annotation accuracy with protein evidence reached in GeneMark-EP lead to an increase in genome annotation accuracy by BRAKER2.

The BRAKER2 project locations are <https://github.com/Gaius-Augustus/BRAKER> and <https://github.com/gatech-genemark/BRAKER2>.

## **W285: Computational Gene Discovery**

### **BUSCO v4 and OrthoDB v10 Perspective on Genes and Genomes**

**Evgeny M. Zdobnov**, University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland

BUSCO (<https://busco.ezlab.org>) is a popular tool for completeness assessment of genome assemblies, transcriptomes, and predicted gene sets [1]. It is based on scoring the presence and coverage of marker genes, a set of universal single-copy orthologs expected to be found in one copy across the majority of genomes. The BUSCO scores thus provide an evolutionarily sound measure of gene content completeness, and it is complementary to assembly contiguity measures like the N50 value. Beyond the assessment of genomic completeness, BUSCO is useful for training gene predictors, phylogenomics, and for the evaluation of haplotypes [2-3]. The BUSCO package includes both software and datasets. The software is optionally made available as a docker container, which streamlines the installation of all dependencies. The provided BUSCO datasets correspond to selected gene markers for different phylogenetic clades, where a greater number of markers for more closely related organisms enable more precise evaluation. In this major upgrade, BUSCO release 4, we further expand the datasets and upgrade the code to feature new functionalities. We validated this update by comparing BUSCO v4 with v3 on metazoa and fungi. Estimates of completeness on bacteria and archaea of BUSCO v4 are comparable to those of CheckM. BUSCO is not limited to bacteria, however, and thus will properly evaluate microscopic eukaryotes often sequenced in the context of metagenomic analyses. It also consumes substantially less computing resources, in particular it requires less memory than CheckM, enabling BUSCO assessments to be run on laptops. We added an option to automatically select the most specific dataset to simplify BUSCO usage like in CheckM. BUSCO datasets are derived from OrthoDB (<https://www.orthodb.org>), the hierarchical catalogue of orthologs [4]. The latest release v10 provides the broadest sampling of the genomic space from all domains of life as well as some large viruses, and so BUSCO now covers the following datasets: 83 bacterial, 16 archaeal, 27 viral, 24 fungal, 7 protozoan, 15 vertebrate, 8 arthropod, 9 plant, plus mollusca and nematod sets. OrthoDB is not only the largest database of orthologs but we also have collated a lot of available functional gene annotations and computed evolutionary descriptors for the orthologs, with all the data publicly available via web GUI,

web REST service, data files, and Sparkle RDF. Moreover, the OrthoDB website allows user-submitted data analysis online, including the BUSCO evaluations.

1. Simão F.A., et al. 2015. "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs". // *Bioinformatics*. Vol.19. No.31. P.3210-2. doi:10.1093/bioinformatics/btv351
2. Waterhouse R.M., et al. 2017. "BUSCO applications from quality assessments to gene prediction and phylogenomics". // *Molecular Biology and Evolution*. Vol.35. No.3. P.543–548. doi:10.1093/molbev/msx319
3. Seppey M., et al. 2019. "BUSCO: Assessing Genome Assembly and Annotation Completeness". // M. Kollmar (eds.). *Methods in Molecular Biology: Eukaryotic Gene Prediction*. Springer Nature. doi:10.1007/978-1-4939-9173-0\_14
4. Kriventseva E.K., et al. 2019. "OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs". // *Nucleic Acids Res.* doi:10.1093/nar/gky1053

## **W286: Computational Gene Discovery**

### **Toward Complete Genome Assemblies**

**Pavel Pevzner**, Computer Science Dept., UCSD, La Jolla, CA

Long-read assemblies improved over the short read assemblies because of their greater ability to disambiguate genomic repeats. However, most algorithms for assembling long reads construct contiguous genomic segments (contigs) but do not provide the repeat characterization (*repeat graph*) necessary for producing optimal assemblies. We present the Flye algorithm for assembling long reads that does not attempt to construct contigs at the initial assembly stage but instead generates arbitrary paths (*disjointigs*) in the unknown repeat graph and constructs a repeat graph from these error-riddled disjointigs. Counter-intuitively, this seemingly reckless approach results in an accurate repeat graph and improves on the state-of-the-art long-read assemblers with respect to contiguity and speed. We further describe the development of the Flye assembly toolkit that includes metaFlye and tandemFlye.

This is a joint work with Mikhail Kolmogorov, Jeffrey Yuan, and Yu Lin.

## **W287: Computational Gene Discovery**

### **Toward Complete Genome Assemblies 2**

**Pavel Pevzner**, Computer Science Dept., UCSD, La Jolla, CA

Although variations in centromeres have been linked to cancer and infertility, centromeres still represent the "dark matter of genomes" and remain an enigma for both biomedical and evolutionary studies. Since centromeres have withstood all previous attempts to develop an automated tool for their assembly, recent efforts attempted to manually assemble centromeres using long error-prone reads. We describe the centroFlye algorithm for centromere assembly using long reads, apply it for assembling some human centromere, and use the constructed assembly to gain insights into centromere evolution. Our analysis opens a possibility to automatically close the remaining multi-megabase gaps in various genome and reveals many remaining challenges in centromere assembly.

This is a joint work with Andrey Bzikadze.

## **W288: Connecting Crop Phenotype and Genotype Data**

### **Transforming Breeding through Enterprise Breeding System and Analytics**

**Star Yanxin Gao**<sup>1</sup>, Thomas John Hagen<sup>2</sup>, Kelly Robbins<sup>1</sup>, Elizabeth Jones<sup>1</sup>, Marko Karkkainen<sup>3</sup>, Kate A. Dreher<sup>2</sup>, Pete Selby<sup>1</sup>, Eng Hwa Ng<sup>4</sup>, Michael Olsen<sup>5</sup>, Jens Riis-Jacobsen<sup>2</sup>, Susanne Dreisigacker<sup>2</sup>, Joshua N. Cobb<sup>6</sup>, John Damien Platten<sup>3</sup>, Maria Ymber V. Reveche<sup>3</sup>, Jean-Marcel Ribaut<sup>7</sup>, Jan Erik Backlund<sup>8</sup>, Lukas Mueller<sup>9</sup>, Andrzej Kilian<sup>10</sup>, George Kotch<sup>11</sup>, Eduardo Covarrubias<sup>11</sup>, Gustavo Teixeira<sup>11</sup>, Sarah Hearne<sup>2</sup>, Michael Quinn<sup>11</sup>, Marianne Banziger<sup>2</sup> and Gary Atlin<sup>12</sup>, (1)Cornell University, Ithaca, NY, (2)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (3)International Rice Research Institute, Los Baños, Philippines, (4)Excellence in Breeding, Laguna, Philippines, (5)CIMMYT (International Maize and Wheat Improvement Center), Nairobi, Kenya, (6)International Rice Research Institute, Los Baños, Laguna, Philippines, (7)Integrated Breeding Platform, Texcoco, Mexico, (8)Integrated Breeding Program, Texcoco, Edo de México, Mexico, (9)BTI, Ithaca, NY, (10)Diversity Arrays Technology, Canberra, Australia, (11)Excellence in Breeding, CIMMYT, Texcoco, Mexico, (12)Bill & Melinda Gates Foundation, Seattle, WA

As world population, urbanization of farm lands, and climatic change increase, food insecurity becomes an increasing threat in densely populated, low income countries. To solve food security, the Bill & Melinda Gates Foundation has funded numerous projects, with the ultimate goal that CGIAR and partner breeding programs can deliver crop varieties with increased genetic gain (yield) and improved on-farm results. These projects include the Integrated Breeding Platform (**IBP**), Transforming Rice Breeding (**TRB**), Next Generation Cassava, and the Genomic Open-source Breeding informatics initiative (**GOBii**), that have developed the **BMS**, **B4R**, **Breedbase**, and **GOBii-GDM** etc. data management systems, respectively. These data management systems will be provided by the Excellence in Breeding (**EiB**) Platform to CGIAR and partner breeding programs along with **BrAPI** and many other cutting-edge tools, high throughput phenotyping and genotyping services, and best practices.

From 2020 on, a high quality modern crop-agnostic web-based enterprise breeding system will be continuously built under the EiB platform, called Enterprise Breeding System (**EBS**), which is designed for breeding institutions to adopt the software at low costs and with high impact. EBS' core breeding operations are based on B4R while genotypic data management is based on the GOBii-GDM. The EBS supports major domains of crop breeding programs including Breeding Program Planning, Experiment Creation, Seed Inventory Management, Sample Workflow Management, Breeding Trial Analytics, Decision Making, and Reporting.

Successful execution of the EBS IT project largely depends on synergistic leadership and close coordination between the development teams and application teams at IRRI, CIMMYT, and Cornell University. The success of transforming breeding can't be accomplished just by the successful development and launch of the EBS alone without the adoption by CGIAR and partner breeding programs. EiB collectively will support the adoption of systems and tools as well as implementation of best practices to enable integrated breeding workflows.

## **W289: Connecting Crop Phenotype and Genotype Data**

### **Development and Validation of a Low-Cost, Rapid NIRS-Based Phenotyping Approach for Improving Cassava Root Quality Traits**

**Jenna M. Hershberger**<sup>1</sup>, Bryan Ellerbrock<sup>2</sup>, Lukas A. Mueller<sup>2</sup>, Robert Kawuki<sup>3</sup>, Eder Oliveira<sup>4</sup>, Ismail Rabbi<sup>5</sup> and Michael A. Gore<sup>1</sup>, (1)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (2)Boyce Thompson Institute, Ithaca, NY, (3)National Root Crops Resources Research Institute, Namulonge, Uganda, (4)Embrapa, Brasília, Brazil, (5)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

Over 800 million people across the tropics rely on cassava as a major source of calories. While the dry matter content (DMC) of this starchy root is important for both growers and consumers, characterization of DMC by traditional methods is time-consuming and laborious. Near-infrared spectroscopy (NIRS) has been proposed as an alternate phenotyping method, but while it is highly predictive of DMC in cassava roots, spectrometers that have been validated for high accuracies are prohibitively expensive, limiting

their usefulness. For this reason, we investigated the use of a low-cost, handheld NIR spectrometer (SCiO, ConsumerPhysics) for field-based DMC prediction in cassava roots. Pilot investigation into the predictive effects of preprocessing techniques, number of roots sampled per plot, and within-root sampling location were used to develop a scanning and sampling protocol. Following this protocol, oven-dried measurements of DMC were paired with scans of roots of diverse clones from IITA (Nigeria), NaCRRRI (Uganda), and Embrapa (Brazil) and grouped into training and test sets based on prediction scenarios common to plant breeding programs. Partial least squares regression models were evaluated for predictive ability, which ranged from  $R^2=0.51-0.85$  depending on the cross-validation scheme. With appropriate calibration, this spectrometer will allow for field-based collection of spectral data with a smartphone for accurate DMC prediction, a step that could be easily integrated into existing harvesting workflows of cassava breeding programs. These and other NIRS models will be hosted on BreedBase to facilitate further analysis and incorporated into the PhenoApps suite of Android applications for plant phenotyping.

## **W290: Connecting Crop Phenotype and Genotype Data**

### **Prioritization of Genetic Variants by Biological and Evolutionary Annotation: Functional Assessments in Diverse Maize Populations**

**Guillaume Ramstein**<sup>1</sup>, Sarah Jensen<sup>2</sup>, Hai Wang<sup>3</sup>, M. Cinta Romay<sup>1</sup> and Edward S. Buckler<sup>1</sup>,  
(1)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (2)Cornell University, Ithaca, NY,  
(3)Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Ithaca, NY

Genomic prediction is a useful tool for estimating genetic values and quickly selecting individuals within breeding programs. However, genomic prediction models typically identify effects of genomic markers that are confounded by genetic linkage to the actual causal variants. Therefore, such models typically cannot generate useful insight about the genetic architecture of traits, and tend to lose accuracy across population backgrounds (e.g., prediction in a bi-parental breeding population based on a model trained in a different diverse population). Based on two panels of 1,106 and 1,640 maize hybrids, we show the usefulness of annotations about gene action (additive and dominance effects), genic regions (distance to genes), and evolutionary constraint (cross-species allelic conservation) for genomic prediction across population backgrounds, whereas annotations about marker effects from genome-wide association studies could not contribute to any increase in prediction accuracy. We further show the limitations of current methods for incorporating annotation information in genomic prediction models, due to low density of marker assays, and limited statistical power for estimating the importance of various annotations. To address these limitations, we present an innovative imputation method based on haplotype graphs (the Practical Haplotype Graph), which should provide an exhaustive assay of genetic variability around genes. Furthermore, we present machine learning methods for incorporating various biological annotations of genetic variants (recombination rate, chromatin openness, expression effects, gene ontology, etc.) based on evolutionary constraint. Our results show promising avenues for effectively incorporating biological information into genomic models and increasing their robustness in cross-population prediction.

## **W291: Connecting Crop Phenotype and Genotype Data**

### **Empowering Global Rice Breeding Programs using Genomic Selection**

**Parthiban Prakash**<sup>1</sup>, Juan David Arbelaez Velez<sup>2</sup>, Holden Verdeprado<sup>3</sup>, Vitaliano Lopena<sup>4</sup>, Rose Imee Zhella<sup>4</sup>, Jerome Bartholome<sup>5</sup>, Jessica Rutkoski<sup>6</sup>, Rosemary Murori<sup>7</sup>, Alexis Ndayiragije<sup>8</sup>, Mohammad Rafiqul<sup>9</sup>, Sanjay Kumar Katiyar<sup>10</sup> and Joshua N. Cobb<sup>11</sup>, (1)International Rice Research Institute, Los Baños, Philippines, (2)University of Illinois Urbana Champaign, (3)International Rice Research Institute, Los Baños, Philippines, (4)International Rice Research Institute, Philippines, (5)International Rice Research Institute | CIRAD, Philippines, (6)University of Illinois Urbana Champaign, Los Baños, (7)International Rice Research Institute, Tanzania, United Republic of, (8)International Rice Research Institute, Mozambique, (9)International Rice Research Institute, Bangladesh, (10)International Rice

Research Institute (IRRI), Manila, India, (11)International Rice Research Institute, Los Baños, Laguna, Philippines

Rice is the staple food for half of the global population and irrigated rice contributes 70% of total rice produced. Given the strategic importance of this market segment to global food security, the irrigated rice breeding program at IRRI is mandated to breed varietal rice and empower rice breeding programs in South Asia and Eastern and Southern Africa. As public sector breeding budgets are insufficient to adequately test all new lines across such a varied environmental landscape, advanced molecular tools such as genomic selection (GS) are critical to achieving high levels of selection intensity and selection accuracy within each region. For the past two years, the irrigated lowland breeding program at IRRI has distributed carefully selected 'estimation sets' containing a few hundred breeding lines to global partners in Africa and South Asia with the intention of using genomic prediction to evaluate several thousand selection candidates that have only been genotyped. Combining different software tools such as B4R (phenotypic data management), GOBii (genotypic data management) and ASReml-R (modeling) into an analytical workflow, we have generated initial prediction accuracies for grain yield, plant height, and flowering time which are routinely used in the program to make optimized breeding decisions for each of these disparate global environments. An overview of the results will be presented as well as a discussion about the future integration of sparse testing designs, multi-environmental models, use of weather data through machine learning, and phenotyping with unmanned aerial vehicles (UAV).

### **W292: Connecting Crop Phenotype and Genotype Data**

#### **Utilizing Phenomics, Genomics, and Deep Learning Methods to Accelerate the Development of Climate Resilient Crop Varieties**

**José Crossa**, International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico, Osva Montesinos-Lopez, University of Colima, Colima, CL, Mexico, Abelardo Montesinos-Lopez, University of Guadalajara, Guadalajara, JA, Mexico, Jaime Cuevas, University of Quintana-Roo, Chetumal, QR, Mexico, Paulino Perez-Rodriguez, Colegio de Postgraduado, Texcoco, EM, Mexico, Margaret Karuse, Researcher, Texcoco, EM, Mexico, Juan Burgueño, CIMMYT, El Batán, Mexico and Fernando Toledo, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Hyperspectral reflectance phenotyping and genomic selection are two technologies that have the potential to increase prediction accuracy for grain yield under different environmental conditions. Hyperspectral cameras quantify canopy reflectance associated with numerous biophysical and biochemical processes in plants, and genomic selection models utilize dense molecular markers or/and pedigree information to predict the breeding values of lines. Authors have proposed Bayesian functional regression models that take into account all available bands, genomic or pedigree information, the main effects of lines and environments  $\times$  environment interaction ( $G \times E$ ) and wavelength  $\times$  environment interaction ( $B \times E$ ). Results show that models with  $B \times E$  interaction terms were the most accurate models, whereas the functional regression models (with B-splines and Fourier basis) and the conventional models performed similarly in terms of prediction accuracy. However, the functional regression models are more parsimonious and computationally more efficient because the number of beta coefficients to be estimated are lower (number of basis) than estimating all regression coefficients for all bands. Also other studies have proposed hyperspectral reflectance-derived relationship matrices to model the genotype  $\times$  environment ( $G \times E$ ) interactions across environments. Multi-kernel models combining marker/pedigree information with hyperspectral reflectance phenotypes had the highest prediction accuracies. Last years several genomic-enabled predictions models have been developed with increasing prediction accuracy over the genomic best linear unbiased estimator (GBLUP). These methods are the Gaussian kernel (GK) and Arc-cosine (AK) kernel. The AK kernel method emulates the deep machine learners (DL) and are much easier to implement and with less computational time and similar or slightly higher genomic-enabled prediction accuracy than the GK kernel. Genomic DL methods have been useful for assessing big data comprising multi-trait multi-environment. Implementing the multi-trait DL models is feasible but challenging due to the large number of hyper-parameters involved.

### **W293: Connecting Crop Phenotype and Genotype Data**

## **Integrating Genotype and Phenotype in Real-Time Trait-Targeted Wheat Breeding**

**Mary J. Guttieri**, USDA Agricultural Research Service, Manhattan, KS

Breeding information management systems (BIMS) have been a focus of significant public and private investment. The benefits of these systems to working breeding programs are many, including consistent data structures, well-defined trait ontologies, data security, and sophisticated analysis pipelines. These information systems now interact with mobile devices for data collection. Increasingly, available BIMS include components that incorporate genotyping information. The primary focus of BIMS has been the management of inbred line and hybrid phenotyping trial information. Such germplasm is typically genotypically characterized only once and on a standard platform, inbred lines and hybrids are considered genetically reproducible, and phenotypic heterogeneity is rare. Breeding information management during early generation selection, and particularly during trait-targeted, marker-assisted crossing cycles, presents additional challenges for BIMS. Accessions are inherently transient as progeny are not equivalent to the maternal plant. Phenotypic data typically are heterogeneous. Moreover, a wide array of genotypic information may be used for selection. Incorporating genotypic information for selection within sequential breeding generations of annual crops requires rapid turnaround of genotyping information, which may be obtained internally in breeding programs. Breeding management systems can provide data structures that support interaction with genotypic data collection activities and visualization tools. The rapidly evolving nature of genotyping and phenotyping technologies challenges the adoption of new information management systems. New data management technologies are adopted in real time in breeding programs, affecting ongoing operations. Breeding programs therefore have an inherent reluctance to adopt entirely new information management systems. Modular information management tools can facilitate the transition to systems which integrate phenotype and genotype data for crop improvement.

## **W294: Cool Season Legumes**

### **Genomic Selection in Cool Season Legumes**

Luke Pembleton<sup>1</sup>, Hossein Kahrood<sup>1</sup>, Shimna Sudheesh<sup>1</sup>, Babu Pandey<sup>2</sup>, Laura James<sup>2</sup>, Dianne Noy<sup>2</sup>, German Spangenberg<sup>1,3</sup>, Garry Rosewarne<sup>2</sup> and **Sukhjiwan Kaur**<sup>1</sup>, (1)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, Australia, (2)Agriculture Victoria, Grains Innovation Park, Horsham, VIC, Australia, (3)School of Applied Systems Biology, La Trobe University, Bundoora, VIC, Australia

**Sukhjiwan Kaur**, Agriculture Victoria

Genomic selection (GS), which incorporates genome wide marker information to predict the breeding values of individuals in a breeding population, is rapidly becoming an important selection tool in plant breeding. One of the major advantages of using GS in plant breeding is its ability to predict the phenotypic performance of individuals early in the breeding cycle, hence, reducing the generation interval and thus increasing genetic gain. Here we evaluate its efficiency for breeding lines of lentil and field pea from Australian pulse breeding programs. A total of c. 2,000 and 1,500 advanced breeding lines were obtained from Pulse Breeding Australia (PBA) lentil and field pea breeding programs, respectively. These lines were evaluated infield annually in a range of environments from 2010-2017 for lentil and from 2013-2017 for field pea, for a range of economically important traits including grain yield, grain weight, disease resistances and abiotic stress tolerances. A genotyping-by-sequencing approach was used to genotype the breeding material and over 150,000 SNPs were identified from both lentil and field pea. The ability to genomically predict the observed phenotypic performance was explored by forward prediction and applying a range of genomic selection models as well as incorporating GxE components. Genomic selection has now been fully implemented into the lentil breeding program and prediction equations for yield, grain weight, boron and salt tolerance as well as botrytis grey mould resistance have been derived with moderate to high prediction accuracies (0.35-0.70). Optimal crossing schemes have also been designed to increase genetic gain and reduce the overall generation interval. Significant progress has been made towards establishment of GS in field pea program with prediction equations

derived for ascochyta blight and downy mildew resistance, boron and salt tolerance with prediction accuracies ranging from 0.23-0.52.

## **W295: Cool Season Legumes**

### **Comparison of Alternative Methods of GWAS for Mapping Phenology Traits in Lentil**

**Teketel A. Haile**, University of Saskatchewan, Saskatoon, SK, Canada

Understanding the genetic control of quantitative traits is an important step for marker assisted breeding. Association mapping, or linkage disequilibrium (LD) mapping, has been used extensively to dissect complex traits in plants. Association mapping is commonly performed by testing each marker for association with the trait of interest. Advances in high-throughput genotyping and sequencing technologies have resulted in the availability of abundant single nucleotide polymorphisms (SNPs), which requires testing hundreds of thousands of variants for their association with the phenotype. Simultaneously testing large number of variants which exhibit LD leads to the problem of multiple testing. The commonly used approaches to correct for multiple testing are either too stringent because they fail to account for LD or computationally intensive. Selecting a smaller subset of tag SNPs that represent blocks of haplotypes or grouping markers in strong LD into multi-marker haplotypes, have been proposed as ways to reduce the number of tests without reducing power. Alternatively, Bayesian regression methods that are used for genomic prediction have been proposed for GWAS to simultaneously estimate the effects of all markers without significance testing. Our objective was to compare single marker, haplotype-based, and Bayesian approaches to identify loci associated with phenological traits in a lentil diversity panel of 324 genotypes collected from around the world. These lines were phenotyped for phenological traits across five environments in Saskatchewan and genotyped using a custom exome capture assay. Significant marker-trait associations and the potential of each approach to reduce false positives will be discussed.

## **W296: Cool Season Legumes**

### **Putting the Pulse Back into Chickpeas – How Exploiting Wild Germplasm to Expand the Genetic Diversity could Revamp the Adaptive Potential of Domesticated Chickpea**

**Silke Jacques**<sup>1</sup>, Toby Newman<sup>1</sup>, Christy Grime<sup>1</sup>, Fiona Kamphuis<sup>1</sup>, Yuphin Khentry<sup>1</sup>, Robert Andrew Syme<sup>2</sup>, Robert Lee<sup>1</sup> and Lars G Kamphuis<sup>1</sup>, (1)Centre for Crop and Disease Management, Curtin University, Perth, WA, Australia, (2)Canadian Centre for Computational Genomics, Montreal, QC, Canada

Chickpea (*Cicer arietinum* L.) is an important legume crop, which is produced and consumed worldwide, with Australia being the second largest producer and the largest exporter of chickpea [1]. However, chickpea production is constrained by several biotic and abiotic stresses, as well as its particularly narrow genetic base [2].

In order to overcome these limitations and, ultimately, improve the currently stagnant yields of chickpea for Australian growers, wild relatives can be exploited. To this end, we are utilising a collection of wild *Cicer reticulatum* and *Cicer echinospermum* accessions, which are the direct progenitor and a sister species of chickpea, respectively [3].

With a view to expand the genetic diversity of chickpea, diverse wild genotypes have been crossed with an elite Australian chickpea variety, namely PBA HatTrick. Segregating populations derived from these crosses have been assessed in multiple field trials covering different soil types over different years for various agronomically important traits, such as flowering time and growth habit. Genotyping of these populations via Diversity Arrays Technology (DART sequencing) enabled us to investigate the genetic basis of these valuable traits to ultimately provide the breeders with genetic markers associated with these traits of interest.

In addition to this, we have taken a more targeted approach to improving resistance to *Ascochyta* blight, a devastating fungal disease caused by *Ascochyta rabiei*. Screening of the wild collection has revealed some promising accessions that displayed resistance against a diverse set of isolates, hinting at possible sources of durable ascochyta blight resistance. A genome wide association study using restriction site associated DNA sequencing (RAD Sequencing) pointed to new loci associated with ascochyta blight resistance. Mapping populations are being generated to elucidate the underlying QTLs associated with resistance.

### **References:**

[1] FAOSTAT data, 2017, <http://www.fao.org/faostat/en/#data>

[2] Abbo S. et al., 2003, *Functional Plant Biology*, 30(10), pp.1081-1087.

[3] von Wettberg E.J.B. et al., 2018, *Nature Communications*, 9(1), p.649.

## **W297: Cool Season Legumes**

### **The Faba Bean Pan-Transcriptome**

**Leandro Escobar-Herrera**<sup>1</sup>, Jonathan Kreplak<sup>2</sup>, Marcin Nadziejka<sup>3</sup>, Morgane Terezol<sup>4</sup>, Catherine Desmetz<sup>4</sup>, Gregoire Aubert<sup>2</sup> and Stig Uggerhøj Andersen<sup>5</sup>, (1)Aarhus University, Aarhus C, Denmark, (2)INRAE, UMR1347 Agroécologie, Dijon, France, (3)Aarhus University, Denmark, (4)INRAE, UMR1347 Agroécologie, France, (5)Aarhus University, Aarhus, Denmark

Faba bean (*Vicia faba*) is a widely adapted nitrogen fixing grain legume that produces seeds with high protein content. It is an agriculturally attractive crop with a good disease resistance profile, but targeted breeding is hampered by the lack of a reference genome and gene set. Here we develop a faba bean gene expression atlas and pan-transcriptome. The pan-transcriptome was generated using data from four different accessions (Hedin, Hiverna, 153b, 2378), including data from both aerial and root tissues. The quality and utility of the pan-transcriptome was assessed by mapping reads from accessions that had not contributed data to the pan-transcriptome. Here, the high mapping rates suggested that the pan-transcriptome represents a comprehensive faba bean reference gene set and that it will be a useful resource for differential gene expression analyses and genome annotation. In addition, the gene expression atlas provides valuable information to support gene function discovery.

## **W298: Cool Season Legumes**

### **Integrated Omics Approach Reveals Drought Tolerance Mechanism in Chickpea (*Cicer arietinum* L)**

**Himabindu Kudapa**, ICRISAT, Hyderabad, India

Food legume crops play important role in contributing to food security worldwide. However, legume production has been adversely affected by drought stress in many regions of the world. Chickpea (*Cicer arietinum* L.), is a cool season legume crop providing nutritious food for the expanding world population. Drought in chickpea is one of the most devastating stress adversely affecting yield and yield components. Understanding how the underlying genome sequence translates into specific plant phenotypes under drought stress and information on molecular mechanisms combating drought stress is crucial. To investigate changes in transcripts, relative proteins and metabolites in response to drought, an integrative -omics approach combining transcriptomics, proteomics and metabolomics has been performed. Root tissues of four contrasting drought responsive chickpea lines (ICC 4958, JG 11 and JG 11+, drought-tolerant and ICC 1882, drought-sensitive) under control as well as drought stress conditions were considered for the study to identify differential accumulation of transcripts, proteins and metabolites. These lines represent parents of two intraspecific populations, ICC 4958 × ICC 1882 and ICC 4958 × JG 11, while JG 11+ was a near isogenic line derived from ICC 4958 × JG 11, which

possessed the “QTL-hotspot” region. This study led to identification of a total of 3,956 differentially expressed genes, 876 differentially expressed proteins and 34 metabolites across four genotypes. Furthermore, antibiotic biosynthesis, galactose metabolism and isoflavonoid biosynthesis pathways were significantly associated with drought tolerance mechanism. In summary, the study established integrated view of the tolerance mechanisms of chickpeas towards drought stress and identification of candidate genes/ proteins/ metabolites for use in chickpea breeding. In addition, it also offered novel insights into the molecular processes involved and emphasize the importance of systems approach to uncover stress regulatory mechanisms in chickpea.

## **W299: Cool Season Legumes**

### **The Pea Genome, Now and after ...**

**Jonathan Kreplak**<sup>1</sup>, Gregoire Aubert<sup>1</sup>, Valentin Delefortrie<sup>2</sup>, Virginie Bourion<sup>2</sup>, Karine Gallardo<sup>1</sup>, Judith Burstin<sup>1</sup> and The Pea Genome Consortium, (1)INRAE, UMR1347 Agroécologie, Dijon, France, (2)INRAE, UMR1347 Agroécologie, France

Having a genome sequence available is a critical step towards unravelling functional diversity and establishing genome-enabled breeding. The recently generated pea genome sequence represents a great tool for genomicists, geneticists and breeders not only for the pea community but also for legume research. In the genome project, re-sequencing data revealed the considerable diversity present in the *Pisum* genus. In the PeaMUST and GRASP project, an unprecedented effort was made to genotype large pea collections using the exome capture technology. This high density SNP data was exploited in genome-wide association studies (GWAS) on a large number of traits related to yield, symbiose, as well as response to biotic and abiotic stresses. Furthermore, transcriptomics experiences were able to unravel or confirm mechanisms linked to nutrient remobilization in leaves or seeds. We will present snapshots of these results using the genome, discuss limits of this assembly and how we can improve it.

## **W300: Cool Season Legumes**

### **Long Read Assemblies Reveal the Complexity of *Lens* Genomes**

Larissa Ramsay<sup>1</sup>, Chu Shin Koh<sup>2</sup>, **Carolyn T. Caron**<sup>1</sup>, Clarice J Coyne<sup>3</sup>, David J Konkin<sup>4</sup> and Kirstin Bett<sup>1</sup>, (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)USDA-ARS-WRPIS, Pullman, WA, (4)National Research Council Canada, Saskatoon, SK, Canada

The *Lens culinaris* genome has gone through multiple iterations of sequencing over a wide range of sizes and technologies. Our previous short-read assembly (v1.2) covered 2.7GB of 4.3GB expected sequence but was riddled with collapsed sequences and we experienced many difficulties with sequence ordering. The most recent assembly (v2.0) uses long-reads combined with HiC to provide massive improvements in contiguity and identification of full repeats. The new assembly spans nearly 3.7GB with a contig N50 of 1,352,216 base pairs and has been made accessible for BLAST and visualization on JBrowse through KnowPulse (<https://knowpulse.usask.ca>).

The genomes for *L. lamottei*, *L. odemensis* and *L. ervoides* have also been completed. In the near future, there will be complete assemblies for *L. tomentosus*, *L. orientalis* and *L. nigricans*. Early comparisons among these genome assemblies show structural variations worthy of further investigation in both inter and intraspecific populations.

## **W301: Crop Evolution Genomics & Future Agricultural Productivity**

### **Discovery of Chromosomal Inversions in Domesticated Barley**

**Martin Mascher**, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and The International Barley Pangenome Consortium, IPK Gatersleben, 06466, Germany

Chromosomal inversions occur in natural populations of many species and may underlie reproductive isolation and local adaptation. We have recently developed a method to use three-dimensional contact probabilities between genomic loci as assayed by chromosome conformation capture sequencing (Hi-C) to detect multi-megabase polymorphic inversions. We applied our method for inversion discovery to a barley diversity panel comprising 70 accessions, mainly domesticated genotypes sampled from the German *ex situ* genebank. In this panel, we identified dozens of inversions relative to the Morex reference genome, most of them occurring at low frequencies and residing in the low-recombining, proximal regions of the chromosomes. Two notable exceptions were detected on chromosomes 2H and 7H. The 7H inversion spans a 140 Mb interval in the distal region of that chromosomes and occurs at a low frequency in current elite cultivars. The 2H inversion is present only in barleys of Western origin, where it segregates at high frequency. Its boundary is close to *HvCEN*, a well-known regulator of flowering time in barley, hinting at a possible association to range expansion. In summary, Hi-C-based inversion discovery has broadened the spectrum of sequence variation accessible to population genomic studies in crops.

### **W302: Crop Evolution Genomics & Future Agricultural Productivity**

#### **Long-Read Sequence Evidence of Structural Variants in Wild Barley**

Li Lei, DOE Joint Genome Institute, Berkeley, CA, Skylar R. Wyant, University of California, Giulia Frascarelli, Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy, Samuel Hamann, University of Minnesota, Adrian O. Stec, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, Chaochih Liu, Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN and **Peter L. Morrell**, Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN

Chromosomal structural variants can make important contributions to adaptive genetic variation. Inversions in particular can preserve locally adaptive variants, even in the presence of extensive gene flow. Evidence for the presence of inversions or other large structures variants is often encountered in population-level studies of SNP genotyping or resequencing datasets. Putative structural variants are often identified because of the observation of extensive chromosomal regions with elevated linkage disequilibrium, allele frequency differentiation between populations, or an excess of common variants. Long read DNA sequencing can provide physical evidence for putative structural variants. We use linked reads and nanopore resequencing to examine structural variants in a wild barley (*Hordeum vulgare* ssp. *spontaneum*) accessions where putative structural variants have been associated with adaptation to variation in precipitation and temperature.

### **W303: Crop Evolution Genomics & Future Agricultural Productivity**

#### **Separating the Barley from the Chaff: Studying Adaptation in Structured Wild Barley Populations**

**Sariel Hubner**, Galilee Research Institute, Kiryat Shmona, Israel

Global crop production is challenged by rapid population growth, declining natural capital and dramatic climatic turnovers. These challenges urge plant breeders to explore new ways to enhance adaptation and sustainability in crops. Crop wild relatives have been long prized by breeders due to their enhanced tolerance to biotic and abiotic stress, yet their implementation in breeding programs is limited. To extend the use of crop wild relatives in breeding, high quality information on the genetics, physiology, and environmental context of the wild species is required. In this talk, I will present a new comprehensive wild barley germplasm collection comprised of 300 accessions that represent a wide ecological spectrum including extreme environments. This collection was characterized phenotypically in a common garden and whole genome sequence data was generated for all accessions. The unique sampling design of this collection allows for the identification of signals of selection at high resolution in spite of the strong ecotypic structure observed among wild barley populations.

### **W304: Crop Evolution Genomics & Future Agricultural Productivity**

#### **Evolution of *Sorghum halepense* and its Potential for Perennial Agriculture**

**Wenqian Kong**<sup>1</sup>, Pheonah Nabukalu<sup>2</sup>, T. Stan Cox<sup>2</sup> and Andrew H. Paterson<sup>3</sup>, (1)University of Georgia, Athens, GA, (2)The Land Institute, Salina, KS, (3)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

*Sorghum halepense* (2n = 40), a tetraploid hybrid derived from a natural cross of *S. bicolor* (2n = 20) × *S. proinquum* (2n = 20), is the one of the world's most noxious weeds, having reached all continents except Antarctica. It has extensive subterranean stems (rhizomes) making it capable of overwintering in temperate latitudes and regrowing in early spring before cultivated sorghum can be seeded. The ability of *S. halepense* to cross with *S. bicolor* makes it a paradigm for 'crop- to-weed' gene escape, but also an opportunity to breed for perennial crops, a topic recently gaining attention to alleviate soil degradation to improve ecosystem services. Previous effort has been focused on understanding gene escape and discovering agriculturally important traits in tetraploid populations derived from *S. bicolor* × *S. halepense*. Recent success in creating diploid progeny from *S. bicolor* × *S. halepense* crosses opens new avenues to study transmission genetics in sorghum and facilitates breeding perennial sorghum in both temperate and tropical environments.

### **W305: Crop Evolution Genomics & Future Agricultural Productivity**

#### **Co-Evolution of Crops, Birds, and Farmers in African Agroecosystem Linked by Secondary Metabolites**

Yuye Wu<sup>1</sup>, Tingting Guo<sup>2</sup>, Qi Mu<sup>2</sup>, Jinyu Wang<sup>2</sup>, Xin Li<sup>2</sup>, Yun Wu<sup>2</sup>, Bin Tian<sup>3</sup>, MingLi Wang<sup>4</sup>, Guihua Bai<sup>5</sup>, Ramasamy Perumal<sup>1</sup>, Harold N. Trick<sup>3</sup>, Scott Bean<sup>6</sup>, Ismail Dweikat<sup>7</sup>, Mitchell Tuinstra<sup>8</sup>, Geoffrey P. Morris<sup>1</sup>, Tesfaye T. Tesso<sup>1</sup>, Jianming Yu<sup>2</sup> and **Xianran Li**<sup>2</sup>, (1)Kansas State University, Manhattan, KS, (2)Iowa State University, Ames, IA, (3)Kansas State University, Department of Plant Pathology, Manhattan, KS, (4)USDA-ARS, Griffin, GA, (5)USDA, Hard Winter Wheat Genetics Research Unit, Manhattan, KS, (6)USDA-ARS, Manhattan, KS, (7)University of Nebraska-Lincoln, Lincoln, NE, (8)Purdue University, WEST LAFAYETTE, IN

Domestication is an evolutionary process of humans selecting desired characterizes from wild progenitors to fit human needs and local environments. All three components, human (domesticator), crop (domesticate), and environment are essential for domestication to take place.

Among major cereals domesticated as staple foods, only sorghum (domesticated in Africa) has a high proportion of cultivars with condensed tannins, which can trigger bitter perception in animals by binding to type 2 taste receptors (TAS2Rs). We identified a pair of duplicate recessive genes (*Tan1* and *Tan2*) underlying the presence of tannins in sorghum grains. Three loss-of-function alleles from each gene were identified in non-tannin sorghum desired as a palatable food. Following the serendipitous observation that condensed tannins effectively prevented sparrows from consuming sorghum grain, we uncovered parallel geographic distributions between tannin sorghum and the red-billed quelea bird, supporting the role of tannins in fighting against this major herbivore threat in Africa. Association between geographic distributions of human *TAS2R* variants and tannin sorghum suggested that people in areas where tannin sorghum are predominantly grown are more likely to carry the non-taster alleles in *TAS2Rs*, presumably better tolerance to the bitter taste from tannins.

With the uncovered genetic evidence and the parallel distributions, condensed tannins probably played important roles in African agroecosystem. According to local environments and human taste sensitivity, the balance between natural and artificial selection resulted in contrasting directions: condensed tannins were selected for to reduce herbivore damages in East and South Africa and were selected against to produce palatable food in West Africa. Crop domestication is an intricate process of dynamically balancing the interactions among the plant-human-environment triangle.

## **W306: Crop Evolution Genomics & Future Agricultural Productivity**

### **Can Wild Beans be a Source of Higher Yield and Drought Tolerance for Domesticated Beans?**

Jorge C. Berny Mier y Teran<sup>1</sup>, Eneas R. Konzen<sup>2,3</sup>, Viviana Medina<sup>1</sup>, Antonia Palkovic<sup>1</sup>, Andrea Ariani<sup>1</sup>, Matthew E. Gilbert<sup>1</sup>, Siu M. Tsai<sup>4</sup> and **Paul Gepts**<sup>1</sup>, (1)University of California, Davis, CA, (2)Universidade de Sao Paulo, Brazil, (3)Universidade Federal do Rio Grande do Sul, Brazil, (4)University of São Paulo - CENA, Piracicaba, Brazil

Domestication has reduced genetic diversity in common bean (*Phaseolus vulgaris* L.). This observation suggests that additional, agronomically useful variation may be present in the wild ancestor. Yet, utilization of this variation for bean improvement is difficult because of yield evaluation difficulties, identification of useful variation, and linkage drag. The lack of adaptation to cultivation conditions and the existence of highly structured populations make association mapping of diversity panels suboptimal. Joint linkage mapping of nested populations avoids the later constraint, while populations crossed with a common domesticated parent allow the evaluation of wild variation within a more adapted background. Three domesticated by wild backcrossed-inbred-line populations (BC1S4) were developed using three wild accessions representing the full range of rainfall of the Mesoamerican wild bean distribution crossed to the elite drought tolerant domesticated parent SEA 5. These populations were evaluated under field conditions in three environments, two fully irrigated trials in two seasons and a simulated terminal drought in the second season. The goal was to test if these populations responded differently to drought stress and contained progenies with higher yield than SEA 5, not only under drought but also under water-watered conditions. Results revealed that the two populations derived from wild parents of the lower rainfall regions produced lines with higher yield compared to the domesticated parent in the three environments, i.e., both in the drought-stressed environment and in the well-watered treatments. Several progeny lines produced yields, which on average over the three environments were 20% higher than the SEA 5 yield. Twenty QTLs for yield were identified in 13 unique regions on eight of the 11 chromosomes of common bean. Five of these regions showed at least one wild allele that increased yield over the domesticated parent. The variation explained by these QTLs ranged from 0.6 to 5.4 % of the total variation and the additive effects ranged from -164 to 277 kg ha<sup>-1</sup>, with evidence suggesting allelic series for some QTLs. Separate greenhouse experimentation showed that wild beans originating from central and north-west Mexico and Oaxaca, in the driest parts of their distribution, produced more biomass and were deeper-rooted. Compared with domesticated beans, wild beans showed a more limited reduction and delay in growth and development in response to drought stress. Our results underscore the potential of wild variation, especially from drought-stressed regions, for bean crop improvement as well the identification of regions for efficient marker-assisted introgression. They also show that successful introgression from wild beans depends on the careful choice of wild accessions and traits to be introgressed.

## **W307: Crop Genomics for Global Food Security**

### **Identification and Mapping of Heterotic Loci (HLs) from *Oryza rufipogon* to Enhance Yield Potential of Hybrid Rice Parental Lines**

**Revathi Ponnuswamy**<sup>1</sup>, V Mahesh Jagadale<sup>2</sup>, Sandeep G<sup>2</sup>, Satendra Mangrauthia<sup>3</sup> and Meenakshi Sundaram Raman<sup>2</sup>, (1)Crop Improvement-Hybrid rice section ICAR - Indian Institute of Rice Research (IIRR), Hyderabad, India, (2)ICAR-IIRR, India, (3)ICAR-Indian Institute of Rice Research, Hyderabad, India

Rice is the most important staple food crop for more than half of the world's population. After the deployment of semi-dwarf varieties, hybrid rice technology has been the major strategy for raising further genetic yield potential of rice. India produces 106 million tonnes rice from 44 million hectares as compared to China which produces 137 million tonnes from 30 million hectares. This is due to the large-scale adoption of hybrid rice by China, while in India the progress in adopting hybrid rice has been slow. Hybrid-rice technology has the potentiality to further increase the rice production through enhanced productivity, thus ensuring global food security. The yield improvement in recently released hybrids in India is negligible over the varieties indicating yield stagnation in the hybrids. This could be due to very narrow genetic diversity among the parental lines. Utilizing the huge variability available within wild

species of rice and identification and transfer of superior alleles to cultivated species has emerged as a novel option so as to enhance the productivity traits. Marker assisted backcross breeding (MABB) and AB-QTL mapping to identify yield enhancing heterotic QTLs from wild species to develop pre-breeding material to widen the genetic base of parental lines is under progress at IIRR. Though the superiority of wild introgression lines over recurrent parents have been reported by many studies but the heterotic effect of introgression in F<sub>1</sub> hybrids is very limited. We had developed *O. rufipogon* wild introgression lines in the background of restorers and maintainers of hybrid rice. The wild introgression lines of BC<sub>2</sub>F<sub>2</sub> and BC<sub>2</sub>F<sub>4</sub> generation selected lines were crossed with WA based CMS lines viz., APMS 6A, CRMS 32 A and IR 68897A to produce F<sub>1</sub>s hybrids. The selected 192 lines were genotyped with 50K rice SNP chip. About 400 F<sub>1</sub>s rice hybrids along with checks were evaluated for heterosis and combining ability in the L X T design for identifying lines possessing heterotic loci (HL) for yield contributing traits. Some of the rice hybrids expressed higher heterosis for yield in comparison with check hybrids and wild introgression lines. This indicates that wild introgression lines in homozygous phenotype may not express positively for yield related traits because of replacement of a cultivated genome segment. However, in heterozygous condition, the recipient complement is conserved and minor negative effect associated with wild segment is likely to be masked there by expressing higher heterosis. In this study, genotypic and phenotypic data of BC<sub>2</sub>F<sub>2</sub> and BC<sub>2</sub>F<sub>4</sub> has been utilized to map yield enhancing QTLs and heterotic loci. Thus, the present study emphasis on identification and mapping of heterotic loci (HLs) from wild rice for improving heterosis in rice.

### **W308: Crop Genomics for Global Food Security**

#### **Comparison of 16 Wild and Domesticated Sorghum Genomes Reveals Extensive Variation in the Sorghum Pan-Genome**

**Yongfu Tao**<sup>1</sup>, Hong Luo<sup>2</sup>, Jiabao Xu<sup>3</sup>, Alan W Cruickshank<sup>4</sup>, Fei Teng<sup>3</sup>, Tracey Shatte<sup>4</sup>, Xiaoyuan Wu<sup>2</sup>, Yuanming Liu<sup>2</sup>, Haichun Jing<sup>2</sup>, David Jordan<sup>5</sup> and Emma Mace<sup>5</sup>, (1)The University of Queensland, Warwick, QLD, Australia, (2)Institute of Botany, Chinese Academy of Sciences, Beijing, China, (3)BGI, SHENZHEN, China, (4)Department of Agriculture and Fisheries, Queensland Government, Warwick, QLD, Australia, (5)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, QLD, Australia

Sorghum provides a major source of food, feed, fibre and biofuel. Cultivated sorghum and its inter-fertile wild relatives have a broad geographical distribution and constitute the primary gene pool for sorghum. The rich diversity within this resource is the raw material upon which human selection act during domestication and modern breeding to reshape its phenology and productivity to meet human needs. To better understand diversity in the sorghum primary gene pool, we have constructed the sorghum pan-genome using 16 de novo genome assemblies. We de novo assembled 13 sorghum genomes including cultivated sorghum and its wild relatives via a hybrid sequencing strategy combining Illumina short reads and PacBio long reads. Analysis of the sorghum pan-genome was conducted using these 16 assembled genomes, which revealed a sorghum pan-genome consisting of 61,564 predicted gene families. Of these, approximately 30% (18,141 gene families) were present in ≥15 sorghum genomes, and defined as core genes, 33% (20,439) were present in 2-15 sorghum genomes, and defined as shell genes, and 37% (22,984) were present in only one sorghum genome, and defined as cloud genes. Compared to core genes, shell genes were found to be shorter with less exons, but with higher SNP density and higher non-synonymous/synonymous substitution ratios, indicating that shell genes are less functionally conserved. Comparisons with the reference genome identified 1.23-5.31 million SNPs and 20,386-148,669 PAVs in each comparison. Genes affected by PAVs were enriched with defence response functions. Selection signals on PAVs during sorghum domestication were identified and associations between PAVs and agronomical traits were detected. This study presents the first pan-genome analysis in sorghum with thorough investigation on PAVs. Results from this study will be a good resource for genetic and genomic analysis in sorghum and other crops.

### **W309: Crop Genomics for Global Food Security**

#### **Capturing Cereal Biodiversity for Food Security**

**Robert J. Henry**, University of Queensland/QAAFI, Brisbane, QLD, Australia; Queensland Alliance for Agriculture and Food Innovation, Sta Lucia, Australia

Food security requires increasing efforts to deliver the quantities of food required by increasingly discerning consumers. The adaptation of agriculture to climate change is also essential for food security. Wild crop relatives provide a resource for improving current crops to achieve these objectives. The domestication of new plant species is another option. The potential to work with grasses to deliver both improved and new domesticated cereal species will be explored. Recent progress in defining useful diversity in the genomes of species of *Oryza* and *Sorghum* will be described. Capture of this genetic variation requires an understanding of evolutionary relationships, reproductive barriers and characterization of desirable alleles in wild populations.

### **W310: Crop Genomics for Global Food Security**

#### **Genomic Prediction to Accelerate the Rate of Genetic Gain in Chickpea for Providing Nutritional Food Security**

**Manish Roorkiwal**<sup>1</sup>, Bharadwaj Chellapilla<sup>2</sup>, Nicholas Santantonio<sup>3</sup>, GP Dixit<sup>4</sup>, Diego Jarquin<sup>5</sup>, Srinivasan Samineni<sup>1</sup>, Elizabeth Jones<sup>3</sup>, Kelly Robbins<sup>3</sup>, Narendra P Singh<sup>4</sup>, José Crossa<sup>6</sup> and Rajeev K Varshney<sup>1</sup>, (1)ICRISAT, Hyderabad, India, (2)ICAR-IARI, Delhi, India, (3)Cornell University, Ithaca, NY, (4)ICAR-Indian Institute of Pulses Research, India, (5)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, (6)International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico

Chickpea (*Cicer arietinum*) is the second most important food legume globally, which plays a key role in ensuring the nutritional food security. Average chickpea productivity has been restricted to ~ 1 t h<sup>-1</sup> due to several biotic and abiotic stresses. Prolonged use of conventional breeding approaches have started to fall short of meeting the yield and nutrition demands. To address the issues related to complex traits such as yield which is controlled by multiple QTLs, genomic selection (GS) approach can be very useful in crop breeding to capture several genes with minor additive effects. GS offers breeders to select lines prior to field phenotyping using genotyping data, resulting in reduced cost and shortening of selection cycles. Initial results on GS in chickpea using 320 elite breeding lines suggested high prediction accuracies for diverse yield and yield related traits. Inclusion of G x E effects in GS models has shown significant improvement of prediction accuracies in breeding programs. In order to assess the potential of GS in chickpea breeding program 6000 F<sub>5</sub> lines from 12 different crosses from ICRISAT and IARI breeding programs were selected and genotyped using DArTseq platform. After merging the markers from the training and prediction sets, which were run on different DArT marker platforms (DArTseq and LD DArT), about a thousand markers were used to run the prediction models. The cross validation prediction accuracy were run with a 10-fold consolidation scheme. Each cross validation was repeated 10 times with new random folds, and the mean of the prediction accuracies was calculated. To compare the potential of GS models, two set of ~200 lines each were identified based on visual selection by breeder and based on genomic prediction based GEBVs. Both of these set were evaluated in the field conditions during 2018-19. Selection efficiency of GS over visual phenotypic selection was found significantly better. Genomic prediction based line selection over visual selection saves time and cost involved in large scale screening of populations.

### **W311: Crop Genomics for Global Food Security**

#### **Genomic, Metabolic, and Transcriptomic Responses of the Extremophile Grass *Paspalum vaginatum* to Nutrient Deficit Stress**

**Guangchao Sun**<sup>1,2</sup>, Nishikant Wase<sup>2,3</sup>, Toshihiro Obata<sup>2,3</sup> and James C Schnable<sup>1,2</sup>, (1)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, (2)Center for Plant Science Innovation, University of Nebraska-Lincoln, (3)Department of Biochemistry, University of Nebraska-Lincoln

Many wild grass species have been shaped by natural selection to thrive under environmental conditions or resource constraints far outside the range experienced by crop species. And improved understanding

of the molecular and evolutionary strategies natural selection employed to achieve these changes in stress tolerance and nutrient use efficiency in crop wild relatives has the potential to aid efforts to engineer resilient and low input crops and advance food security. Here we focus on *Paspalum vaginatum* a crop wild relative which is multiple abiotic stress tolerant and exhibits greater tolerance to both nitrogen and phosphorous deficiency. *Paspalum* exhibits no significant decrease in biomass accumulation under nitrogen or phosphorous deficient conditions that significantly impact the biomass accumulation of maize and sorghum. All three species exhibit increased root elongation and branching in response to nitrogen deficit. Metabolomic and transcriptomic analyses identified many commonalities in the molecular responses to stress in all three species. However, uniquely, *paspalum* exhibits significantly increases accumulation of trehalose under nutrient deficit conditions, and genes involved in metabolic pathways leading to trehalose production are experiencing more rapid protein sequence evolution in the lineage leading to *paspalum* than in other grass species. Efforts are underway to experimentally test the link between *paspalum*'s unique strategy of accumulating trehalose in response to stress and *paspalum*'s resilience to nutrient deficit.

### **W312: CSSA: Translational Genomics**

#### **Marker Set Optimization for Genomic Selection**

**Avjinder Kaler**, University of Arkansas, Fayetteville, AR

Advancement in high-throughput genotyping and sequencing technologies provided fast and low-cost molecular markers, particularly single nucleotide polymorphisms (SNPs), covering the whole genome. Genomic selection uses all markers for estimating the genomic estimated breeding values by regressing phenotypic values on the markers. Marker density is one of major factors that affects the prediction accuracy. Basic assumptions in genomic prediction is that the markers are scattered throughout the genome to attain sufficient coverage so that at least one marker is in linkage disequilibrium (LD) with QTLs. Both larger and smaller number of markers can affect the accuracy. In this study, different subset approaches, including stride, thinning based on the physical positions, linkage disequilibrium, hybrid (adding GWAS markers in the set), and random selection methods, were compared at the different marker density levels using related and unrelated target sets in four datasets in three crop species (Maize, Rice, and Soybean). We observed prediction accuracy increases with increasing marker density in all four datasets until it reaches a plateau. Stride selection for soybean data and thinning selection for maize and rice data performed better than other methods. As expected of higher accuracy in the related than distant set, however, increasing marker density did not show any increase in prediction accuracy in the distant target. Low heritability traits need more markers to reach a plateau. Adding GWAS markers give a boost in prediction accuracy, but as marker density increases, GWAS markers do not show any effect.

### **W313: CSSA: Translational Genomics**

#### **Maize Yabby Genes Drooping leaf1 and Drooping leaf2 Regulate Plant Architecture**

**Erik W. Vollbrecht**, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA

### **W314: CSSA: Translational Genomics**

#### **CRISPR-Cas12 Genome Engineering Systems in Plants**

**Yiping Qi**, University of Maryland, College Park, MD

Cas12a (formerly Cpf1), a type V-A CRISPR effector RNA-guided DNA endonuclease, has been widely used for plant genome editing in recent years. Cas12b is a type V-B CRISPR effector RNA-guided DNA endonuclease. Like Cas12a, Cas12b recognizes T-rich PAMs and generates staggered DNA double strand breaks. While Cas12a only requires crRNA to function, Cas12b requires both crRNA and tracrRNA or their engineered fusion known as single guide RNA (sgRNA). This feature resembles the CRISPR-Cas9 system and enables guide RNA engineering, making Cas12b advantageous over Cas12a

in certain genome engineering applications. Here, we describe our recent efforts on developing a new plant genome engineering platform based on Cas12b. We first compared multiple Cas12b orthologs of different bacterial origins for genome editing in rice, an important crop. Among them, we identified a potent ortholog for targeted mutagenesis, which was further demonstrated in multiplexed genome editing in stable transgenic lines. Next, we engineered three Cas12b based repressors and showed that they could mediate targeted transcriptional repression at different levels. Finally, we compared over a dozen transcription activation systems based on Cas12b in plants. We found the most potent transcription activation system relies on both Cas12b protein and engineered sgRNAs for the recruitment of different transcriptional activators. With the demonstration of Cas12b for genome editing, CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa), our work comprehensively establishes Cas12b as the third promising CRISPR system, after Cas9 and Cas12a, for plant genome engineering.

### **W315: CSSA: Translational Genomics**

#### **Unbiased Genome-Wide Screening of Regulatory Elements Provides Targets for CRISPR-Mediated Optimization of Gene Expression Levels in Plants**

**Lotte Westerhof**, Ferdinand Los and Rudi Ariaans, Hudson River Biotechnology, Wageningen, Netherlands

CRISPR is only as powerful as the genetic information available. For plants, although great strides have been made in terms of gene annotations, currently there is still limited information on regulatory elements controlling gene expression (GREs; promoters and enhancers). Access to such information would elevate CRISPR from an on/off switch to a volume regulator for gene expression. This greatly enhances possibilities for crop improvement via CRISPR, especially since complete knockout of trait influential genes may be detrimental to plant health overall, and many causative SNPs used as breeding markers are often found in between genes rather than inside open-reading frames. HRB pursues technologies that enable genome-wide identification of GREs, such as SuRE (Survey of Regulatory Elements), a technology owned by Gen-X that we have jointly translated for use in plants. Combining GRE data with e.g. gene expression data can provide novel targets for CRISPR, enabling targeted mutations that alter the level of gene expression rather than shutting off a gene completely. Moreover, these technologies enable the screening of populations of plants in combination with large-scale phenotyping data to identify causal mutations. We will discuss how we analyzed the tomato genome for GREs and demonstrated the validity of this data.

### **W316: CSSA: Translational Genomics**

#### **Incorporating Epigenetic Variation into Directed Breeding for Enhanced Crop Performance**

**Sally A Mackenzie**, Pennsylvania State University, University Park, PA

Epigenetic variation is known to exist in plants, but the extent that it influences plant phenotype has not been comprehensively investigated. We have developed epigenetically altered populations of Arabidopsis and sorghum to investigate the range and behavior of phenotypic plasticity that is associated with epigenetic effects. Investigation of genome-wide DNA methylation repatterning identified four integrated gene networks that participate in epigenomic reprogramming, and mutant analysis has identified central network hubs and pathways that underpin the reprogramming process. Field studies of epigenetically modified sorghum lines demonstrated that epi-types that performed well under optimal conditions were distinct from lines that performed well under severe stress. We will present a strategy that combines genetic selection with epigenetic variation to enhance environmental resilience for crop performance under unstable conditions.

### **W317: CSSA: Translational Genomics**

#### **Mapping Transcription Factor Binding Sites in C4 Grasses for Predictable Crop Engineering**

**Steven Burgess**, University of Illinois at Urbana Champaign, Urbana, IL

Crop engineering has traditionally involved simple manipulations, such as introducing herbicide or pesticide resistance, but as efforts expand to target complex processes it will become increasingly important to limit transgene expression to specific cell types to avoid unwanted side-effects. One example is engineering photosynthesis: in C4 grasses, photosynthesis gene expression is compartmented between mesophyll and bundle-sheath cells and this must be taken into consideration when introducing transgenes. However, there are only a small number of commonly used promoters for cereal engineering and these are predominantly constitutive. In order to begin to understand how to design cell specific promoters, we performed DNaseI sequencing on mesophyll and bundle sheath tissues to identify genome-wide profiles of transcription factor binding in leaves of the C4 grasses *Zea mays*, *Sorghum bicolor*, and *Setaria italica* as well as C3 *Brachypodium distachyon*. In C4 species, while bundle-sheath strands and whole leaves shared similarity in the broad regions of DNA accessible to transcription factors, the short sequences bound varied. Transcription factor binding was prevalent in gene bodies as well as promoters, and many of these sites could represent duons that influence gene regulation in addition to amino acid sequence. Although globally there was little correlation between any individual DNaseI footprint and cell-specific gene expression, within individual species transcription factor binding to the same motifs in multiple genes provided evidence for shared mechanisms governing C4 photosynthesis gene expression. These data therefore provide insight into the architecture associated with C4 photosynthesis gene expression in particular and characteristics of transcription factor binding in cereal crops in general.

### **W318: Cucurbit Genomics**

#### **Resequencing of 1,175 Melon Accessions Reveals Multiple Domestication Events and Loci Influencing Agronomic Traits**

Guangwei Zhao<sup>1</sup>, Qun Lian<sup>2</sup>, Zhonghua Zhang<sup>3</sup>, Qiushi Fu<sup>3</sup>, **Jordi Garcia-Mas**<sup>4</sup>, Yongyang Xu<sup>1</sup> and Sanwen Huang<sup>2</sup>, (1)Chinese Academy of Agricultural Sciences, Zhengzhou, China, (2)Chinese Academy of Agricultural Sciences, Shenzhen, China, (3)Chinese Academy of Agricultural Sciences, Beijing, China, (4)Center for Research in Agricultural Genomics, Barcelona, Spain

The genetic basis and history of melon domestication is largely unknown. The resequencing of 1,175 accessions representing the diversity of the species allowed obtaining a comprehensive melon genomic variation map, revealing that three independent domestication events occurred in melon. We also detected two independent sets of domestication sweeps during melon breeding, corresponding to the two subspecies *melo* and *agrestis*. Potential domestication loci were identified related to fruit mass, flesh bitterness and acidity. Genome-wide association studies for 16 agronomic traits identified 208 loci significantly associated with fruit mass, quality and morphological characters. Among them, candidate genes for fruit traits as rind sutures, peel and flesh colour were identified. This large-scale genome resequencing study provides a valuable resource for genomics-assisted breeding in this important crop.

### **W319: Cucurbit Genomics**

#### **Cloning of a Melon Fruit Shape QTL Contributes to Identify a Common Mechanism that underlies Organ Morphological Diversity in Plants**

Cecilia Martinez-Martinez<sup>1</sup>, Pablo Sipowicz<sup>2</sup>, Irene Martínez-Fernández<sup>3</sup>, Maria José Gonzalo<sup>3</sup>, Maria Dolores Gomez<sup>3</sup>, Miguel Angel Perez-Amador<sup>3</sup>, Mohamed Zouine<sup>4</sup>, Konstantinos Alexiou<sup>5</sup>, Jordi Garcia-Mas<sup>6</sup>, Carlos Romero<sup>3</sup> and **Antonio J. Monforte**<sup>3</sup>, (1)Universidad de Almería, La Cañada de San Urbano, Spain, (2)INTA. Estación Experimental Agropecuaria Manfredi., Córdoba, Argentina, (3)IBMCP. CSIC-UPV, Valencia, Spain, (4)INPT-ENSAT, Toulouse, France, (5)IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Bellaterra, Spain, (6)IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Barcelona, Spain

Melon cultivars display a panoply of fruit morphology diversity. We have studied thoroughly a fruit shape (FS) QTL (FSQS8.1) detected in a cross between the Indian accession PI124112 (elongated shape) and the Spanish cultivar "Piel de Sapo" (PS, oval shape). FSQS8.1 is particularly interesting because is responsible of transgressive segregation for FS in this cross (PI124112 allele induces round shape),

interacts epistatically with a repressing its pleiotropic effects on fruit elongation, and maintains its effects on different genetic backgrounds. An introgression line (CALC\_8-1) was developed introgressing FSQS8.1 into PS background, yielding round fruit (FS index=1). Effects of FSQS8.1 were already apparent at early stages of ovary development. Fine-mapping showed that FSQS8.1 is located in a genomic region with important structural variation, that was verified by *de novo* assembling of CALC\_8-1 genome by 10 X Genomics Chromium sequencing technology. Ultra-resolution mapping suggested that FSQS8.1 is encoded by *CmOFP13* a member of the Ovate Family Proteins (OFP). RT-PCR analysis showed a differential expression of *CmOFP13* during early ovary development between PS and CALC\_8-1, suggesting that the variation in morphology is caused by differential expression. *CmOFP13* is likely the homologous of *SIOFP20* in tomato. *SIOFP20* interacts with the member of TONNEAU1 Recruiting Motif family (TRM) *SITRM5* in the pathway to determine FS. *SIOFP20* and *SITRM5* homologues have been also found in other species underlying variability in organ morphology. These findings suggest that *OFPs* and *TRMs* are acting in a common mechanism involved in organ growth among in all species.

### **W320: Cucurbit Genomics**

#### **Fruit Size and Shape in Cucurbits: A Comparative Perspective**

Yiqun Weng, USDA ARS/ University of Wisconsin, Madison, WI

The cucurbits (family Cucurbitaceae) are known for their diverse fruit sizes and shapes. In recently years, many studies have been conducted to identify QTL for fruit size (FS), shape (FSI), and fruit weight (FW) in major cucurbits. We reviewed the literature on FS/FSI/FW QTL identified in cucumber, melon and watermelon, from which consensus QTL for these traits were inferred and their positions were compared. Genome-wide survey of the three cucurbit genomes identified homologs of eight classes of fruit or grain size/weight-related genes cloned in Arabidopsis, tomato and rice including those encoding proteins containing the CNR (cell number regulator), CSR (cell size regulator), CYP78A (cytochrome P450), SUN, OVATE, TRM (TONNEAU1 Recruiting Motif), YABBY, and WOX domains. Alignment of the consensus QTL with candidate gene homologs suggested structure and function conservation of fruit size/shape gene homologs in cucurbits. The *andromonoecy* and the *carpel number* (for CLAVATA3) loci have been shown to have pleiotropic effects on cucurbit fruit shape, which may complicate identification of fruit size/shape candidate genes. Findings from this investigation may facilitate identification and cloning of fruit size/shape QTL in cucurbits through the candidate gene approach.

### **W321: Cucurbit Genomics**

#### **QTL and Transcriptomic Analyses Implicate Cuticle Transcription Factor Shine as a Source of Natural Variation for Epidermal Traits in Cucumber**

Stephanie Rett-Cadman<sup>1</sup>, Marivi Colle<sup>1</sup>, Ben N. Mansfeld<sup>2</sup>, Cornelius S. Barry<sup>2</sup>, Yuhui Wang<sup>3</sup>, Yiqun Weng<sup>4</sup>, Lei Gao<sup>5</sup>, Zhangjun Fei<sup>5</sup> and **Rebecca Grumet**<sup>1</sup>, (1)Department of Horticulture, Michigan State University, East Lansing, MI, (2)Michigan State University, East Lansing, MI, (3)Department of Horticulture, University of Wisconsin, Madison, WI, (4)USDA ARS/ University of Wisconsin, Madison, WI, (5)Boyce Thompson Institute, Cornell University, Ithaca, NY

The fruit surface is a unique tissue with multiple roles influencing fruit development, post-harvest storage and quality, and consumer acceptability. Serving as the first line of protection against herbivores, pathogens, and abiotic stress, the surface can vary markedly among species, cultivars within species, and developmental stage. In this study we explore developmental changes and natural variation of cucumber (*Cucumis sativus* L.) fruit surface properties using two cucumber lines which vary greatly for these traits and for which draft genomes and a single nucleotide polymorphism (SNP) array are available: Chinese fresh market type, Chinese Long '9930' (CL9930), and pickling type, 'Gy14'. Thin-section samples were prepared from the mid-region of fruit harvested at 0, 4, 8, 12, 16, 20, 24 and 30 days post pollination (dpp), stained with Sudan IV and evaluated for cuticle thickness, depth of wax intercalation between epidermal cells, epidermal cell size and shape, and number and size of lipid droplets. 'Gy14' is characterized by columnar shaped epidermal cells, a 2-3 fold thicker cuticular layer

than CL9930, increased cuticular intercalations between cells and a larger number and larger sized lipid droplets. In both lines maximal deposition of cuticle and increase in epidermal size coincided with exponential fruit growth and was largely completed by approximately 16 dpp. Phenotyping and quantitative trait locus mapping (QTL) of fruit sampled from an F<sub>7</sub>:F<sub>8</sub> Gy14 × CL9930 recombinant inbred line (RIL) population identified QTL regions on chromosomes 1, 4 and 5. Strong QTL for epidermal cell height, cuticle thickness, intercalation depth, and diameter of lipid droplets co-localized on chromosome 1. SSR markers on chromosome 1 were used to screen for recombinants in an extended RIL population to refine the QTL region. Further fine mapping by KASP assay combined with gene expression profiling suggested a small number of candidate genes. Tissue specificity, developmental analysis of expression, allelic diversity and gene function implicate the regulatory factor *CsSHINE1/WIN1* as a source of natural variation for cucumber fruit epidermal traits.

## **W322: Cucurbit Genomics**

### ***De novo* Genome Assembly of Sweet Watermelon Relatives and a Pan-Genome of *Citrullus* Species**

**Shan Wu**<sup>1</sup>, Lei Gao<sup>1</sup>, Sandra E. Branham<sup>2</sup>, Patrick Wechter<sup>3</sup>, Shaker Kousik<sup>4</sup>, Amnon Levi<sup>4</sup>, Yong Xu<sup>5</sup> and Zhangjun Fei<sup>1</sup>, (1)Boyce Thompson Institute, Cornell University, Ithaca, NY, (2)Clemson University, Charleston, SC, (3)USDA-ARS, Charleston, SC, (4)USDA, ARS, U.S. Vegetable Laboratory, Charleston, SC, (5)Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

Sweet watermelon (*Citrullus lanatus*) is among the most important vegetable crops in the world. It belongs to the *Citrullus* (2n=2x=22) genus, which includes six other species, namely egusi watermelon (*C. mucospermus*), citron watermelon (*C. amarus*), colocynth (*C. colocynthis*), *C. ecirrhosus*, *C. rehmi* and *C. naudinianus*. To maximize the capture of genome variations within and among these *Citrullus* species and to identify novel agronomically important alleles for facilitating watermelon breeding, we have generated high quality reference genomes using selected individuals in *C. lanatus*, *C. mucospermus*, *C. amarus* and *C. colocynthis*. Structural variations were identified between species, some of which could affect local recombination and underlie the observed segregation distortion in populations derived from interspecific crosses. We are in the process of constructing pan-genomes for these *Citrullus* species using whole genome resequencing data. A pan-genome for each of the four species will be constructed by combining the reference genome and *de novo* assembled novel non-redundant sequences. Presence-absence variations (PAVs) of protein-coding genes will be analyzed, and collections of core and dispensable gene sets for each species will be identified. Finally, comparative analysis of the four pan-genomes will be performed and a catalog of orthologous gene relationships across four species will be identified to highlight syntenic regions and species-specific variations. The *Citrullus* pan-genome resource will enable us explore genes and alleles underlying disease resistance and other important horticultural traits and enhance the efficiency of watermelon breeding.

## **W323: Cucurbit Genomics**

### **Genomics-Assisted Breeding of Fusarium Wilt Resistance in Watermelon**

**Sandra E. Branham**, Clemson University, Charleston, SC, Amnon Levi, USDA, ARS, U.S. Vegetable Laboratory, Charleston, SC and Patrick Wechter, USDA-ARS, Charleston, SC

Fusarium wilt, caused by the soil-borne, fungal pathogen *Fusarium oxysporum* f.sp. *niveum*, is one of the most devastating diseases of watermelon (*Citrullus lanatus*). Multiple independent evaluations of the USDA *Citrullus* plant introduction (PI) collection for resistance to *Fon* race 2 have identified resistant *Citrullus amarus* (citron melon) germplasm, with a number of these screens failing to find resistance in any other species within the genus. Citron melon, although recently reclassified from a watermelon subspecies (*C. lanatus* subsp. *citroides*) to a separate species (*Citrullus amarus*), readily crosses with cultivated watermelon and has substantially higher genetic diversity. Although the first resistant citron melon was identified in 1989, there are no edible watermelon cultivars available with resistance to *Fon* race 2. Breeding has been hampered by polygenic inheritance, strong environmental effects and the necessity of interspecific introgression from wild citron melon. *Fon* race 2 resistance QTL have been

identified from multiple citron melon sources through both traditional QTL mapping and a genome-wide association study. Interspecific Kompetitive Allele-Specific PCR markers linked to a major *Fon* race 2 QTL were designed from whole genome resequencing data. We will describe our initial efforts to move resistance QTL into cultivated watermelon through marker-assisted selection.

### **W324: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **CyVerse 2020**

**Eric Lyons**, The University of Arizona, Tucson, AZ

CyVerse is a National Science Foundation supported project whose mission is to design, develop and deploy a national cyberinfrastructure to enable basic and applied research, and to train scientists in its use. With more than 70,000 users worldwide in all domains of science, CyVerse is enabling advances across diverse scientific disciplines including human health, animal and plant agriculture, geology, ecology and the environment, astronomy, and defense. Regardless of the field of science, CyVerse provides a highly functional infrastructure for data management, data analysis, training and collaborations, with more than 4PB of private research data being managed and actively analyzed. A major challenge for CyVerse is meeting the diverse needs of scientists using its infrastructure encompassing the full range of data and computational expertise. This talk will provide an overview of CyVerse, review the major features released in CyVerse over the past year, and discuss our future development roadmap for 2020.

### **W325: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **i5k and CyVerse: Tools for Predicting Genome-Wide Function**

**Fiona McCarthy**<sup>1</sup>, Amanda M. Cooksey<sup>1</sup>, Anna Childers<sup>2</sup> and Monica Poelchau<sup>3</sup>, (1)University of Arizona, Tucson, AZ, (2)USDA-ARS, BELTSVILLE, MD, (3)USDA-ARS, National Agricultural Library

The goal of the i5k Project is to sequence 5,000 arthropod genomes and to develop resources and best practices for the arthropod community. We utilize CyVerse resources to provide accessible suites of tools that functionally annotate proteomes from newly assembled genomes, providing Gene Ontology and KEGG pathway annotation. These tools are available on the CyVerse Discovery Environment and as containers that can be installed and deployed locally, along with detailed documentation. We are currently in the process of functionally annotating a diverse set of arthropod proteomes and determining community guidelines for this process. During the next year we will co-ordinate with the i5k community and interested arthropod researchers to provide functional annotations and support use of this suite of tools.

### **W326: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **Phylogenomic Analyses of the Thousand Plant Transcriptomes**

**Siavash Mirarab**, University of California, San Diego, La Jolla, CA and One Thousand Plant Transcriptomes Initiative

Green plants (Viridiplantae) include around 450,000–500,000 species of great diversity and have important roles in terrestrial and aquatic ecosystems. The One Thousand Plant Transcriptomes Initiative sequenced the vegetative transcriptomes of 1,124 species that span the diversity of plants in a broad sense (Archaeplastida), including green plants (Viridiplantae), glaucophytes (Glaucophyta) and red algae (Rhodophyta). Here, we discuss the robust phylogenomic framework used for examining the evolution of green plants. The framework relies on many recent techniques, including species tree methods that account for gene tree discordance. Most inferred species relationships were well supported across multiple species tree and supermatrix analyses, but discordance among plastid and nuclear gene trees at a few important nodes highlights the complexity of plant genome evolution, including polyploidy, periods of rapid speciation, and extinction. Incomplete sorting of ancestral variation, polyploidization and

massive expansions of gene families punctuate the evolutionary history of green plants. In addition to the results, we discuss the availability of the dataset through CyVerse.

### **W327: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **Hosting Data at CyVerse to Visualize and Analyze on the UCSC Genome Browser**

**Brian Lee**, Univ. Calif. Santa Cruz, Santa Cruz, CA

CyVerse allows hosting data to access and to view on the UCSC Genome Browser through a “Send to Genome Browser” link creation option. CyVerse’s support of byte-range requests makes it an optimal location for storing genomic data, as well as being the location to host data in assembly hub files to represent novel organisms for visualization and for data extraction through tools such as UCSC’s new Application Programming Interface (API) with JavaScript Object Notation (JSON) output. With data hosted at CyVerse, UCSC’s new API provides direct access to different data including annotations and sequences of the novel organisms represented by the user-generated assembly hub files. A complete list of how the API works can be seen on the help page: <http://genome.ucsc.edu/goldenPath/help/api.html>

### **W328: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **CyVerse/DNA Subway: Incorporating Both Fundamental Coding Skills and Modern RNA-Seq Tools into the Undergraduate Curriculum**

**Judy A. Brusslan**, California State University, Long Beach, Long Beach, CA

Incorporating large transcriptome data analysis into the undergraduate curriculum is essential for preparing undergraduates for the modern workforce or graduate-level programs. This task can be difficult when both teachers and students have minimal computational experience. CyVerse and DNA Subway provide tools that allow inexperienced students to complete an RNA-seq pipeline using simple command line steps followed by more sophisticated GUI steps. Students use Jupyter notebooks to trim and filter raw data using Trimmomatic. Students need to change input files and output folder names and modify trimming parameters, thereby learning the fundamentals of command line logic. In addition, they can see how their changed parameters affect the trimmed data output. Students then transition to the DNA Subway Green Line and used Kallisto and Sleuth for rapid pseudoalignment and statistical analysis to identify differentially expressed genes (DEGs). DNA Subway is flexible and teachers can design analyses based on their area of interest or if a wet-lab confirmation of DEGs is the next step. In addition, GO enrichment analysis for Biological Processes can be performed using available GUI tools. Jupyter notebooks provide explanations and allow teachers and their students to understand and use command line while the more sophisticated steps are completed on the robust Green Line of the DNA Subway. Teaching tools will be available to guide teachers as they bring these essential skills to a broad group of undergraduate students.

### **W329: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **Deploying Scalable, Interactive Bioinformatics Analyses *via* VICE**

**Peter W Rose**, UC San Diego, La Jolla, CA

We demonstrate how we deployed a prototypes of our mmtf projects for scalable data mining of the Protein Data Bank (PDB) on VICE (Visual Interactive Computing Environment), an extension to the CyVerse Discovery Environment (DE).

The Protein Data Bank (PDB) represents the core data resource for Structural Bioinformatics. The rapid growth of the PDB (> 150,000 structures) enables large-scale data mining, such as development of knowledge-based potentials, docking and scoring functions, and machine learning for protein structure and function prediction.

We have developed efficient data representations (MacroMolecular Transmission Format) and a scalable framework to mine the PDB using state-of-the-art Big Data Technologies (mmtf-pyspark). We have deployed applications of this framework in Jupyter Notebooks that are hosted on CyVerse.org, enabling researchers to publish documented workflows that are reproducible and that can be re-run, modified, or used as starting points for new structural analyses.

We demonstrate these capabilities by mapping and visualization of post-translational modifications from proteomics experiments and genomic variations to 3D structures in the context of protein-protein/nucleic acid/ligand/drug interactions. We also cover best practices of deploying these workflows to enable reproducibility and reuse.

### **W330: CyVerse - Software, Tools, and Services for Data-Driven Discovery and Learning**

#### **PhytoOracle: A Scalable, Modular Framework for Phenomics Data Processing and Trait Extraction**

**Sateesh Peri**, Genetics Graduate Interdisciplinary Group, University of Arizona, Tucson, AZ

As the capacity of phenomics to generate larger and higher dimensional data sets improves, there is an urgent need to develop and implement robust data processing pipelines to handle the data volume so that biological insight can be leveraged from them. Current phenomics data pipelines lack extractor modularity and distributed computing, leading to significant bottlenecks in data processing. To address these challenges, we have developed PhytoOracle, a modular, scalable data pipeline that aims to improve data processing for phenomics research. PhytoOracle refines the TERRA-REF data pipeline by integrating CCTools' Makeflow and Workqueue frameworks for distributed task management. Briefly, PhytoOracle distributes data processing tasks to either local, cloud or high-performance computing (HPC) systems. These systems include CyVerse, JetStream and other XSEDE resources, local/private HPC centers, and commercial cloud providers. Each tool and pipeline is available as containers providing portability as well as modularity, enabling researchers to swap between available extractors or integrate new ones suited to their specific research needs. The future scope and applications of phenomics will largely depend on the capabilities of data pipelines. PhytoOracle handles increasing rates of data collection while also enabling easy development, modification, and customization. As a result, researchers using this pipeline can quickly process data and extract phenotypic information, thereby enabling faster elucidation of genetic components of complex traits. Code, containers, and documentation is available: [github.com/uacic/PhytoOracle](https://github.com/uacic/PhytoOracle).

### **W331: Data Resource Sustainability and Funding**

#### **Sustainability Approaches for Biological Databases and Tools**

**Eva Huala**, Phoenix Bioinformatics, Fremont, CA and **Maureen A. O'Leary**, Stony Brook University, Stony Brook, NY

Phoenix Bioinformatics is a nonprofit founded with a mission of exploring and making available a range of sustainable funding options for databases that have lost grant funding. Our initial efforts were focused on support of TAIR and we found that a partially open subscription model could successfully support this resource. The concept of partially open encompasses ways to make data as widely available as possible while still preserving an incentive to subscribe. This model has been applied successfully to provide support for BioCyc and other Phoenix partners. More recently we have launched an effort to support our latest partner, MorphoBank, using a different sustainability model that provides unrestricted data access to all. The voluntary membership model is informed by the successful example of ArXiv. It relies on membership fees from institutions that obtain high value from the resource and support open access.

### **W332: Data Resource Sustainability and Funding**

#### **Global Core Data Resources and the Global Biodata Coalition**

**Charles E Cook**, Global Biodata Coalition, Hinxton, United Kingdom

Life science and biomedical research is crucially dependent on a global infrastructure of open access data resources that store primary experimental data and provide added-value to those primary data through curation and computational tools. This infrastructure is supported by public and charitable funders across the globe, but that support is not coordinated internationally and is often short-term, leading to some risk for the infrastructure.

The Global Biodata Coalition (GBC) is a forum created by and for biomedical and life sciences funders to aid those funders in better coordinating support for biodata resources and to ensure sustainable funding for the global infrastructure of biodata resources. A major activity of the GBC will be to identify a set of Global Core Data Resources (GCDRs) that are crucial for this infrastructure. These resources are fundamental to the entire biomedical and life science research effort globally and ensuring that these resources are sustainable and available for researchers will help stabilise the global biomedical and life science research enterprise.

A process for selecting the GCDRs is still under development, but will be modeled on the process used by ELIXIR to select European Core Data Resources (CDRs) that used a basket of qualitative and quantitative indicators to assess the scientific focus and impact, quality, governance, and importance of the resources for the infrastructure.

### **W333: Data Resource Sustainability and Funding**

#### **Flybase Community Financial Support**

**Susan Russo Gelbart** and Norbert Perrimon, Harvard University, Cambridge, MA

### **W334: Data Resource Sustainability and Funding**

#### **Working Toward Sustainability for the Genomics Education Partnership**

**Sarah C. R. Elgin**, Washington University in St. Louis, St. Louis, MO, Anna Allen, Howard University, Washington, DC, Andrew Arsham, Bemidji State University, Bemidji, MN, Anya Goodman, Cal Poly - San Luis Obispo, San Luis Obispo, CA, Catherine Reinke, Linfield College, McMinnville, OR, Anne Rosenwald, Georgetown University, Washington, DC, Douglas L. Chalker, Washington University in St. Louis, St. Louis, MO and Laura K. Reed, The University of Alabama, Tuscaloosa, AL

The Genomics Education Partnership is a nation-wide consortium of >100 faculty who work collaboratively to provide their undergraduate students with a research-based introduction to genomics, centered on genome annotation. The lead scientist, responsible for the overall goals of the research effort, gets valuable help; the faculty partners get support in teaching a genomics-based CURE; and the students get a research experience, learn how to use large databases, and grow their knowledge of genes and genomes. There are two major financial costs: support for the central IT infrastructure and support for new faculty training and an annual all-member workshop. Founded in 2006, the GEP has been supported continuously by grants from HHMI, NSF, and NIH. This enables us to utilize the administrative and physical infrastructure of our host institutions (University of Alabama and Washington University in St. Louis) to manage employee pay/benefits and host workshops. While new grant funding has been obtained recently, indefinite grant funding seems unlikely. Therefore, we are currently working to acquire non-profit (501c3) status. To try to manage the necessary decisions and steps in the most cost-effective manner, we have joined with other nation-wide genomics education projects to form the Genomics Education Alliance (GEA), currently funded by an NSF RCN-UBE grant. We are testing the proposition that GEA can provide IT support for many such projects. If this is successful, GEA might become a 501c3 umbrella organization for all partners. GEP is currently funded by NSF IUSE-1915544 and NIH IPERT-1R25GM130517-01 to LKR.

### **W335: Data Resource Sustainability and Funding**

## **Eukaryotic Pathogen, Vector & Host Informatics Resources (VEuPathDB) - a Merged International Approach Towards Achieving Sustainability**

**Jessica Kissinger**, University of Georgia, IOB, CTEGD & Genetics, Athens, GA and on Behalf of VEuPathDB

VEuPathDB.org is a family of free online resources that support Omics-data mining for eukaryotic pathogens, fungi, and invertebrate vectors of infectious diseases. The resources, which include 15 area-specific databases contain almost 200 organisms and nearly 500 genome sequences within the protists and vectors as well as numerous oomycetes, fungi and several vertebrate host species. These databases support a wide range of datatypes including genomic, population biology, microbiome and epidemiology. VEuPathDB facilitates the discovery of meaningful biological relationships or testing of hypotheses from large volumes of integrated pre-analyzed Omics data with advanced search capabilities, data visualization and analysis tools. The graphic interface allows users to mine the data without the need for computational training. Data types include genome sequence, annotation, transcriptomics, proteomics, epigenomics, metabolomics, population resequencing, clinical data, abundance data and host-pathogen interactions. Data are analyzed using bioinformatics workflows and in-house analyses including domain predictions and orthology profiles across all genomes which permit inferences from data-rich organisms to organisms with limited or missing data. Our resources offer several perspectives for data mining – record pages compile all data for genes, pathways, etc; a JBrowse genome browser for visualizing sequence data aligned to a reference genome; a search strategy system that queries pre-analyzed data and returns genes or features with shared biological characteristics; a geospatial visualization tool for finding sample; a private Galaxy workspace for analyses of user data and viewing in context with public data already integrated into VEuPathDB. Our active user support offers an email help desk, social media, video tutorials and a worldwide program of workshops. These free resources easily merge evidence from diverse data and across species to place the power of bioinformatics with every scientist. Recent expansion includes ClinEpiDB.org, a site which facilitates the exploration and analysis of epidemiologic studies and MicrobiomeDB.org.

Despite the different communities and data types served by the 15 area-specific databases, the underlying architecture is the same/re-used, providing significant economies of scale. Core funding for VEuPathDB components is provided by NIH contract (NIH HHS 75N93019C00077) while ClinEpiDB is supported by the Bill & Melinda Gates Foundation (OPP1151701). Curation for specific species and additional tool development are supported by funding from the Wellcome Trust via resource grants (212929/Z/18/Z & 218288/Z/19/Z). As our user base is global, our funding model is as well. This said, we also recognize that different funders and countries have different priorities and constraints. In an attempt to serve the largest number of users and keep pace with the data deluge produced by the communities we serve, we have taken on the onus of applying and reporting to multiple funding sources and we are truly grateful for their support. However, this approach does not scale as the administrative burden generated by asynchronous reporting in multiple formats to different agencies and tracking of distinct deliverables is high.

### **W336: Data Resource Sustainability and Funding**

#### **Sustainability for USDA's Ag Data Commons Catalog and Repository**

**M.J. Woodward-Greene**, USDA Agricultural Research Service, National Agricultural Library, Beltsville, MD, Cynthia Parr, USDA Agricultural Research Service, National Agricultural Library, BELTSVILLE, MD and Susan McCarthy, USDA Agricultural Research Service, National Agricultural Library

The National Agricultural Library (NAL) launched the Ag Data Commons in 2015 in preparation for supporting US government policies requiring public access to research data produced with federal funding. Developments in public policy, journal policies, and data science in the last five years have reinforced the need for a sustainable, well-coordinated program to support discovery, added value, and preservation of research data in agriculture. However, resources to implement the vision must be requested from key stakeholders and spent wisely. Thus, in the last year NAL engaged in a number of

activities to explore business use cases and models, as well as success metrics. This talk will describe our engagement with Phoenix Bioinformatics, which conducted market research and provided recommendations in support of a mixed funding model for the Ag Data Commons. We present a refined model of NAL's role in the research data landscape, including plans for deeper relationships with AgBioData and other disciplinary databases for long term data stewardship and crop data harmonization tools. Finally, we share results from an exercise by earth and environmental sciences data repositories to identify useful and implementable measures to track costs and benefits of repository operations, in order to justify sustained investment.

### **W337: Data Resource Sustainability and Funding**

#### **Database and Resource Sustainability from the NSF Viewpoint**

**Peter McCartney**, National Science Foundation, Alexandria, VA

Following the outcome of a 2007 Committee of Visitors review, the Division of Biological Infrastructure at NSF has sought to address the growing challenges of sustaining the availability of cyberinfrastructure resources upon which biological research has become increasingly dependent. What we've learned is that this involves significant cultural changes across the board: within the funding agencies, the resource providers, and the communities they serve. Actions we have taken to date include introduction of funding programs explicitly for supporting operations and maintenance of resources, training opportunities for resource providers to incorporate business planning into their sustainability goals, and changes within NSF's program structure to recognize sustainability as one component of our overall portfolio. We now have several active awards that are successfully drawing at least some of their revenue through the services they provide and they have become models for other resources to develop more direct fiscal relationships with their user communities.

### **W338: Decadal Vision for Plant Systems Science: Challenges and Breakthroughs**

#### **Overview of the Plant Science Research Network and the PSRN Decadal Vision**

**Rebecca Grumet**, Department of Horticulture, Michigan State University, East Lansing, MI

The Plant Science Research Network (PSRN) was established in 2015 to implement the Plant Science Decadal Vision goals and chart a vision for the future by integrating research (discovery), people (workforce and workplace) and technology (infrastructure improvement). To this end, the PSRN sponsored a series of workshops on cyberinfrastructure, postgraduate training, broadening participation, and envisioning future research environment scenarios. In 2019, PSRN hosted the Plant Summit III including 50 scientists representing diverse plant science communities working across scales from molecule to plant to field to ecosystem to ask: how can plant sciences help address the grand challenges facing agricultural and natural ecosystems from population growth, climate change, and escalating agricultural demands? This collection of early, mid and late-career researchers from academia and industry, reached across disciplinary niches to synthesize research, training and technology goals for next Decadal Vision, 2020-2030. Four overarching themes emerged: (1) Harness plants for planetary resilience through deep understanding of ecological diversity and evolutionary change; (2) Advance technology and diversity-driven sustainable plant production systems; (3) Develop 21<sup>st</sup> century applications of plant science to improve nutrition, health and well-being; and (4) Develop a 'transparent plant' tool to discern mechanistic insights and solve vexing and urgent problems. These goals, which will require new technologies, hardware, and ability to process Big Data, must be approached in the context of workplaces that nurture and support diverse scientists, and elicitation of broader engagement by the public with plant sciences.

### **W339: Decadal Vision for Plant Systems Science: Challenges and Breakthroughs**

#### **Reimagining the Workplace to Nurture and Support Adaptive and Diverse Scientists**

**Katie Rogers**, American Society of Plant Biologists

People provide the foundation and motivation for plant science research and its applications. A diverse workforce that reflects our society is needed and will be achieved by developing research environments that are inclusive, equitable, and geared toward incentivizing, supporting and rewarding collaborative research.

Collaborative achievements increasingly deliver the far-reaching scientific insights society requires. To encourage collaboration, a resolution is needed to resolve the contradiction in academic metrics, wherein individual accomplishments are the primary drivers of faculty hiring and evaluation.

Recognizing that a minority of postgraduate trainees desire or obtain careers as academia, training and mentoring should focus on providing support and preparation for a variety of career choices. Customizable, flexible and modular training possibilities will provide an array of transferable research and cultural skills.

In the future, technology will continue to facilitate seamless, global collaboration and workforce diversity by delivering accessibility to specialized knowledge and equipment, for both research and training purposes. Virtual workplaces and meetings will help decrease the carbon footprints of travel and strengthen the global plant science community.

### **W340: Decadal Vision for Plant Systems Science: Challenges and Breakthroughs**

#### **Harnessing Plant Systems for Planetary Resilience**

**Pamela S. Soltis**, University of Florida, Gainesville, FL

A rapidly changing planet is generating challenges for plant life worldwide. Rising temperatures, altered precipitation patterns and fire regimes, and unpredictable weather are affecting natural ecosystems and agricultural lands alike. To preserve natural areas, maintain ecosystem services, and feed 9 billion people in 2050, new approaches to conservation and crop breeding are needed. The challenges of the 21<sup>st</sup> century require increased focus on biodiversity science and greater understanding of ecology and evolution. Emerging cyberinfrastructure and new data sources provide unparalleled opportunities for mobilizing and integrating massive amounts of information from organismal biology, ecology, genetics, genomics, climatology, and other disciplines. Key among these data sources is the rapidly growing volume of digitized specimen records from natural history collections. With over 120 million specimen records currently available online, these data provide excellent information on species distributions, changes in distributions over time, phenology, and a host of traits. Case studies that link and analyze specimen data and related heterogeneous data sources to address a range of evolutionary and ecological problems will explore the specific challenges encountered and how these challenges may be overcome. Linking ecological, evolutionary, and genomic data will provide keys to engineering resilience in a changing planet.

### **W341: Decadal Vision for Plant Systems Science: Challenges and Breakthroughs**

#### **The Role of Science in Future Farming**

**Gerardus W. van der Heijden**, Corteva Agriscience, Johnston, IA

Farmers generally know which crops grow well on their fields and how they need to manage their crop. But a lot is still unknown: what is the exact influence of the previous crop or cover crop on the current crop, how do soil properties and soil preparation as well as weather influence the growth of the crop this year? What is the optimal amount of fertilizer for my local conditions? How much will disease and insect pressure affect my yield this year? How can I be as profitable and sustainable as possible?

Science can help the farmer with some of these questions and uncertainties. A lot of the research can and must be done in the lab or in the field under well-controlled conditions with advanced sensitive equipment, but it is also fascinating to think about how the new ways in which we collect, exchange and

store data can help us to do our research to better understand the farming system and provide recommendations to the farmer.

More and more data become available: machine data regarding planting, spraying and harvest are stored in the cloud combined with sensor data from advanced and novel IoT devices. Daily satellite and weather data are available at a high spatial resolution. Molecular data are collected on crops and pathogens and so on. If we have the right amount of (meta)-data we can consider every farmer's field as an experimental unit for data-analysis. And we can draw inferences from the data. This is not going to be easy, as there are many variables, with small effects and complex unknown interactions. If crucial information is missing in the data, we will mostly observe noise. Attention to detail, knowledge of which data are crucial, data curation, creative minds and complex learning algorithms are just some of the aspects needed to truly learn and discover from the wealth of data that we are collecting.

Eventually we should be able to better understand the farming system from that data. We can then build better models to predict the future of the crop and ultimately help the farmer to improve the profitability and sustainability of the farm.

### **W342: Decadal Vision for Plant Systems Science: Challenges and Breakthroughs**

#### **Towards the Transparent Plant: Contemplating Interdisciplinary, Comprehensive, Cross-Lab Development of Resources for Understanding Plant Environmental Resilience**

Sally A Mackenzie, Pennsylvania State University, University Park, PA and **Brett Tyler**, Oregon State University, Corvallis, OR

Numerous governmental, National Academy of Science, and scientific community reports have attempted to draw public attention to the critical need for focused research that addresses the growing threat of climate instability to agricultural and ecological systems worldwide. Yet, to date, a collective response in the form of actionable goals has not been formulated by the plant research community. A 2019 meeting of the Plant Science Research Network was held to begin this process, and several approaches to a coordinated, international plant research effort were debated. One of four research priorities identified was dubbed "The Transparent Plant" which referred to developing and modeling a detailed, integrated understanding of plant evolutionary, structural, and metabolic processes essential to engineering plants of the future for productive and resilient growth. Here, we present a crucial application of the transparent plant strategy, directed towards modeling at the plant:water interface. Water is a vital and increasingly scarce component of plant growth, and a detailed understanding of plant processes that are impacted by water will be essential to engineering plants of the future for productive growth under limited, more saline, or recurring flood to drought scenarios for water. This detailed knowledge will require high resolution, spatiotemporal gene network integration under various environmental scenarios and across a core collection of plant species. High-resolution dissection of protein-protein interaction and chromatin behavior, evolutionary decision-making, intracellular machinations and genotype-to-phenotype processes must integrate into a standardized, quality-controlled public resource designed to expedite plant resilience research. Finding agreement on design and priorities within the plant research community, from fundamental to agricultural to ecological, resourcing this grand challenge, and convincing policymakers and industry partners of its vital importance are some of the issues we now face.

### **W343: Decadal Vision for Plant Systems Science: Challenges and Breakthroughs**

#### **Panel Discussion**

**Sreekala Chellamma**, Corteva Agriscience™, Johnston, IA and **Brett Tyler**, Oregon State University, Corvallis, OR

The speakers and the chairs will participate in a discussion with the audience on the plant decadal vision, covering its seven goals.

- 1: Harness plant systems for planetary resilience through deep understanding of ecological diversity and evolutionary change
- 2: Advance technology and diversity-driven sustainable plant production systems
- 3: Develop 21st-century applications of plant systems science to improve nutrition, health and well-being
- 4: Develop a “Transparent Plant Systems” platform to discern mechanistic insights and solve both vexing and urgent problems.
- 5: Reimagine the workplace to nurture and support adaptive and diverse scientists
- 6: Building pervasive “plant awareness” to enhance engagement with plant science
- 7: Build a participatory community of learners and citizen scientists

### **W344: Degraded DNA and Paleogenomics**

#### **Spatiotemporal Genetic Diversity of Lions**

**Caitlin Curry**<sup>1</sup>, Brian W. Davis<sup>1</sup>, Laura Bertola<sup>2</sup>, Paula A. White<sup>3</sup>, William J. Murphy<sup>1</sup> and James N. Derr<sup>1</sup>, (1)Texas A&M University, College Station, TX, (2)City College of New York, (3)Zambia Lion Project, University of California, Los Angeles, Los Angeles, CA

We determined the genetic architecture of both historical and modern lions to identify changes in genetic diversity that occurred over 100 years of landscape and anthropogenic change. We surveyed microsatellite and mitochondrial genetic variation from 143 high-quality museum specimens of known provenance and combined them with data from recently published nuclear and mitochondrial studies. Analysis of variation at 9 microsatellites and 280 polymorphic mitogenome SNPs indicate the presence of male-mediated gene flow and recent isolation of local subpopulations, likely due to habitat fragmentation. Nuclear markers showed a significant decrease in genetic diversity from the historical ( $HE=0.833$ ) to the modern ( $HE=0.796$ ) populations, while mitochondrial genetic diversity was maintained ( $Hd=0.98$  for both). While the historical population appears to have been panmictic based on nDNA data, hierarchical structure analysis identified four tiers of fine structure in modern populations, able to detect most sampling locations. Mitochondrial analyses identified 4 clusters: Southern, Mixed, Eastern, and Western; and were consistent between modern and historically sampled haplotypes. Within the last century, habitat fragmentation caused lion subpopulations to become more isolated as human expansion changed the African landscape. This resulted in an increase in fine-scale nuclear genetic structure and loss of genetic diversity as subpopulations became more differentiated, while mitochondrial structure and diversity was maintained over time.

### **W345: Degraded DNA and Paleogenomics**

#### **Ghost Genomics: The Persistence of Red Wolf Ancestry in Southeastern Canids**

**Bridgett vonHoldt**, Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ

The red wolf (*Canis rufus*), a legally recognized and severely endangered wolf, is known to interbreed with coyotes (*C. latrans*). Declared extirpated in the wild in 1980, red wolves were reintroduced to North Carolina nearly a decade later. Between 1973 and 1977 (“pre-recovery”), up to 400 canids were trapped from the American Gulf Coast regions of Louisiana and Texas, and evaluated for canonical red wolf features. Of these, only 17 individuals were retained as red wolf founders for the Species Survival Plan’s captive breeding program. Following nearly a decade of captive breeding, red wolves were reintroduced to the Albemarle Peninsula in northeastern North Carolina in 1987. Interbreeding with coyotes was largely believed to threaten red wolf recovery. However, red wolf ancestry has recently discovered in

canids along the American Gulf Coast, igniting a broader survey of endangered ancestry in southeastern canid populations. Here, we examine geographic and temporal patterns of genome-wide red wolf ancestry in 260 canids across the southeastern United States at over 164,000 SNP loci. We consistently identified two distinct geographic occurrences of red wolf ancestry in southeastern coyotes, concordant with previous reports of historical and on-going hybridization between coyotes and red wolves across Texas and Louisiana, and in North Carolina, respectively. Assignments to the red wolf cluster were highest in pre-recovery canids (1-51%), followed by canids in the Gulf Coast and North Carolina coyotes (2- 45%). We observed similar trends in the private allele sharing analysis, as reference red wolves shared the most private alleles with pre-recovery Texas canids. Alleles private to historic and contemporary Gulf Coast canids may represent variation that was unique to the historical red wolf population, which were lost in the bottlenecked of establishing their captive breeding colony (i.e., ghost alleles). Such populations that carry admixtures of red wolf and coyote ancestries likely represent an important reservoir of red wolf genetic diversity that has been lost in the captive breeding and wild populations due to demographic contraction over the last century. Furthermore, it is important to recognize that estimates of red wolf ancestry are based entirely on the genetic diversity that is found in the captive population; thus, estimates of red wolf ancestry in wild canids are likely to be underestimated as the captive population is not expected to represent the full diversity of wild red wolves prior to extirpation.

### **W346: Degraded DNA and Paleogenomics**

#### **The Identification of the First Storegga Slide Tsunami Deposits in the Southern North Sea using a Multi-Proxy Approach**

**Roselyn Ware**, University of Warwick, Coventry, United Kingdom

Doggerland was a fertile land-mass in what is now the southern North Sea, prior to its submersion and subsequent excision of Britain from Europe, and retains valuable evidence about local Mesolithic communities and the local environment. The inundation of Doggerland is considered a consequence of sea level rises in the mid-Holocene, with the catastrophic tsunami triggered by the Storegga landslides speculated to have played a pivotal role. There is extensive evidence for this tsunami observed across the northern areas of the North Sea, however, no evidence has been found in the southern North Sea, despite this area having been predicted to within the expected range of the impact of the tsunami.

Sediment cores were taken from across Doggerland, guided by extensive seismic mapping. Evidence based on established methods of traditional paleoenvironmental analysis, geochemical analysis, coupled with sedaDNA analysis, has enabled the first identification of tsunami deposits. This is supported by dating (radiocarbon and OSL) which places these deposits as contemporary to the Storegga slide. The dynamics of palaeoenvironmental change were reconstructed using metagenomic sequencing of ancient sediment DNA (sedaDNA). We identified sweeping shifts in plant communities, with an influx of incongruous plant taxa within the deposit, consistent with the catastrophic effects of a tsunami.

As well as the identification of the first example of Storegga tsunami deposits in the southern North Sea, integration of these diverse datasets has enabled us to identify further tsunami deposits, assess the environmental impacts, and establish how the morphology of the landscape impacted the tsunami's progression. We will also discuss the toolkit we are developing for the analysis of sedaDNA.

### **W347: Degraded DNA and Paleogenomics**

#### **Hybridization Capture of Ancient DNA Reveals Harvest of Winter-Spawning Herring Populations by Coast Salish Fisheries over 900 Years**

**Eleni L. Petrou**<sup>1</sup>, Camilla Speller<sup>2</sup>, Robert Kopperl<sup>3</sup>, Dana Lepofsky<sup>4</sup>, Madonna L. Moss<sup>5</sup>, Antonia T. Rodrigues<sup>4</sup>, Dongya Yang<sup>4</sup> and Lorenz Hauser<sup>1</sup>, (1)School of Aquatic and Fishery Sciences, University of Washington, (2)Department of Anthropology, University of British Columbia, Vancouver, BC,

Canada, (3)Willamette Cultural Resources Associates Ltd, (4)Department of Archaeology, Simon Fraser University, Canada, (5)Department of Anthropology, University of Oregon

The extent to which different populations or phenotypes contribute to ecosystem goods over long time scales is mostly unknown. Temporal studies of population diversity are particularly important in forage fish such as Pacific herring, as they are foundational to coastal food webs, cultures, and economies. In this study, we investigated the relative contributions of genetically distinct herring populations to food supplies over the last millennium, using ancient DNA extracted from herring bones ( $N = 44$ ) excavated from the Burton Acres archaeological site in Puget Sound, Washington. We applied hybridization capture techniques to genotype approximately 5,000 SNPs in ancient samples and identify herring populations harvested by Coast Salish fishers over a period of approximately 900 years. Our genetic data reveal that recent catches at Burton Acres were dominated by winter-spawning herring, a pattern which matches current spawning distributions of these fish. However, late spring-spawning herring were detected in the older Burton Acres assemblage (915-680 ybp), and a mixed stock analysis also indicated that catches at this temporal layer were more diverse and consisted of mixed populations. Our results suggest that people at the Burton Acres site used a portfolio of herring populations and benefited from the ecological resource wave created by different spawning phenotypes.

### **W348: Degraded DNA and Paleogenomics**

#### **Tracking Early Horse Domestication Stages in the 3rd and 4th Millennium BCE**

**Ludovic Orlando**<sup>1</sup>, Antoine Fages<sup>2</sup>, Naveed Khan<sup>3</sup>, Pablo Librado<sup>3</sup>, Alan K Outram<sup>4</sup> and The ERC PEGASUS consortium<sup>5</sup>, (1)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Denmark, Copenhagen, Denmark, (2)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Toulouse, France, (3)CNRS, University Paul Sabatier, Toulouse, France, Toulouse, France, (4)University of Exeter, Exeter, United Kingdom, (5)Laboratoire d'Anthropologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse III Paul Sabatier, Toulouse, France

Today, horses remain essential to the agricultural production of developing countries. In the rest of the world, they are, however, often limited to the sport industry or for providing animal lovers with some of their best companions. Yet, not so long ago, horses were essential to human societies. By providing us with speed, they revolutionized not only the way we traveled, but also the way our genes, diseases, goods and languages circulated across the planet, effectively globalizing the world for the first time. They also fundamentally changed the way we made war and represented key military assets for past civilizations. Yet, the process of early horse domestication is difficult to reconstruct based on archaeological data alone. The horse is indeed generally not-so-common in the Neolithic assemblages of the 4th and 3rd mill BCE and no massive size shifts have accompanied the transformation of wild animals into domesticates, in contrast to other herbivore species. In the last few years, we have undertaken to leverage both standard archaeological approaches and ancient DNA data to clarify the temporal and geographic loci underlying horse domestication. The genetic data have been essential to overcome the limitations of a too often fragmentary record, and represent the largest genome time-series for a non-human organism generated thus far. They reveal strong patterns of geographic differentiation amongst wild horses and across Eurasia at the time of domestication and a striking genomic turnover coincidental with massive human migration.

### **W349: Degraded DNA and Paleogenomics**

#### **The Ever Changing Genepool of Scandinavia**

**Anders Gotherstrom**, Stockholm University, Stockholm, Sweden

Over the recent years we have been generating ancient human genomic data from Scandinavia for specific archaeogenetic studies. Consequently, there is data from various chronologies from this corner

of the world. This provides for the possibility to do a genetic transect of the region. While most of the demographic events mirrors a larger and by now well-known demographic development, there are some interesting specific cases. I will highlight a few of these. There are individuals from the Mesolithic as well as from the Neolithic periods, and also from the Viking period that do not easily fit in with the genetic and isotopic patterns, and they provide personal biographies. And the sailors from the royal Man of war Kronan, that capsized in the Baltic sea 1676 sheds some light on the latter process of establishing the gene pool that is present in Scandinavia today. Only with the depth of the Scandinavian dataset, it becomes possible to single out specific cases and processes like these.

### **W350: Degraded DNA and Paleogenomics**

#### **Genetic History of Northern and Southern East Asians from 9,500-3,900 Years Ago**

**Melinda Yang**, University of Richmond, Richmond, VA and Qiaomei Fu, Molecular Paleontology Lab, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, China

Study of the genetic history of humans in East Asia in the early to mid-Holocene have typically relied on present-day populations, or ancient individuals from peripheral regions such as the Tibetan Plateau, Siberia, or Southeast Asia. In China, the early Holocene coincides with a shift to an agriculturally-based economy, with archaeological evidence for increasing population size and site complexity. Morphological evidence have been used to suggest two 'layers' of ancestry in East Asia - first of Paleolithic hunter-gatherers, including those associated with Hòabínhian culture in Southeast Asia and Jomon culture in Japan, replaced by populations more closely associated with present-day northern East Asians with the expansion of rice and millet agriculture. We sampled ancient DNA from the nuclear genome from 26 individuals from northern and southern East Asia (Inner Mongolia, Shandong, Fujian, Taiwan) dating to ~9,500-3,900 years ago. We found that they share a close relationship to each other and present-day East Asians, where ~9,000-7,500-year-old individuals from southern China do not possess ancestry related to previously sampled Paleolithic hunter-gatherers. However, we did observe higher genetic differentiation 9,500-3,900 years ago between northern and southern East Asians relative to populations within the same regions today, related to an increase in ancestry associated with ancient northern East Asians. The results indicate gene flow southward from northern East Asia played a prominent role in the composition of today's East Asians.

### **W351: Degraded DNA and Paleogenomics**

#### **An Ancient Genomic Perspective of the Tiger Evolution**

**Shu-Jin Luo**, Peking University, Beijing, China

Genome-wide evolutionary analysis has affirmed intraspecific distinction that supported six living subspecies in the tiger, whose free-ranging population has dwindled from over 100,000 individuals in the 1900s to fewer than 4,000 to date. The oldest tiger fossil was dated at 2-3 Mya, however TMRCA of modern tigers was traced to only 110 kya, suggesting a Late Pleistocene bottleneck and complex demographic dynamics, which can only be unraveled through analysis of the extinct tigers. Here we retrieved the first ancient tiger genome, at an 8.3x genome coverage, of a ~10,000-year-old bone (radiocarbon date of 10,582–10,399 cal BP with  $\pm 2\sigma$  at 95.4% probability) uncovered from the Russian Far East (RFE), as well as whole genome information from centuries-old South China tigers (*P. t. amoyensis*, N=12) and Caspian tigers (*P. t. virgata*, N=3) to represent historical tiger populations across mainland Asian. In conjunction with published genome sequences from 32 extant voucher tiger specimens, genome-wide phylogeny supported *P. t. amoyensis* as a statistically robust clade relative to other subspecies, albeit its mitochondrial paraphyly, resolving the long-lasting taxonomic controversy in the South China tiger. The ancient RFE tiger carried a basal mitochondrial lineage distinct from modern Amur tigers (*P. t. altaica*), however clustered in the autosomal tree within the northeast Asian phylogroups including *P. t. amoyensis* and *P. t. altaica*. At last, *P. t. virgata* of central Asia originated from tigers in the RFE expanding westbound via Siberia and subsequently had genetic introgression from Bengal tigers to the south. Ancient tiger genomes illuminated that mainland China served as a Late Pleistocene refugia for relic tiger lineages and multiple range contraction-expansion-isolation cycles

during the Last Glacial Period have led to phylogeographic partitioning and distinction of living tiger subspecies.

### **W352: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**

#### **The Long Road to the Golden Banana**

**James Dale**<sup>1</sup>, Jean-Yves Paul<sup>2</sup>, Stephen Buah<sup>3</sup>, Jimmy Tindamanyire<sup>4</sup>, Amba Lawrence<sup>2</sup>, Priver Namanya<sup>3</sup>, Jerome Kubiriba<sup>5</sup>, Robert Harding<sup>6</sup> and Wilberforce Tushemereirwe<sup>7</sup>, (1)Centre for Tropical Crops and Biocommodities, Brisbane, QLD, Australia, (2)Queensland University of Technology, Brisbane, QLD, Australia, (3)National Agricultural Research Organisation, Kampala, Uganda, (4)NARL, Kawanda, Uganda, (5)National Agricultural Research Organisation, Kampala, Uganda, (6)Queensland University of Technology, Brisbane, Australia, (7)National Agriculture Research Laboratories, Kampala, Uganda

Vitamin A deficiency (VAD) continues as one of the major public health challenges in the developing world. Biofortification provides the opportunity to alleviate VAD through elevation of pro-vitamin A (PVA) levels in staple crops with Golden Rice as the prime example. In Uganda, where there are significant levels of VAD particularly in rural areas, more than 70% of the population depend on a cooking banana, East African Highland bananas as their primary starch source. In 2005, we commenced a project to develop enhanced levels of PVA in this staple through genetic modification. We inserted a phytoene synthase gene from a Fe'i banana that has naturally very high levels of PVA. In the elite line selection confined field trials at one site in Uganda, we have selected 12 lines that have been progressed through to the regulatory national performance trials at four different agro-ecological zones in Uganda. The 12 lines are comprised of 6 lines of Nakitembe, an East African Highland banana, and 6 lines of Hybrid M9, a conventionally bred cooking banana with specific disease resistances. The criteria for the selection of these elite lines was firstly a level of PVA greater than the target of 20mg/g dry weight equivalents of  $\beta$ -carotene as well as minimal or no yield penalty. This followed selection on the basis of transgene copy number, absence of plasmid backbone and an integration site that did not interrupt ORFs or create new functional ORFs. Our plan is to release these bananas in 2022.

### **W353: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**

#### **Genome Editing in Grass Plants**

**Bing Yang**, University of Missouri, Columbia, MO

Programmable nucleases (e.g., CRISPR RNA guided Cas nucleases) have been successfully engineered to induce site-specific mutations at genomic loci in grass plants such as rice, maize, wheat, sorghum, etc. The genome editing tools have significantly advanced our basic understanding of gene function and engineering beneficial traits in grass plants. In my presentation, I will provide our experience in developing and utilizing CRISPR/Cas9 technologies for targeted mutagenesis in several important grass crops.

### **W354: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**

#### **Nanomaterials Enable Delivery of Genetic Material without Transgene Integration in Mature Plants**

**Markita P. Landry**, University of California Berkeley, Berkeley, CA

Genetic engineering of plants is at the core of sustainability efforts, natural product synthesis, and agricultural crop engineering. The plant cell wall is a barrier that limits the ease and throughput with which exogenous biomolecules can be delivered to plants. Current delivery methods either suffer from host range limitations, low transformation efficiencies, tissue regenerability, tissue damage, or

unavoidable DNA integration into the host genome. Here, we demonstrate efficient diffusion-based biomolecule delivery into tissues and organs of intact plants of several species with a suite of pristine and chemically-functionalized high aspect ratio nanomaterials [1]. Efficient DNA delivery and strong protein expression without transgene integration is accomplished in mature *Nicotiana benthamiana*, *Eruca sativa* (arugula), *Triticum aestivum* (wheat) and *Gossypium hirsutum* (cotton) leaves and arugula protoplasts [2]. We demonstrate that our platform can be applied for CRISPR-based genome editing for transient expression of Cas9 and gRNAs. We also demonstrate a second nanoparticle-based strategy in which small interfering RNA (siRNA) is delivered to mature *Nicotiana benthamiana* leaves and effectively silence a gene with 95% efficiency. We find that nanomaterials both facilitate biomolecule transport into plant cells, while also protecting polynucleotides such as RNA from nuclease degradation. DNA origami and nanostructures further enable siRNA delivery to plants at programmable nanostructure loci [3]. Our work provides a tool for species-independent, targeted, and passive delivery of genetic material, without transgene integration, into plant cells for diverse plant biotechnology applications.

1. Demirer, G.S., Zhang, H., Goh, N.S., Grandio, E.G., Landry, M.P. Carbon nanotube-mediated DNA delivery without transgene integration in intact plants. *Nature Protocols* (2019). DOI: 0.1038/s41596-019-0208-9
2. Demirer, G.S., Zhang, H., Matos, J., Goh, N., Cunningham, F.J., Sung, Y., Chang, R., Aditham, A.J., Chio, L., Cho, M.J., Staskawicz, B., Landry, M.P. High Aspect Ratio Nanomaterials Enable Delivery of Functional Genetic Material Without DNA Integration in Mature Plants. *Nature Nanotechnology* (2019). DOI: 10.1038/s41565-019-0382-5NNANO-18081684
3. Zhang, H.\*, Demirer, G.S.\*, Zhang, H., Ye, T., Goh, N.S., Aditham, A.J., Cunningham, F.J., Fan, C., Landry, M.P. Low-dimensional DNA Nanostructures Coordinate Gene Silencing in Mature Plants. *PNAS* (2019). DOI: 10.1073/pnas.1818290116

## **W355: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**

### **Genome Editing of Wheat Grain-Regulatory Genes**

**Wanlong Li**, South Dakota State University, Brookings, SD

## **W356: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**

### **GAENTRY, a Novel *Agrobacterium*-Based Transgene Stacking System for Improved Crop Biotechnology**

**Roger Thilmony**, Leyla Hathwaik, Ray Collier and James Thomson, USDA-Agriculture Research Service, Albany, CA

Genetic engineering is an important tool for the rapid genetic improvement of crops and will enable future improvements of complex traits like yield and nutritional quality through the introduction and coordinated expression of multiple genes. GAENTRY (Gene Assembly in *Agrobacterium* by Nucleic acid Transfer using Recombinase technology) is a flexible and effective system for stably stacking multiple genes within an *Agrobacterium* virulence plasmid Transfer-DNA (T-DNA). The system utilizes unidirectional site-specific recombinases *in vivo* and an alternating antibiotic selection scheme to sequentially assemble multiple genes into a single, stable transformation construct. To demonstrate GAENTRY's capabilities, we have assembled large T-DNAs carrying 10 or more cargo sequences, >28 kilobases in size, and have used these constructs to generate transgenic rice and *Arabidopsis* plants. Frequently, the GAENTRY-generated transgenic plants were high quality events that carried the entire T-DNA, expressed all the functional transgenes and were free of unintended sequence from outside the T-DNA. Our research results demonstrate that GAENTRY is a powerful, yet simple to use, new tool for transgene stacking and crop biotechnology.

## **W357: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**

### **Expanding Multiplex Genome Editing Tools for Accelerated Breeding in Industrial Crops**

**Diego Orzaez**, IBMCP-CSIC Q2818002D, Valencia, Spain

Plant breeding is one of the trump cards at human reach to respond to climate change threats with sustainable growth. The intense research effort that followed the discovery of CRISPR/Cas9 and its exciting envisioned applications has led, as in a self-fulfilling prophecy, to the development of new CRISPR-related tools with expanded applications in plant breeding. This has been possible thanks to the striking ability of CRISPR/Cas ribonucleoproteins to accept the attachment of new modules encompassing additional functions, either attached to the protein itself or to the gRNA structure using an extended scaffold of RNA aptamers. In this talk we will review the expanded applications of CRISPR/Cas proteins in plant breeding, from multiplex mutagenesis to programable transcriptional regulation, and will present a new web-based genome engineering toolbox (GoldenBraid4.0), which facilitates the modular assembly of most CRISPR/Cas extended constructs. Furthermore, we will present examples of functional validation of GoldenBraid4.0 tools in fast breeding of *Nicotiana tabacum* and *Nicotiana benthamiana* as industrial crops for molecular farming uses.

## **W358: Domestication Genomics**

### **Exploitation of Semi-Domesticated Tomato Accessions for Crop Improvement Traits**

**Esther van der Knaap**<sup>1</sup>, Lei Zhang<sup>2</sup>, Lara Pereira<sup>2</sup>, Manoj Sapkota<sup>3</sup>, Hamid Razifard<sup>4</sup> and Ana Caicedo<sup>5</sup>, (1)Department of Horticulture, Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, (2)Center for Applied Genetic Technologies, Athens, GA, (3)Institute of Plant Breeding, Genetics and Genomics, (4)Biology Department, University of Massachusetts, Amherst MA, (5)Biology Department, University of Massachusetts, Amherst, MA

The trajectory of fruit and vegetable domestication is rarely thoroughly understood. Moreover, the selection of traits that were associated with their domestication may have differed over evolutionary time. Tomato was initially domesticated from wild relatives in Ecuador and Peru<sup>1</sup>. After that, semi-domesticated accessions traversed north to Central America and southern North America for further domestication into the varieties we know today<sup>1</sup>. Diversity studies of the subgroups show large genetic variation in certain semi-domesticated accessions, particularly from Ecuador and Peru. The genetic variation in these semi-domesticates is similar to wild accessions, contrary to highly domesticated and elite germplasm. Even though the fruits of these genetically diverse semi-domesticates are relatively small and sometimes unpalatable, we developed mapping populations to determine the inheritance of fruit quality traits that could be used for crop improvement. We discovered novel loci for fruit weight that led to the transition from one phylogenetic subclade to another. Moreover, we found evidence that some beneficial alleles for fruit quality may have been left behind during domestication.

Funded by NSF IOS 1564366.

1. Razifard, Ramos, Della Valle, Goetz, Manser, Li, Zhang, Visa, Tieman, van der Knaap, Caicedo. Population genomic evidence for complex domestication history of the cultivated tomato in Latin America. *In Press* Molecular Biology and Evolution.

## **W359: Domestication Genomics**

### **Genomic Diversity of Saamaka Rice Varieties in Suriname**

**Jacob B Landis**<sup>1,2</sup>, Rosemary Glos<sup>2</sup>, Margaret Royall<sup>2</sup>, Yuna Seo<sup>2</sup>, Sandra Harrington<sup>2</sup>, Iriel E Edwards<sup>2</sup>, Chelsea D Specht<sup>3</sup>, Susan McCouch<sup>3</sup> and Erika Styger<sup>2</sup>, (1)University of California Riverside, Riverside, CA, (2)Cornell University, (3)Cornell University, Ithaca, NY

Rice is a staple crop among the Saamaka Maroons of Suriname, a culturally, politically, and economically independent people from the upper Suriname River. Rice cultivation and consumption are intimately linked to Saamaka cultural identity and oral history dating back to the escape from slavery in the 17<sup>th</sup> century. Traditional rice cultivation is under threat from a variety of environmental and socio-economic factors, including land degradation, increasing consumption of commodity rice, and loss of rice-growing knowledge. Previous reports have suggested that the Saamaka grow both Asian rice (*Oryza sativa*) and African rice (*O. glaberrima*), but few Saamaka rice varieties have been studied at the genetic level. We collected samples of rice grown by Saamaka farmers in 17 field sites along the Suriname River, and were given permission to bring seeds back to Cornell University through a formal agreement with the Saamaka Paramount chief and Council. We grew seed from and resequenced a total of 432 lines to investigate the extent of genomic diversity and to address four initial questions: 1) Are Saamaka varieties maintained as pure lines or as heterogeneous populations? 2) Do variety names identify the same genetic entities in different farmers' fields? 3) Is there any evidence of hybridization or admixture resulting from intentional cross pollination? 4) What is the interplay between geographic distribution of field sites and genetic diversity of Saamaka rice varieties? The dataset generated as part of this collaborative project provides the most comprehensive view of rice diversity managed by Maroon farmers anywhere in the Guianas.

### **W360: Domestication Genomics**

#### **Assessing the Impact of Inbreeding on Demographic Estimates of Domestication**

**Paul D Blischak**, Michael S. Barker and Ryan N Gutenkunst, University of Arizona, Tucson, AZ

The evolutionary history of domesticated species reflects the common process of human-mediated selection for traits of interest from a wild ancestor, resulting in a reduction in genetic diversity. Inferring the severity, duration, and potential recovery from this bottleneck is a crucial step toward gaining a better understanding of the impact of domestication and can also serve as a null model to guide tests for selection while accounting for demography. Estimating demography from genomic data is a common practice in studies of domestication; however, most methods fail to account for the fact that modern breeding stocks are often inbred, which may lead to incorrect estimates of historical events. In this study, we develop a model for co-estimating demography and inbreeding and use simulations to test how inbreeding affects estimates of demography in the canonical bottleneck scenario. We then estimate demography in cabbage (*Brassica oleracea* var. *capitata*) using a model that includes a bottleneck followed by exponential population size change. In our simulations, we find that ignoring inbreeding generally leads to underestimation of both population size and duration of the bottleneck. For cabbage, inbreeding is estimated to be  $F_{IS}=0.58$ , and the model fit without inbreeding estimates a much larger reduction in population size followed by further population contraction. Given these results, there is evidence to suggest that estimates of severe bottlenecks in domesticated species are driven in part by inbreeding, which has important consequences for downstream analyses aimed at identifying domestication-related genes and/or genes under selection.

### **W361: Domestication Genomics**

#### **Genetics of Breed Formation and Associated Traits in the Domestic Dog**

**Elaine A. Ostrander**<sup>1</sup>, Heidi G. Parker<sup>1</sup>, Jaemin Kim<sup>1</sup>, Jocelyn Plassais<sup>1</sup>, Andrew Hogan<sup>1</sup>, Dayna L. Dreger<sup>2</sup>, Alex Harris<sup>1</sup> and Muhammad Basil Ali<sup>1</sup>, (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)Purdue University College of Veterinary Medicine, West Lafayette, IN

More than 400 dog breeds exist in the world and we have now sampled more than half of these. While most breeds were developed within the last 200 years, excessive haplotype sharing between breeds and reassignment of breeds to different clades upon addition of new breeds to our dataset suggests that multiple ancestral dog populations were used to construct many of the modern domestic breeds observed today. To recreate the ancestry of early dogs leading to modern domestic dogs we developed a dataset of 224 breeds from Europe, Asia, Africa, Australia, North and South America. Samples were genotyped at 150K SNPs using the Illumina CanineHD bead chip and analyzed using distance matrices

and neighbor-joining methods in Phylip, IBD haplotype sharing with the program Beagle, with statistical calculations performed in R. Genome-wide haplotype analysis done with common polymorphisms reveals the most likely modern versions of original ancestor dogs used to create modern breeds. We find, further, that the inclusion of regional breeds reveals previously hidden relationships between diverse dog clades, enabling the identification of “proto-breeds” that represent the initial breed types that preceded the domestic dogs of today. Finally, we demonstrate the importance of geography in breed development and the role it plays in conclusions regarding ancient versus modern breeds.

### **W362: Domestication Genomics**

#### **The Genetic Architecture of Adaptation in the Long Term Barley Composite Cross Experiments**

Jacob Landis<sup>1</sup>, Christopher Fiscus<sup>1</sup>, Angelica Guercio<sup>1</sup> and **Daniel Koenig**<sup>2</sup>, (1)UC Riverside, (2)University of California-Riverside, Riverside, CA

Understanding adaptation in natural populations can be complicated by an unknown evolutionary history. We seek to complement studies of adaptation in natural systems using one of the worlds longest on going evolutionary experiments, the barley Composite Cross II (CCII). The CCII began in 1927 by intercrossing 28 barley varieties in a half diallele design. The progeny of these crosses were then competed against each other under field conditions in Davis, CA with minimal artificial selection for sixty generations. We leverage modern technological tools to understand how phenotypes have shifted in this population over evolutionary time, and to pinpoint the molecular changes that drive these changes. We find that strong natural selection has targeted genes involved in floral development, including the well characterized domestication locus *Vrs1*. We also document the nonrandom assembly of adapted multilocus haplotypes that dominate the population in later generations. Finally, I will compare the process of adaptation in this population in parallel experiments with different founding genetic diversity and conducted at different sites. These experiments provide us with a unique window into the process of adaptation under domestication.

### **W363: Domestication Genomics**

#### **Counter-Domestication: Maintenance of Atavisms under Domestication**

Travis Allen Parker, Jorge C. Berny Mier y Teran, Antonia Palkovic and **Paul Gepts**, University of California, Davis, CA

An atavism refers to the recurrence or reappearance of ancestral traits. In domestication studies, the presence of atavistic traits is an apparent contradiction of the concept of domestication syndrome. There are multiple causes for the presence of atavistic traits in a domesticated gene pool, among which gene flow from wild relatives. An alternative reason is heterogeneity in the genetic control of domestication trait, further enhanced by environmentally dependent expression of the trait. We illustrate this situation with the case of pod dehiscence (PD) in common bean (*Phaseolus vulgaris*). Based on evaluations of the PD trait in a biparental population and two diversity panels, we determined that different genes were selected in each domestication or ecogeographic race of common bean. In the Mesoamerican domesticated gene pool, a dirigent-like gene, involved in lignin biosynthesis, showed a base-pair substitution leading to a loss-of-function that is associated with decreased PD. This substitution arose after domestication after millions of years of purifying selection and underlies the expansion of Mesoamerican domesticates from the Central-Mexico

domestication region into northern Mexico, where arid conditions promote PD. Hybridizations between genotypes without PD but belonging to different gene pools or ecogeographic races lead to complementation restoring the PD atavism, which is further maintained in moist environments in which PD is not expressed. We suggest that this mechanism – complementation combined with environmental influence on expression – is responsible for the maintenance of atavistic traits in common bean and, potentially, other crops.

#### **W364: Duckweed Research and Applications**

##### **The Secret of Duckweed Survival: Molecular Characterization of Turion Formation in *Spirodela polyrhiza***

**Buntora Pasaribu**<sup>1</sup>, Kenneth Acosta<sup>2</sup> and Eric Lam<sup>2</sup>, (1)Department of Plant Biology, Rutgers, the State University of New Jersey, New Brunswick, NJ, (2)Department of Plant Biology, Rutgers, the State University of New Jersey, New Brunswick, NJ

*Spirodela polyrhiza*, commonly known as Greater Duckweed, exhibits a prominent survival behavior under conditions of abiotic stress. In unfavorable environments, such as low temperature, light and nutrient limitation, clones of *S. polyrhiza* and other species in the duckweed family are able to differentiate their shoot apical meristem to form a novel dormant structure that detaches from the mother frond. This so-called “turion” contains a high starch content and upon detachment sinks to the bottom of the water body until appropriate conditions for germination return and growth resumes. While some of the factors that affect this remarkable survival strategy are known, little has been clarified at the molecular and mechanistic levels. In the present study, we begin our molecular characterization of *S. polyrhiza* turion formation under phosphate limitation, a well-established environmental signal for turion induction. Leveraging previous studies on natural variations in specific turion yield (STY) of diverse populations, 5 genotypes were selected for a comparative analysis to examine the physiological response of these clones under phosphate limitation. Both the timing of turion appearance and maximum yield were quantified at different induction time points, as well as the starch content in their fronds. The results support and extend previous work by showing a good correlation between the timing of starch increase in the fronds and that of turion formation. Our analysis revealed the starch production in 5 selected *S. polyrhiza* clones were massively increased within 2 to 3 weeks after the start of phosphate deficiency, with high STY clones showing the most rapid response. We then carried out RNA-seq studies on *S. polyrhiza* clones containing different STY at different time points to delineate the gene expression networks in fronds and turions during this induction process. To enable this aim, we successfully developed an RNA extraction method to produce good quality RNA from *S. polyrhiza* fronds and turions containing high levels (up to 75%) of starch. We describe initial transcriptome comparative analyses between fronds and turions in *S. polyrhiza* clones 9512 and 9509, with high and low STY respectively, to identify candidate genes and the gene expression network that underlie this survival strategy of meristem transformation in the Greater Duckweed. (The support by a grant from the Department of Energy (DE-SC0018244) to E.L. for this project is gratefully acknowledged).

#### **W365: Duckweed Research and Applications**

##### **A Gene-Centric View of the Genomes of Lemnaceae**

**Shifra Ben-Dor**, Weizmann Institute of Science, Rehovot, Israel and Marvin Edelman, Weizmann Institute of Science, Israel

In recent years, genomes and transcriptomes of various members of Lemnaceae have become publically available. We have either one or both for *Spirodela*, *Lemna*, *Landoltia* and *Wolffia*, in some cases more than one isolate per genera, which allows us to perform cross-genome comparison. While comparisons

have been published on the whole genome level, we took a 'gene-centric' approach, manually curating large gene families to see what we could learn of the genome builds.

The two test cases we used are CDPKs (calcium dependent protein kinases) and aquaporins. By using the manually curated genes, we can detect polyploidy, conserved gene structure, and family size differences as compared to other plant genomes. Another important feature is the detection of genomic gaps even in the most polished versions of the current genomes. In light of the gaps, we found that curating the genes from both transcriptome and genome of a given species allowed for a more complete picture of the gene families.

## **W366: Duckweed Research and Applications**

### **Developing Duckweed as a Model System to Study Plant-Pathogen Interactions**

**Erin Baggs**, Mark Tiersma and Ksenia Krasileva, Department of Plant and Microbial Biology UC Berkeley, Berkeley, CA

The study of plant microbe interactions has historically relied heavily on the dicot model species *Arabidopsis thaliana* with favourable size, growth and reproductive cycles which has allowed research to advance rapidly. However, it has not been without its limitations, including the inability to transfer concepts derived from *Arabidopsis* to crop species within the distantly related monocots. Furthermore, there is a lot of redundancy in the *Arabidopsis* immune system which is comprised of over 200 NLR immune receptors and several redundant signaling pathways.

In a recent study by our lab we identified that *S. polyrhiza* (duckweed) has far fewer NLRs - only ~60 compared to an average of 600 in cereal monocots. Furthermore, *Spirodela* has lost a major immune signaling pathway which had previously been obscured by redundant immune pathways. In addition *S. polyrhiza*'s fast growth, clonal lineages and small size, make it an ideal monocot model organism for plant immunity. We have begun to develop assays to study the plant-pathogen interaction between *S. polyrhiza* and the generalist pathogen *Pseudomonas syringae*. We demonstrated that *P. syringae* is able to infect *S. polyrhiza* and dissected the contribution of phytohormones and effectors to *P. syringae*'s pathogenicity.

In developing *S. polyrhiza* pathosystems, we hope to create a tool for rapid identification of immunity components whose role or function was previously obscured by redundancy or absence in dicot model systems.

## **W367: Duckweed Research and Applications**

### **The Metabolome of the Duckweed Family – Metabolic Diversity and Biosynthetic Pathway Discovery**

**Uwe Heinig**, Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel and Nir Shahaf, Dong Yonghui, Liron Feldberg, Asaph Aharoni, all Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel

Recent year's renaissance in Duckweed research was centered on its great potential use as a nutritionally rich food source, sustainable biofuel and phytoremediation. However, the special features of duckweeds could be partially attributed to their metabolic repertoire. Despite its significance, only little is known with respect to duckweed metabolome composition and the biosynthetic pathways generating the diversity of unique chemicals in these species.

I will present deep metabolomics analysis of representatives of the five genera of duckweed, highlighting the metabolic diversity within the family. Furthermore using advanced stable isotope labelling techniques combined with high resolution mass spectrometry we were able to reconstruct *Lemna* specific natural product biosynthetic pathways and resolve their spacial localization within the plant.

To examine the *Lemnaceae* metabolome, extensive metabolic profiling was performed with 40 *Lemna* species including members of all 5 genera of the family. Besides identification of several hundred lipid species, 40 primary metabolites and 25 terpenoids, using LC/MS, a myriad of different secondary metabolites could be detected. For identification comparison to an in-house generated LC/MS reference library of authentic standards, WEIZMASS, was used. Searching against this LC/MS library resulted in high-confidence identification of 88 secondary metabolites from 9 different natural product classes. Newly developed analysis methods give insights into the natural product landscape of duckweed.

In recent years, the Aharoni lab developed a new method called DLEMMA-MS-Imaging, in which feeding of stable isotope labelled metabolite precursors allows high-confidence identification of metabolites and prediction of metabolic networks originating from the labelled precursor molecule. Once applied to *Spirodela polyrrhiza* this method facilitated the discovery of approx. 60 new metabolites and the reconstruction of an extensive metabolic network. Furthermore combining this labelling approach with MALDI-imaging mass spectrometry, made it possible to map metabolites in duckweed tissues and resolve between structural isomers based on their spatial distribution pattern in tissues.

In summary, this study describes the first comprehensive metabolic map of the duckweed family. Together with the recently released high quality genomes of several duckweed species it is now possible to identify genes/enzymes of duckweed natural product production and construct genome wide metabolic models to rationally engineer duckweed for the production of valuable natural compounds.

### **W368: Ecological Genomics**

#### **Precision Agroforestry for Carbon, Ecosystem and Climate Resilience**

**Justin O. Borevitz**, The Australian National University, Canberra, ACT, Australia

The 21st century is getting hot and crowded threatening food security and biodiversity. The land sector has the potential to switch from being a source to a sink for carbon to stabilize climate (ref) with benefits for food, fibre, soil fertility and habitat. Transforming agriculture, forestry and other land use will require improved practices that integrate precision genomics, phenomics and environmental mapping to predict best practices. These methods rely on bringing trees into the landscape, to secure ridge, creek and contour lines to define rotational cropping and grazing zones that best match climate and soil types. I will present results from a 1000 Eucalyptus genome2phenome2environment project which sets a foundation for this transformation.

### **W369: Ecological Genomics**

#### **Fifty Years after Ohno: Catostomid Fishes as a Model for Understanding the Role of Whole Genome Duplication in Ecological Diversification**

**Trevor Krabbenhoft**, University at Buffalo, Buffalo, NY, Tianying Lan, University at Buffalo, Amherst, NY, Hannah M. Waterman, University at Buffalo and Thomas E. Dowling, Wayne State University

In 1970, Susumu Ohno published his classic book "*Evolution by Gene Duplication*," in which he used catostomid fishes as one of his few examples of allopolyploidy in animals. Fifty years later, we present the first chromosome-scale catostomid genome assemblies and revisit Ohno's arguments in light of modern data. Catostomid fishes (suckers) diverged from diploid cypriniform relatives following an ancestral whole genome duplication event approximately 50 million years ago. Catostomids retain a duplicated karyotype ( $2n=100$  in suckers versus  $2n=50$  in most other cypriniforms), and in striking contrast to other polyploid fishes (e.g., salmonids), tetrasomy has not been observed. Based on these observations, Ohno and others have suggested that suckers are allotetraploids resulting from inter-specific hybridization, a common occurrence in cypriniform fishes. We assembled reference-quality genomes of Chinese sucker (*Myxocyprinus asiaticus*) – the sister lineage of all other catostomids, along with a morphologically-derived species (razorback sucker, *Xyrauchen texanus*), and a diploid outgroup (*Gyrinocheilus aymonieri*) using nanopore sequencing and Hi-C scaffolding. We compare patterns of

retention of duplicated genes across these species to test whether duplicate gene retention and expression is related to ecological diversification. We also examine whether there are genomic hot spots of gene loss in catostomids compared to diploid relatives and discuss the evolutionary significance of non-random gene loss across the catostomid phylogeny. Catostomid fishes represent a largely overlooked clade for studying the effects of polyploidy on patterns of ecological and phenotypic diversification, speciation, sex-determination, and hybridization dynamics.

### **W370: Ecological Genomics**

#### **Using Genomics Approaches to Understand Mechanisms of Response to Complex Environmental Conditions in Non-Model Plants**

**Christina Richards**, University of South Florida, Tampa, FL

### **W371: Ecological Genomics**

#### **Genome Evolution in Polar Fishes**

**Scott Hotaling**<sup>1</sup>, Thomas Desvignes<sup>2</sup>, Luana Lins<sup>3</sup> and Joanna L. Kelley<sup>1</sup>, (1)Washington State University, Pullman, WA, (2)University of Oregon, Eugene, OR, (3)CSIRO, Canberra, NSW, Australia

Many organisms have adapted to the harsh, subfreezing waters of the polar seas. The most ubiquitous and well-studied adaptation associated with this adaptation is the evolution of antifreeze proteins, a group of proteins that bind ice crystals to lower the freezing point of body fluids. Antifreeze proteins have arisen repeatedly at both poles in multiple fish families. The genomic basis of adaptation to polar seas, however, is almost certainly more complex than the rise of a single group of proteins. Yet, few genome-wide perspectives of adaptation to polar seas have been presented for fishes, limiting our general understanding of the process that underlies contemporary biodiversity in both the Arctic and Southern Oceans. Here we report the draft genome assembly of the eelpout *Ophthalmolycus amberensis* (family Zoarcidae), a resident of the Southern Ocean. Ours is the first genome sequence for the eelpouts, a cosmopolitan fish group comprising a substantial portion of marine fish diversity at high latitudes. Eelpouts are particularly interesting because unlike other commonly studied polar fish groups -- e.g., cod (family Gadidae; Arctic Ocean) and icefishes (family Channichthyidae; Southern Ocean) -- eelpouts have adapted both the Arctic and Southern Ocean. The *O. amberensis* genome contains 680.7 million basepairs (Mb) in 1,828 contigs with a contig N50 of ~1 Mb. The assembly contains 95.7% of the single-copy, conserved orthologous genes (BUSCOs) in the reference set for Actinopterygii. We annotated 22,572 genes in the genome and compared this new resource to previously sequenced polar fish genomes to gain new insight into the genomic basis of adaptation to polar seas.

### **W372: Ecological Genomics**

#### **Within- and Trans-Generational Plasticity in Response to Climate Cues**

**Mariano Alvarez**, Duke University, Durham, NC, Gabriela Auge, Fundación Instituto Leloir, Ciudad Autónoma de Buenos Aires, Argentina and Kathleen Donohue, Duke University

Phenotypes are generated by both the genetic variation within an individual as well as the environments that they experience. Crucially, environmental influence on phenotype can come from both environments that are experienced by an individual and from environments that have been experienced by prior generations, and the ability to respond to these cues may be adaptive. The transmission of environmental cues between generations is known as parental effects or transgenerational plasticity; when these changes include modifications at the molecular level, the phenomenon is known as epigenetic inheritance. Using three generations of *Arabidopsis thaliana*, we describe complex patterns of transgenerational plasticity in phenology, biomass, and fitness. We also identify candidate genes that may underlie the propensity to express these effects, and outline new computational methods to search for selection for this likely polygenic trait.

### **W373: Ecological Genomics**

#### **Transcriptome-Wide Gene and Protein Expression Differences in the Intertidal Mussel *Mytilus californianus* Exposed to Field and Lab Treatments**

**Lani Gleason**, California State University, Sacramento, Sacramento, CA

Although rare, datasets examining both gene and protein expression can increase understanding of organisms' responses to environmental stressors. Previous work has identified physiological differences in intertidal *Mytilus californianus* mussels living in unique temperature habitats; however, the gene and protein expression patterns driving these physiological differences have not been investigated. The objectives of this study were 1) to determine whether RNA and protein expression follow similar patterns in *M. californianus*, 2) to identify genes and proteins that change expression across different field and lab treatments, and 3) to identify genes and proteins whose expression correlates with physiological data.

To achieve these aims, we used bioinformatics techniques to analyze RNA-sequencing and proteomics data for ~1500 genes obtained from 51 *M. californianus* individuals exposed to 5 intertidal and lab conditions that vary in environmental stressors such as temperature. Genes with significantly different RNA and protein expression between treatments were identified using the software packages DESeq2 and EdgeR to address objectives 1 and 2, and the package WGCNA was used to address objective 3.

For objective 1, we found that many genes that are differentially expressed at the RNA level are not differentially expressed at the protein level (and vice versa). For objective 2, results indicate that genes related to cilia, motility, and protein folding show differential expression between treatments. No common functions were found in differentially expressed proteins. For objective 3, we found that nucleic acid and ion binding genes and protein folding and proteolysis proteins have expression patterns that correlate with physiological traits such as catalase enzyme activity.

Given the unique expression patterns of RNA vs. protein, we conclude that post-transcriptional and/or post-translational processes should be considered when examining *M. californianus*' response to environmental differences. In addition, we hypothesize that differential expression of cilia genes are related to a differential need for transporting fluid within *M. californianus*' respiratory cavity. Overall, these results provide insight into the genes contributing to physiological differences in *M. californianus* and can enhance understanding of intertidal invertebrates' responses to climate change.

### **W374: Engineering NUE**

#### **Identification of a Key Player on Nitrogen Use Efficiency of Rice**

**Ju-Kon Kim**, Seoul National University, Pyeongchang, Gangwon, South Korea and Jae Sung Shim, Seoul National University, Pyeongchang, South Korea

Crops obtain insufficient N sources from natural supply of soil and rest of N sources are inevitably supplied by N fertilizers. N fertilizers increase crop yields, however, overdose of the N fertilizers results in serious damages to environment. Little is known about the molecular NUE mechanisms to solve the long-standing problem due to lack of a typical N phenotype. *ALLANTOINASE (ALN)* and *UREIDE PERMEASE 1 (UPS1)*, genes that modulate ureide metabolism, were identified to be most responsive to N status. *ALN* was up-regulated under low N conditions, whereas *UPS1* was sensitively up-regulated under high N conditions. We evaluated an activation tagging rice plants overexpressing *UPS1 (UPS1<sup>OX</sup>)* under several N regimes. Under normal field conditions, panicles from *UPS1<sup>OX</sup>* plants showed significant allantoin accumulation. Under hydroponic system at the vegetative stage, *UPS1<sup>OX</sup>* plants displayed higher ammonium uptake in roots compared to wild type (WT). When grown under low-N soil, *UPS1<sup>OX</sup>* exhibited better growth at 50% N showing higher chlorophyll, tiller number and at least 20% increase in shoot and root biomass relative to WT. We propose that *UPS1* is responsible for allantoin partitioning in rice and its overexpression can support plant growth through accumulation of allantoin in sink tissues. Taking advantage of their nature in response to N status, we generated N molecular sensors using *proALN::ALN-LUC2* and *proUPS1::UPS1-LUC2* in transgenic rice plants. Transgenic plants with *proUPS1::UPS1-LUC2* sensor showed strong luciferase activity under high N conditions (> 1 mM

ammonium nitrate), whereas transgenic plants with *proALN::ALN-LUC2* sensor showed strong luciferase activity under low N conditions (< 1 mM ammonium nitrate). We treated the transgenic plants with 1% EMS mutagen to generate an EMS mutant population. With the EMS mutant population (20,000 M<sub>1</sub> lines), we are identifying a key player on NUE of rice using the N molecular sensors.

### **W375: Engineering NUE**

#### **Ms44 Male Sterile Hybrid Improves NUE in Maize**

**Bo Shen**, R & D Trait Discovery; Corteva Agriscience, Johnston, IA

Under low nitrogen condition, tassel development is dominant over ear development. A reduction of tassel dominance via male sterility could improve ear development and increases grain yield under stress conditions. *Ms44* is a dominant male sterile mutant encoding an anther specific protein of unknown function. Male sterile hybrids carrying the *Ms44* allele demonstrated a 4%–8.5% yield advantage when N is limiting, 1.7% yield advantage under drought and 0.9% yield advantage under optimal growth conditions relative to the yield of fertile wild type. Furthermore, we have developed an *Ms44* maintainer line for fertility restoration, male-sterile inbred seed increase and hybrid seed production. This study demonstrates that a reduction in competition between tassel and ear by male sterility improves grain yield under low-nitrogen conditions in maize.

### **W376: Engineering NUE**

#### **Manipulation of Flowering Pathway Genes for High Fruit Productivity**

**Guo-Qing Song**, Michigan State University, East Lansing, MI

Blueberries contain high amounts of antioxidants known to be important for human health. Developing new cultivars with different chilling requirement, high cold/heat tolerance, and for high yield are among the top priorities in blueberry breeding due particularly to the anticipation of climate changes and the rapidly expanding market need of blueberry products. Overexpression of a blueberry *DWARF AND DELAYED FLOWERING 1* (*VcDDF1*) increased freezing tolerance without a trade-off impact on yield. Turning on a blueberry a *RESPONSE REGULATOR 2*-like gene (*VcRR2*) in a mutant caused by the *VcDDF1* transgene insertion resulted in reduced chilling requirement for flowering and a high yield potential. The mutant provides an outstanding material to study chilling-mediated flowering mechanism in woody plants. Transgenic blueberries overexpressing a blueberry *FLOWERING LOCUS T* (*VcFT*) facilitate FAST-TRACK blueberry breeding through transgrafting. Transgrafting on *VcFT*-overexpressing blueberry plants promoted floral bud formation in nontransgenic scions and thus not only demonstrates that hormones are involved in *FT*-induced long-distance transport of the florigenic signals but also provides a new approach to increase blueberry yield. Overexpression of the K-domain of a blueberry *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* gene increased berry productivity through the interaction of MIKCMADS-box genes. This new K-domain technology is being tested in corn and works as well. Overall, we demonstrate that manipulation of flowering pathway gene(s) or hormone synthesis pathway gene(s) is a powerful approach to increase fruit/crop productivity.

### **W377: Engineering NUE**

#### **Nitrogen Phosphate Signaling Interactions in Plants from Arabidopsis to Crops**

**Gabriel Krouk**, CNRS, Montpellier, France

### **W378: Engineering NUE**

#### **Nitrogen Use Efficient Rice Heading into Regulatory Field Trials in Africa**

**Jos van Boxtel**, Arcadia Biosciences, Davis, CA

Through combined AATF/USAID/Arcadia/CIAT/NARS efforts, five years of comparative field trials with Nitrogen Use Efficient (NUE) rice, identified two lead events, of which one is currently being trialed in

four locations in Africa to compile data for a regulatory dossier. The goal is to achieve approval for cultivation and human consumption of GM NUE rice in various countries in the region.

The NUE trait, based on the overexpression of an alanine aminotransferase in NERICA rice, displays grain yield increases compared to controls, under limiting nitrogen availability. NUE rice is characterized by an early establishment of the crop in the field and increased tillering/biomass to support an abundant seed set and grain harvest. Comparative trials, conducted in Colombia, Uganda, Ghana and Nigeria, showed grain yield increases as high as 30% compared with controls, using low rates of nitrogen fertilizer.

NUE rice is an adequate solution for yield increases in low-input agricultural systems to the benefit of small-holder farmers. In high-input agricultural systems, NUE rice brings moderate yield increases, but has also the potential to reduce the loss of unused fertilizer into the environment, leading to N leaching in waterways and emission of greenhouse gases.

### **W379: Equine 1**

#### **Current Knowledge of the Equine Microbiome**

**Marcio Costa**, University of Montreal, Saint-Hyacinthe, QC, Canada

The development of new DNA sequencing technologies has allowed a broader assessment of complex microbial environments, such as the gastro-intestinal tract (GIT). Next generation sequencing has been recently used to investigate the equine GIT microbiota, demonstrating that many factors are associated with changes in bacterial composition in that species. Among other factors, diet, age, feeding management, pregnancy, transportation, intense exercise and stress have been shown to alter the GIT bacteria. Furthermore, many diseases such as colitis, colic, equine grass sickness, equine metabolic disease and parasitism have been associated with changes in the fecal microbiota and with a decreased number of bacterial species (decreased richness and diversity).

At this point, we face the challenge of interpreting these recently generated data in a way that it can be used to improve horse health. Specifically, we need to learn how to manipulate the GIT microbiota to treat and prevent diseases, and to improve performance. The most used methods of microbiota manipulation are diet modulation, pro and prebiotics and fecal microbiota transplantation (FMT), in which the whole ecosystem obtained from one or multiple healthy donors is transferred into a diseased patient. The use of probiotics remains controversial in horses mainly because of conflicting results observed in scientific publications. Furthermore, the use of probiotics requires continuous administration indicating that strains used in those products do not permanently colonize the host. Prebiotics have a great potential to benefit certain groups of microorganisms because they are not digested by the horse. However, the mechanisms by how (and where) those products act in the GIT require further research.

FMT has been used with great success to treat GIT disorders in other species including humans, dogs and laboratory animals. It has been shown that FMT is much more efficient when administered rectally compared to the oral route (or intragastrically) because environmental filters (e.g. gastric pH, enzymatic digestion, transit time) may decrease bacterial viability. However, horses have a long small colon, which precludes the use of enemas to deliver FMT efficiently into their large colon. Although used in horses for many decades, there is no scientific evidence that the procedure is clinically efficient and that the transferred bacteria can in fact colonize their GIT.

In conclusion, this is an exciting field that has great potential to be used for diagnosis, prevention and treatment of many conditions affecting horses, but science based evidences are required before further recommendations can be made.

### **W380: Equine 1**

#### **Equine Hybrids and their Genetic Foundation in History**

Antoine Fages<sup>1</sup>, Naveed Khan<sup>2</sup>, Pablo Librado<sup>2</sup>, Azadeh Mohaseb<sup>3</sup>, Marjan Mashkour<sup>3</sup> and **Ludovic Orlando**<sup>4</sup>, (1)Laboratoire d'Anthropologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse III Paul Sabatier, Toulouse, France, (2)CNRS, University Paul Sabatier, Toulouse, France, (3)CNRS, Natural History Museum of Paris, Paris, France, (4)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Denmark, Copenhagen, Denmark

The domestication of the horse some 5,500 years ago represents a historical turning point in human history as the horse provided mankind with the ability to travel well above its own speed and changed the way we made war. In addition to the many new horse varieties that developed throughout history and formed the basis of famous equestrian civilizations, another type of equine animal soon became instrumental to the economic and military life of past societies as it outcompeted the horse in a number of tasks. This animal was the mule, the first-generation offspring of a donkey jack and a horse mare, which was more resistant, more sure-footed, stronger and lived generally longer than the horse. Mule breeding was, however, extremely demanding as the animal was sterile and could thus not be propagated generation after generation. It required instead significant financial investment for ensuring the maintenance of both horse and donkey parental stocks. Despite the importance of mules in human history, when and where the first mule breeding centers emerged and which strategies underpinned mule production in Antiquity remains highly contentious. This is so due to the difficulty to taxonomically assign archaeological remains that are often fragmentary. Ancient DNA analyses such as those underlying the Zonkey framework can, however, easily identify hybrids, even when only limited amounts of sequences are available. Applying Zonkey as well as other innovative solutions aimed at recovering genome-wide data from mule individuals, our work has deeply revisited both the temporal and geographic loci of the earliest mules produced in (pre)history. The extensive genetic information collected can also provide invaluable information about the horse and donkey parental stocks maintained and utilized by mule breeders. This helps address whether similar mule breeding centers were established by different equestrian civilizations such as the Romans and Byzantines, and whether the two parental species were treated similarly or were instead exposed to different breeding strategies and goals.

## **W381: Equine 1**

### **Equine Y Chromosome Variability**

**Caitlin Castaneda**, Texas A&M University, College Station, TX

Copy number (CN) and copy number variation (CNV) cause phenotypic variations, however CN functionality in horses remains enigmatic despite the updated horse genome EquCab3 and recent completion of the horse Y chromosome reference assembly. Here, we continue CN and CNV studies of Y-linked ampliconic genes between individuals and breeds to compare CNs between reproductively normal males and those with various disorders of sexual development (DSD). Until now, we have generated reproducible data by ddPCR for 8 ampliconic genes and male sex determining gene *SRY* across 20 selected breeds which include Thoroughbred-related commercial breeds, Przewalski's Horse, Oriental and isolated breeds. We also include a cohort of 25 subfertile cryptorchid males with spermatogenic defects and 41 horses of ambiguous sex. We observe high degree of CN conservation across Thoroughbred related breeds with Testis-Specific Protein on Y (*TSPY*) having the greatest range (7-17 copies) and Equine Testis Specific Transcript Protein 2 (*ETSTY2*) having the least variability between normal individuals (3-5 copies). Interestingly, *ETSTY* genes have greater variability between cryptorchid males and those phenotypically normal. Therefore, the observed conservation of gene CNs in normal populations suggests that CN stability of ampliconic genes may be functionally important. Additionally, CN analysis of *SRY* in horses with DSD show a loss in *SRY* results in a corresponding CN loss in neighboring ampliconic genes along with sex-reversal phenotypes. Novel findings of increased *SRY* and neighboring ampliconic CN was observed in phenotypically normal individuals Przewalski's Horses and remote breeds. Increased *SRY* CN occurrence may be higher among remote horse breeds or horse families with DSD individuals. We are currently expanding CN research to compare Y SNP and haplotype variations using the same cohort. The findings are expected to reveal comparative contribution

of CNVs and SNPs to Y chromosome variation in horses, to gain a deeper understanding of horse Y variability.

### **W382: Equine 1**

#### **An 8.7kb Deletion in *MITF* Explains a Novel Splashed White Phenotype in the American Paint Horse**

**Jocelyn L. Tanaka**, University of California - Davis, Davis, CA, Gary Magdesian, University of California Davis, Davis, CA and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Splashed white is a coat color pattern in the horse characterized by extensive white patterning on the legs, belly, and face that is often accompanied by blue eyes and deafness. To date, three mutations, in *Microphthalmia-Associated Transcription Factor (MITF)* and two mutations in *Paired Box 3 (PAX3)* have been identified that explain splashed white patterns (SW1-5). An American Paint Horse stallion with a splashed white phenotype and blue eyes, whose parents were not white patterned, was unexplained by any of the five known splashed white variants or any other known white spotting patterns. We hypothesized this splashed white phenotypes (SW6) was caused by a denovo mutation in *MITF* or *PAX3*. Analysis of whole genome sequencing using Illumina Novaseq technology with 150bp paired end reads to an average depth of 52X coverage identified an 8.7kb deletion in *MITF*. This variant removes 625 coding nucleotides, and is therefore predicted to impair protein function. No SNPs or structural variants were identified in *PAX3*. Sanger sequencing confirmed the stallion was heterozygous for the *MITF* deletion. Genotyping three of his splashed white offspring found that they were also heterozygous for the deletion. One additional offspring did not have the deletion and phenotypically and genotypically was identified as a frame overo like her dam. Given the role of *MITF* in producing white patterning phenotypes, and the predicted deleterious effect of this mutation, this 8.7 kb deletion is the likely causal variant for SW6.

### **W383: Equine 1**

#### **Genome-Wide Association Analyses of Equine Metabolic Syndrome Phenotypes in Welsh Ponies and Morgan Horses**

**Elaine M Norton**<sup>1</sup>, NE Schultz<sup>1</sup>, Ray Geor<sup>2</sup>, Dianne McFarlane<sup>3</sup>, James R Mickelson<sup>1</sup> and Molly E McCue<sup>1</sup>, (1)University of Minnesota, St. Paul, MN, (2)Massey University, Palmerston North, New Zealand, (3)Oklahoma State University, Stillwater, OK

Equine metabolic syndrome (EMS) is a complex trait for which few genetic studies have been published. The study objectives were to perform within breed genome-wide association analyses (GWA) to identify associated loci in two high-risk breeds, coupled with meta-analysis to identify shared and unique loci between breeds. GWA for twelve EMS traits identified 303 and 142 associated genomic regions in 264 Welsh ponies and 286 Morgan horses, respectively. Meta-analysis demonstrated that 65 GWA regions were shared across breeds. Region boundaries were defined based on a fixed-size or the breakdown of linkage disequilibrium, and prioritized if they were: shared between breeds or across traits (high priority), identified in a single GWA cohort (medium priority), or shared across traits with no SNPs reaching genome-wide significance (low priority), resulting in 56 high, 26 medium, and 7 low priority regions including 1,853 candidate genes in the Welsh ponies; and 39 high, 8 medium, and 9 low priority regions including 1,167 candidate genes in the Morgans. The prioritized regions contained protein-coding genes which were functionally enriched for pathways associated with inflammation, glucose metabolism, or lipid metabolism. These data demonstrate that EMS is a polygenic trait with breed-specific risk alleles as well as those shared across breeds

### **W384: Equine 1**

#### **Inherited Hypocalcemia in Thoroughbred Foals is associated with a Nonsense Mutation in *RAPGEF5***

**Victor Rivas**, University of California-Davis, Davis, CA

Inherited Hypocalcemia (IH) is a neonatal disease that occurs only in Thoroughbreds. The disease is characterized by a profound hypocalcemia with inappropriately normal parathyroid hormone (PTH) concentrations, loss of maneuverability of limbs, seizures, muscle fasciculations, ileus, tachycardia, synchronous diaphragmatic flutter, and ataxia. Previously, DNA from two IH-affected foals, their dams, and two unrelated clinically healthy Thoroughbred horses underwent next-generation sequencing and a whole-genome association study was performed. A segregating nonsense variant in exon 28 of *RAPGEF5* was significantly associated with the IH phenotype. In 2019, samples from an additional two suspect IH foals were collected and Sanger sequencing revealed that both foals were homozygous for the nonsense variant. Preliminary screening suggests that this variant is present in low frequency in the Thoroughbred population. Out of 81 genotyped Thoroughbred horses, only three were carriers for the variant ( $q=0.019$ ). In order to test if the IH equine *RAPGEF5* mutation affects protein function, we overexpressed the wild-type and IH variant protein in frog embryos. Overexpression of *RAPGEF5* produced a stereotypical phenotype that was not observed in the IH variant. This assay suggested loss-of-function of the protein with the IH equine mutation. Therefore, there is strong evidence that the nonsense variant in *RAPGEF5* is associated with the IH phenotype.

### **W385: Equine 1**

#### **Whole Genome Sequencing Identifies Missense Mutation in GRM6 as the Putative Cause of Congenital Stationary Night Blindness in a Tennessee Walking Horse**

**Yael Leah Hack**, University of California- Davis, Davis, CA, Elizabeth Crabtree, North Carolina State University College of Veterinary Medicine, Raleigh, NC, Felipe F Avila, UC Davis, Davis, CA, Roger B. Sutton, Texas Tech University Health Sciences Center School of Medicine, Lubbock, TX, Brian C. Gilger, North Carolina State University, Raleigh, NC and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Congenital stationary night blindness (CSNB) is an inherited, non-progressive retinal disorder characterized by absence of vision under low-light conditions. Currently, the only known genetic cause of CSNB in the horse is an insertion in *transient receptor potential cation channel subfamily M member 1* (*TRPM1*, ECA1g. 108,297,929\_108,297,930 ins1378.) However, one Tennessee Walking Horse diagnosed with CSNB did not have this mutation. To identify a causal variant, Illumina Novaseq whole genome sequence data from this case was compared to data from horses from seven other breeds ( $n=33$ ). One hundred and two candidate genes, identified from human and mouse literature, were assessed for coding variants. Variants in these candidate genes homozygous in the case and absent in all other horses were prioritized for further investigation. A single missense mutation in *metabotropic glutamate receptor 6* (*GRM6*) (c.533C>T p.Thr178Met), a gene known to cause CSNB in humans, was identified. This SNP was predicted to be deleterious with 61% confidence by the consensus classifier PredictSNP. Thr178 is highly conserved across vertebrate species and is directly involved in binding the neurotransmitter glutamate, and is thus essential to on-bipolar cell signaling that enables vision in low light conditions. Methionine at position 178 is hypothesized to impair binding of glutamate, which was supported by protein modeling. In screening 80 unrelated Tennessee Walking Horses, the estimated allele frequency was 8.1% with no other homozygotes identified. Taken together, these data provide evidence that this SNP is causal for CSNB in this breed. Additional testing is warranted to evaluate if this variant is present in other breeds with CSNB.

### **W386: Equine 1**

#### **A 16 Kilobase Deletion on ECA13 Is Strongly Associated with Distichiasis in Friesian Horses**

**Erin A. Hisey**<sup>1</sup>, Hanneke Hermans<sup>2</sup>, Zachary T. Lounsberry<sup>3</sup>, Felipe F Avila<sup>4</sup>, Robert A. Grahn<sup>5</sup>, Tomasz Ząbek<sup>6</sup>, Ted Kalbfleisch<sup>7</sup>, Mary Lassaline<sup>8</sup>, Willem Back<sup>9</sup> and Rebecca R. Bellone<sup>3</sup>, (1)University of California, Davis School of Veterinary Medicine, Davis, CA, (2)Utrecht University, Utrecht, Netherlands, (3)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (4)UC Davis, Davis, CA, (5)University of California, Davis, Davis, CA, (6)Laboratory of Genomics National

Research Institute of Animal Production, Balice, Poland, (7)University of Kentucky, Lexington, KY, (8)Department of Clinical Surgical & Radiological Sciences, University of California, Davis, Davis, CA, (9)Ghent University, Merelbeke, Belgium

Distichiasis, an ocular condition reported in Friesian horses, occurs when aberrant lashes grow from the Meibomian glands of the eyelid. These lashes can cause chronic irritation and corneal ulcers, potentially leading to vision loss from corneal scarring. Because of its bilateral nature and prevalence in a breed with known inherited monogenic disorders, this condition is hypothesized to be a simply inherited Mendelian trait. To test this hypothesis, a genome wide association study (GWAS) was performed using the Axiom 670k Equine Genotyping array (MNEc670k) on 14 cases and 38 controls clinically phenotyped for distichiasis. With an additive single locus mixed linear model (EMMAX) approach, a 1.83 Mb locus on ECA5 and a 1.34 Mb locus on ECA13 were identified that reached genome-wide significance ( $p_{\text{corrected}}=0.016$  and  $0.032$ , respectively). Only the locus on ECA13 withstood replication testing ( $p_{\text{combined}}=3.01 \times 10^{-5}$ ). A 371-kb run of homozygosity on ECA13 was found in 13 of the 14 cases providing evidence for a recessive mode of inheritance. A haplotype analysis (hapQTL) narrowed the region of association on ECA13 to 163 kb. Whole-genome, high-throughput sequencing data from 3 cases and 2 controls identified a 16-kb deletion from the ECA13 associated haplotype that contains reported putative regulatory elements. This deletion was strongly associated with distichiasis, as 18 of the 19 cases were homozygous ( $p=6.0 \times 10^{-10}$ ). Further functional analyses of this variant will clarify its role in lash development.

### **W387: Equine 1**

#### **The Frequency of Loss of Function Alleles in the Equine Population**

**Sian Ann Durward-Akhurst**, University of Minnesota, St. Paul, MN

Identification of disease-causing alleles is a fundamental goal of medical genetics and can facilitate disease diagnosis and improve disease understanding. Studies in humans have increasingly demonstrated the value of databases of genetic variation derived from genome sequencing (WGS) for disease-causing allele(s) identification. A surprising finding from these studies is the high number of alleles computationally predicted to lead to loss of function (LOF) of the affected gene in healthy adults. LOF variants with a high frequency in healthy individuals are unlikely to be disease-causing. Here, we investigate LOF alleles present in the equine population.

We mapped WGS from 535 horses to the EquCab3 reference genome. Single nucleotide polymorphisms (SNPs) and short insertions/deletions were identified using GATK-HaplotypeCaller and SAMtools, and annotated using ANNOVAR and SnpEff. The intersect was used to identify LOF variants i.e., nonsense, frameshift, and splice site disrupting variants, or deletions removing the first exon or >50% of protein coding sequence. Average depth of coverage was 11.5x (range 1.4 - 46.7x). 29,882,273 variants were identified, with 8,683 predicted to be detrimental, including 5,673 LOF variants affecting 3,810 genes enriched for olfactory reception and immune related pathways. On average, each horse carried 829 (range 211 - 1,182).

Overall, we demonstrate that similar to humans, LOF alleles are present in the horse population. We will further validate the LOF alleles using hand annotation and produce a list of LOF alleles that can be excluded from candidate disease-causing allele discovery approaches due to their high frequency in the general population.

### **W388: Equine 2**

#### **Mapping CTCF Binding Regions in the Equine Genome**

**Sichong Peng**, University of California, Davis, Davis, CA

CCCTC-binding factor (CTCF) is a zinc finger protein that serves as a core architect protein in chromosome 3D structures, which dictates the chromosome regulatory activities. CTCF binds at chromosomal domain boundaries and often functions as an insulator, blocking inter-domain DNA

interactions. Chromatin immunoprecipitation (ChIP) can be used to pull down CTCF-bound DNA and subsequent sequencing of enriched DNA can identify these bound regions. As part of Functional Annotation of Animal Genomes (FAANG) project, we utilized ChIP-seq to sequence CTCF bound regions on eight prioritized tissues (adipose, brain, heart, lamina, liver, lung, muscle, and ovary) from two healthy adult Thoroughbred mares. The specificity and efficiency of four different anti-CTCF antibodies were evaluated through qPCR using primer sets for both positive (H19 imprinting control region) and negative (myoglobin, exon 2) regions. Libraries were combined into four pools and each pool sequenced using Illumina HiSeq 3000/4000. An average of 28M reads were obtained from each library. Reads were mapped to EquCab3 reference genome after QC and peaks were called using MACS2 narrow peak setting for each library. An average of 53,000 peaks were identified for each tissue type after merging biological replicates (adipose: 51k, brain: 49k, heart: 63k, lamina: 53k, liver: 46k, lung: 52k, muscle: 48k, and ovary: 68k). ChIP-seq for histone marks from the same tissue samples were then incorporated with CTCF data and ChromHMM was used to predict chromatin states in eight prioritized tissues.

### **W389: Equine 2**

#### **Functionally Annotating Regulatory Elements in the Equine Genome using Histone Mark ChIP-Seq**

**Nicole B. Kingsley**<sup>1</sup>, Colin Kern<sup>2</sup>, Catherine Creppe<sup>3</sup>, Erin N. Burns<sup>4</sup>, Huaijun Zhou<sup>2</sup>, Ted Kalbfleisch<sup>5</sup>, James N. MacLeod<sup>6</sup>, Jessica L. Petersen<sup>7</sup>, Carrie J. Finno<sup>8</sup> and Rebecca R. Bellone<sup>9</sup>, (1)University of California, Davis, Davis, CA, (2)Department of Animal Science, University of California, Davis, CA, (3)Diagenode, Liège, Belgium, (4)University of California - Davis, Davis, CA, (5)University of Louisville, Louisville, KY, (6)Department of Veterinary Science, University of Kentucky, Lexington, KY, (7)University of Nebraska-Lincoln, Lincoln, NE, (8)UC Davis, Davis, CA, (9)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

As part of the Functional Annotation of ANimal Genomes (FAANG) initiative, we aimed to characterize tissue-specific regulation within the equine genome. Chromatin Immunoprecipitation Sequencing (ChIP-Seq) was used to map four histone modifications (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) in eight prioritized tissues (adipose, brain, heart, lamina, liver, lung, muscle, and ovary) collected from two adult Thoroughbred mares. Data were generated according to optimized experimental parameters developed during quality control testing. To ensure sufficient IP and successful peak-calling, data and peak-calls were assessed using six quality metrics, replicate comparisons, and site-specific evaluations. Additionally, combinations of marks were used to identify active regions unique to each tissue. We found that liver and lamina were the tissues with the most unique active elements (19,028 and 17,206, respectively) while brain and liver had the most tissue-specific repressed regions (1610 and 1606, respectively). Tissue specificity was further explored by identifying binding motifs within the unique active regions, and motifs were characterized by gene ontology (GO) term and protein-protein interaction network analyses. The histone marks identified in this study represent some of the first publicly available resources for investigating tissue-specific regulation within the equine genome. As such, these annotation data can be used to advance equine studies investigating health, performance, reproduction, and other traits of economic interest in the horse.

### **W390: Equine 2**

#### **Integrating Transcriptomics, Proteomics, and *cis*-Regulatory Networks to Understand Muscle Physiology and Pathophysiology**

**Deborah Velez-Irizarry**, Department of Large Animal Clinical Sciences, Michigan State University, East Lansing, MI

Gene set enrichment analysis (GSEA) is standard practice for large scale omics experiments. GSEA uses a priori gene lists from curated databases to group experimentally derived gene sets into biological pathways in order to infer underlying physiological states. A transcription factor motif enrichment analysis (TFMEA) follows the same basic principles as GSEA except inferences are made on *cis*-regulatory states and functional transcription factor (TF) binding. When applied to multi-omics datasets, TFMEA and GSEA can recognize coordinated physiological and pathophysiological processes by identifying *cis*-

regulatory networks in differentially expressed gene sets. We have applied this integrated approach using transcriptomics and proteomics to study myofibrillar myopathy (MFM, n=16), recurrent exertional rhabdomyolysis (RER, n=15) and glycogen depletion/repletion (GDR, n=40) in horses. We first identified differentially expressed (DE) genes/proteins for each project using weighted linear regression/permutation tests ( $FDR \leq 0.05$ ) followed by TFMEA ( $NES \leq 3.0$ ) and GSEA ( $FDR \leq 0.05$ ) for putative target genes (PTG). We focused on TFs differentially regulated in disease or metabolic state. This approach identified enriched TF DE (eTF-DE) in horses with MFM ( $\uparrow ATF3$ , an anti-inflammatory factor) in association with upregulation of its PTG involved in protein phosphorylation. In RER, eTF-DE ( $\uparrow HSF4$ ) identified PTG associated with the mitochondrial TCA cycle and nitrogen metabolism; *HSF4* is also a potential target of the RER treatment dantrolene. eTF-DE in response to GDR identified  $\uparrow RUNX1$  with PTG enriched for G protein-coupled receptor downstream signaling involved in glucose homeostasis. Our research shows the added value of TFMEA for functional genomic studies of equine skeletal muscle disorders and glycogen metabolism.

## **W391: Equine 2**

### **Single Nucleotide Polymorphism Characterization of Major Histocompatibility Complex Haplotype Diversity in Standardbred Horses**

**Lucia Younger**<sup>1</sup>, Annette M McCoy<sup>2</sup>, Donald Miller<sup>1</sup>, Molly E McCue<sup>3</sup>, James R Mickelson<sup>3</sup> and Douglas Antczak<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)University of Illinois, Urbana, IL, (3)University of Minnesota, St. Paul, MN

Identification of Major Histocompatibility Complex (MHC) haplotypes remains challenging because of the expected high, but unknown, level of diversity within any given species and the difficulty of determining polymorphism across the approximately 4 million base pair extent of the MHC. In the horse, the MHC has been characterized using various methods including serology and microsatellites. Here we explored MHC haplotype diversity in the equine Standardbred breed using 670K SNP chip data from 297 Standardbred horses. Single Nucleotide Polymorphisms (SNP) in the MHC region were extracted and analyzed using the bioinformatic software SHAPEIT. This allowed identification of haplotype-specific SNP patterns and phasing of MHC SNP-based haplotypes in MHC heterozygous horses. We found a total of 80 unique haplotypes comprised of 47 class I and 20 class II blocks. Twenty-five MHC homozygous horses were identified that each carried only one of six common haplotypes. Intra-MHC microsatellite testing is underway to determine the relationship between SNP-based and microsatellite-based MHC haplotypes. Thus far, tests performed on DNA from 31 of the horses identified 17 microsatellite haplotypes, 8 of which had previously been described in Standardbreds, and complete correlation between SNP and microsatellite haplotypes. SNP-based MHC typing holds promise for characterization of polymorphism of the equine Major Histocompatibility Complex. Understanding the amount of diversity in the MHC has many applications in respect to the evolution and physiology of the horse.

## **W392: Evolution of Genome Size**

### **Hybrid Speciation in *Brassica***

**Elvis Tembang Katche**, Department of Plant Breeding, Justus-Liebig University Giessen, Giessen, Germany

Interspecific hybridisation is an important path to generating evolutionary novelty. The genus *Brassica* is an important agricultural genus with a history of interspecific hybridisation. The allotetraploids *B. juncea*, *Brassica carinata*, and *Brassica napus* are hybrids formed by pairwise hybridisation of the diploid progenitors *Brassica rapa*, *Brassica oleracea* and *Brassica nigra*. Though crossing the allotetraploids is possible and has been carried in several studies either to study chromosome behavior or transfer useful traits, attempts to generate novel, stable and fertile synthetic hybrids through this method have not been reported. We generated interspecific hybrids (BBAC = F1 = 35) by crossing *B. juncea* (2n = AABB = 36) × *B. carinata* (2n = BBCC = 34) where A and C genomes lack homologous pairing partners and self-pollinated these hybrids for six generations by selecting for fertility. We intend to answer the question if in

the absence of homologous pairing partners, the A and C genomes can pair, restructure and stabilise to form viable and fertile off springs. Analysis of later generation hybrids show that recombination and genome restricting between the A and C genome can lead to the formation of new, stable and fertile hybrids with stable chromosome transmission across generations. This pathway can be useful for generating evolutionary novelty which can be transferred to other *Brassica* species and also to produce new useful crop types.

### **W393: Evolution of Genome Size**

#### **The Interaction between LTR Retrotransposons and DNA Transposons in Plants**

**Ning Jiang**, Department of Horticulture, Michigan State University, East Lansing, MI

Ning Jiang (presenter), Stefan Cerbin, Dongyan Zhao, and Shujun Ou

Department of Horticulture, Michigan State University, East Lansing, MI 48824

Transposable elements (TEs) are the largest component of plant genomes and their abundance is correlated with genome size. Among them, Gypsy-like LTR retrotransposons are often more abundant than other TEs, particularly in genomes with large sizes. Gypsy-like retrotransposons are located in the heterochromatic regions and nested with themselves and other TEs. The “copy and paste” transposition mechanism combined with the ability to insert into other TEs is likely responsible for their success in plant genomes. Recently we discovered a new type of DNA transposons, called GingerRoot, which is related to Gypsy-like retrotransposons and therefore shares target specificity with Gypsy-like retrotransposons. This element is present in the genomes of multiple plants and animals but its presence is largely correlated with low abundance of LTR retrotransposons. This suggests the competitive relationship between GingerRoot and Gypsy-like elements, and the fate of TEs is determined by their host genomes as well as their con-existing TEs.

### **W394: Evolution of Genome Size**

#### **Gene Expansion in the Tylenchomorpha: Detecting Genes Involved in Plant Host Rewiring**

**Rick Masonbrink**, Iowa State University, Ames, IA

Parasites survive and reproduce by harvesting nutrients from their hosts, thereby exerting a toll on fecundity and prompting host selection. Eventually the host develops resistance, and these same selective forces are mirrored in the parasite until resistance is overcome. These ongoing cycles of reciprocal adaptation between host and parasite have a genomic signature in the form of gene duplications, which are frequently concurrent with these evolutionary outcomes. We studied this phenomenon via 11 publicly available genomes in the Tylenchomorpha, a clade of economically damaging plant parasitic nematodes. Using gene orthology and phylogenetics we discern gene expansions from three phylogenetic nodes: the Tylenchomorpha, a basal node segregating migratory nematodes from sedentary root knot and root cyst nematodes; the Tylenchoidea root-knot and cyst nematodes that have massively different modes of infection with similar consequences; and the Heteroderidae, the divergence of cyst nematode genera in Heterodera and Globodera. We then further our understanding of secreted gene evolution in the development of a feeding site for Heterodera glycines, by leveraging published expression datasets characterizing pre-parasitic and parasitic stages with differing aspects of host compatibility. Altogether, using a particular emphasis on genes with conserved secretion signals in multiple species, we examined the expansion of gene families across these clades to discern conserved mechanisms of host cell manipulation and parasitism in *H. glycines* and provide a resource for developing new modes of nematode resistance.

### **W395: Evolution of Genome Size**

## **Chromosome Number and Success of Polyploid Lineages**

**John Bowers**, University of Georgia, Athens, GA

Polyploidy is ubiquitous and often recursive in plant lineages, occasionally associated with great evolutionary success, yet chromosome numbers have consistently returned to a narrow range. Some angiosperm species exist with, and some even flourish with, haploid chromosome numbers of 100 or more, but most angiosperm species have 13 pairs or less despite multiple polyploidy events. Higher chromosome numbers in angiosperms are disproportionately located on terminal branches in phylogenies, and are seldom the ancestral state of species rich clades. If polyploidy in plants is as advantageous as it appears to be based on the number of paleopolyploidy events, why haven't angiosperm chromosome numbers increased exponentially? Additionally, why do fern lineages appear to flourish with chromosome numbers much higher than commonly observed in angiosperms? A mechanism that could limit the long term success of angiosperms with high chromosome numbers will be presented.

## **W396: Exploring Phytobiomes**

### **Phytobiomes Research for Enhancing the Sustainable Production of Food, Feed, and Fiber**

International Alliance for Phytobiomes Research<sup>1</sup>, Gwyn A. Beattie<sup>2</sup>, Natalie W. Breakfield<sup>3</sup>, Isabelle Caugant<sup>4</sup>, **Kellye Eversole**<sup>4</sup>, Magalie Guilhabert<sup>5</sup>, Jan E. Leach<sup>6</sup>, Matthew J. Ryan<sup>7</sup> and Angela Sessitsch<sup>8</sup>, (1)International Alliance for Phytobiomes Research, Lee's Summit, MO, (2)Iowa State University, Ames, IA, (3)NewLeaf Symbiotics, Saint Louis, MO, (4)International Phytobiomes Alliance, Lee's Summit, MO, (5)Bayer CropScience, Biologics, West Sacramento, CA, (6)Colorado State University, Fort Collins, CO, (7)CABI, Egham, Surrey, United Kingdom, (8)AIT - Austrian Institute of Technology, Vienna, Austria

A major paradigm shift in agricultural production is required to meet the demands of a global world population projected to reach 9.7 billion in 2050. We need to sustainably increase crop productivity, while preserving biodiversity, natural resources, and grower income in the context of climate change. To optimize sustainable productivity and profitability on farms, grasslands, and forests, scientists must embrace a holistic, systems-level approach and focus on the complexity within phytobiomes. The term "phytobiome" refers to a plant growing within a specific environment, or biome, and all of the micro- and macro-organisms living in, on, or around it—such as microbes, animals, insects, and other plants—as well as the geophysical environment, which includes soil, air, water, weather, and climate. By establishing a foundation of knowledge on how phytobiome components interact and affect each other, the Phytobiomes Alliance ([www.phytobiomesalliance.org](http://www.phytobiomesalliance.org)) a non-profit alliance of industry, academic, and governmental partners created in 2016, aims at addressing today's agricultural challenges. The Alliance facilitates and coordinates international efforts toward expanding phytobiomes research in order to accelerate the sustainable production of food, feed, and fiber for food security. Current priority areas of the Alliance include filling the gaps in our knowledge of how microbes interact with other phytobiome components in outdoor and controlled environments as well as building a regulatory science foundation to support rapid commercialization of sustainable, microbial based products that increase the productivity and viability of agricultural production systems.

## **W397: Exploring Phytobiomes**

### **Wheat Virus Identification using Oxford Nanopore Technology**

**John Fellers**<sup>1</sup>, Christian Webb<sup>2</sup>, Madison Fellers<sup>3</sup>, Jessica Shoup-Rupp<sup>2</sup> and Erick DeWolf<sup>2</sup>, (1)USDA ARS, Manhattan, KS, (2)Kansas State University, Manhattan, KS, (3)USDA-ARS, Manhattan, KS

Viral diseases are a limiting factor to wheat production. Viruses are difficult to diagnose in the early stages of disease development and are often confused with nutrient deficiencies or other abiotic problems. Immunological methods are useful to identify viruses, but specific antibodies may not be available or require high virus titer for detection. In 2015 and 2017, wheat plants containing *Wheat streak mosaic virus* (WSMV) resistance gene, *Wsm2*, were found to have symptoms characteristic of WSMV.

Serologically, WSMV was detected in all four samples. Additionally, *High Plains wheat mosaic virus* (HPWMoV) was also detected in one of the samples. *Barley yellow dwarf virus* (BYDV) was not detected and a detection kit was not readily available for *Triticum mosaic virus* (TriMV). Initially, cDNA cloning and Sanger sequencing were used to determine the presence of WSMV; however, the process was time-consuming and expensive. Subsequently, cDNA from infected wheat tissue was sequenced with single-strand Oxford Nanopore sequencing technology (ONT). ONT was able to confirm the presence of WSMV. Additionally, TriMV was found in all of the samples and BYDV in three of the samples. Deep coverage sequencing of full length, single-strand WSMV revealed variation compared to the WSMV Sidney-81 reference strain and may represent a new variant which overcomes *Wsm2*. These results demonstrate that ONT can more accurately identify causal virus agents and has sufficient resolution to provide evidence of causal variants.

### **W398: Exploring Phytobiomes**

#### ***Trichoderma*-Derived Seed Treatments Stabilize Yields and Provide Drought Resilience to *Zea mays* by Modulating Host Gene Expression during Periods of Osmotic Stress.**

**Andrea Marino**, Advanced Biological Marketing, Geneva, NY

Endophytic fungi in the genus *Trichoderma* are utilized as a versatile agricultural input able to improve crop productivity and enhance tolerance to both biotic and abiotic stress. Recent publications suggest multiple mechanisms are at play including direct chemical signaling via fungal metabolite, alterations in rhizosphere diversity, and dynamic transcriptome reprogramming. To characterize responses to *Trichoderma*-based SabrEx seed treatments or *Trichoderma*-derived 1-Octen-3-ol (octenol) seed treatments on *Zea mays*, we conducted (1) yield trials, (2) a field study of rhizosphere diversity and maize foliar gene expression during the 2016 drought, and (3) a sterile soil study of maize root gene expression in response to osmotic stress (PEG). AgriGO's SEA tool was used to identify enriched GO terms. Both SabrEx and octenol increased grain yield, by an average of 6.7% and 4.6%, respectively. The two seed treatments reduced the variance of gene expression relative to base-treated controls, and induced similar gene expression responses to each other, particularly in biological processes involved in the response to abiotic stress. The results of this study suggest that SabrEx and octenol increase and stabilize yields by stabilizing gene expression, particularly during osmotic stress. SEA analyses of DEGs in the sterile soil study show drought-stressed plants treated with biorationale formulations respond uniquely to endogenous stimuli and phytohormones like ABA. Together, these indicate that growers facing worsening drought conditions have an improved opportunity for a return on investment when integrating biorationale seed treatments in their management practices.

### **W399: Exploring Phytobiomes**

#### **Pathogen Detection and the Role of Microbial Communities in the Phyllosphere during Fungal Infection of Wheat**

**Yiheng Hu**<sup>1</sup>, Gamran Green<sup>1</sup>, Andrew W. Milgate<sup>2</sup>, Laszlo Irinyi<sup>3</sup>, Wieland Meyer<sup>3</sup>, Eric Stone<sup>1</sup>, John P. Rathjen<sup>1</sup> and Benjamin Schwessinger<sup>1</sup>, (1)The Australian National University, Canberra, ACT, Australia, (2)NSW Dept of Primary Industries, Wagga Wagga, Australia, (3)The University of Sydney, Sydney, NSW, Australia

Fungal diseases of plants are responsible for major losses in agriculture, highlighting the need for rapid and accurate identification of fungal plant pathogens. Disease outcomes are influenced by diverse microbial communities known as microbiomes at sites of infection. We conducted the first use of metagenomics sequencing with a portable sequencer as a method for detecting fungal pathogens from wheat (*Triticum aestivum*) in a standard molecular biology laboratory. The data revealed that our method is robust and capable of diagnosing stripe rust (caused by *Puccinia striiformis* f. sp. *tritici*), septoria blotch (caused by *Zymoseptoria tritici*) and yellow spot (caused by *Pyrenophora tritici repentis*). We also identified the bacterial genus *Pseudomonas* co-present with *Puccinia* and *Zymoseptoria* but not *Pyrenophora* infections. One limitation of the method is the over-representation of redundant wheat genome sequences from samples and therefore a low coverage of pathogen species which affects the

identification accuracy. We are improving fungal identification by testing different classification strategies on a mock fungal community. Here I will present recent progress in our method development and the biological insight gained during this process. Our work outlines a new approach for detection of a broad range of plant pathogens and associated microbes using a portable sequencer, providing the basis for the development of a high-throughput, large scale and on-site plant disease monitoring system.

#### **W400: Exploring Phytobiomes**

##### **Variability and Plasticity of the Wild Chickpea Microbiome**

**Betsy Alford** and Douglas R Cook, University of California-Davis, Davis, CA

Microbiomes of domesticated crops have come under increasing study for their potential to provide services to plants such as protection from biotic and abiotic stressors and in turn stabilize yields. While significant effort has been made to improve crop microbiome health, little is known regarding the “ground state” of the plant population microbiome. Here, we examine the microbial community structure of the wild progenitor of chickpea, *Cicer reticulatum*, and a sister species, *C. echinospermum*, across its projected native range.

As part of this work, we surveyed 20 sites throughout southeastern Turkey for soil characteristics, plant genetics, and microbial community assortment. In these locations of long standing co evolution, we observe consistent taxonomic microbial guilds corresponding with soil chemistry and plant species. To disentangle these relationships, sites with distinct soil chemistries and plant populations were chosen for a reciprocal transplant experiment. While the wild chickpea roots consistently enriched for microbial guilds regardless of soil type, the exact taxonomic membership is correlated with plant population. Conversely, nodule guild membership does not exhibit a clear plant population signal demonstrating that this relationship may be driven by the availability of compatible microbes in the soil.

#### **W401: Exploring Phytobiomes**

##### **Mighty Microbes: The Tri-Trophic Interactions of Endophytic *Metarhizium* in Maize**

**Imtiaz Ahmad**<sup>1</sup>, María del Mar Jiménez-Gasco<sup>2</sup>, Dawn S. Luthe<sup>3</sup> and Mary E. Barbercheck<sup>1</sup>,  
(1)Department of Entomology, Pennsylvania State University, State College, PA, (2)Department of Plant Pathology and Environmental Microbiology, Pennsylvania State University, State College, PA,  
(3)Department of Plant Science, Pennsylvania State University, State College, PA

Many insect-pathogenic fungi, including *Metarhizium* (Hypocreales: Clavicipitaceae), are also endophytes that can benefit their host plant through plant disease antagonism, plant growth promotion and insect growth suppression. Our research focuses on the transcriptomic and physiological aspects of plant-*Metarhizium*-pest interactions. We isolated *Metarhizium robertsii* from agricultural soil and inoculated maize seeds (*Zea mays* L.) with the conidiospores of *M. robertsii* to establish endophytic colonization of plants. We evaluated maize at V4 growth stage for endophytic colonization of leaf and root tissue, plant height, above-ground biomass, and relative growth rate of second instar black cutworm, *Agrotis ipsilon*. We measured relative expression of selected defense genes in maize in response to endophytic colonization.

We recovered *M. robertsii* from 91% of maize grown from *M. robertsii*-treated seed and more frequently in root sections compared with leaf sections of endophytically colonized plants. Height and above-ground biomass of endophytically colonized plants were significantly greater compared to control plants. In insect feeding bioassays, the relative growth rate of 2<sup>nd</sup> instar black cutworm was lower when fed on maize leaves of endophytic plants compared to control plants. The expression of plant defense genes involved in jasmonic acid and salicylic acid pathways and chitinases differed in endophytically colonized plants compared to control plants. Endophytic colonization by *M. robertsii* had growth promotive effects on maize, growth suppressive effects on black cutworm larvae, and modulated the expression of key defense genes in maize. These effects suggest a potentially important role for endophytic *M. robertsii* in tri-trophic interactions and biological control for sustainable crop production.

## **W402: Extremophile Genomes**

### **Learning from Nature: Multi-Ion Salt Stress Adaptation Explored using Extremophyte Genomics Exemplified by Boron Toxicity**

**Guannan Wang** and Maheshi Dassanayake, Louisiana State University, Baton Rouge, LA

*Schrenkiella parvula* is a close relative of *Arabidopsis thaliana* and *Brassica* crops. It shows adaptations to multiple abiotic stresses that mark its lifestyle in its natural hypersaline habitat. *S. parvula* compared to *A. thaliana*, presents a unique system to identify naturally selected physiological traits, developmental responses, and genome reorganization underlying adaptations to salt stresses. To expand our understanding of plant adaptations to multiple salt stresses, we compared genomic, transcriptomic, metabolomic, and ionic responses of *A. thaliana* and *S. parvula* to excess Na<sup>+</sup>, K<sup>+</sup>, Li<sup>+</sup>, and borate salt-stresses. Each ion induced a set of stress responses that were largely shared among other salt stresses, and also included several specific responses. We observed that boron induced a largely unique set of responses quite distinct from other ions.

Upon boron toxicity stress, *A. thaliana* accumulates more boron as well as induces more differently expressed genes in response to boron stress compared to *S. parvula*. In *A. thaliana*, cell wall metabolism is significantly affected in shoots, while RNA metabolism is severely altered in roots, under boron stress. In support of this observation, we found an accumulation of excess boron in *A. thaliana* shoot cell walls, and metabolic adjustments associated with RNA metabolism in roots upon borate treatments. Clusters of co-expressed orthologs revealed a signature of “stress readiness” in the *S. parvula* transcriptome, where the *S. parvula* orthologs showed basal expression levels comparable to the stress-treated expression of *A. thaliana* orthologs. These “stress-readiness” clusters in roots indicated that the expression of orthologs in RNA metabolism, significantly affected in *A. thaliana*, appeared pre-adapted in *S. parvula*. *S. parvula* showed higher basal-level expression of orthologs for boron toxicity-induced genes in *Arabidopsis* shoot, with mostly unknown functions.

A stress-readiness response was observed in *S. parvula* when treated with excess Na<sup>+</sup>, K<sup>+</sup>, Li<sup>+</sup> salt-stresses compared to that of *A. thaliana*. The rate of accumulation of toxic levels of ions in *S. parvula* was much lower than in *A. thaliana*, and the subsequent response of growth retardation was absent or minimal in *S. parvula* compared to *A. thaliana*. The *A. thaliana* transcriptomes showed an overall greater response compared to that of *S. parvula* under multi-ion salt stresses. Consistently, less metabolites were identified responsive to each stress in *S. parvula* than in *A. thaliana*. *S. parvula* contained a seemingly stress-ready transcriptome, which may have led to a reduced ion accumulation that facilitated continuous growth making it tolerant to all the salt stresses tested. In conclusion, the extremophyte genome of *S. parvula* acts as a repository of genetic changes that have enabled its successful niche adaptation to a multi-ion saline environment. Our comparative analysis with the closely related stress sensitive species, *A. thaliana* facilitated the identification of many of the genes that may have allowed abiotic stress adaptations in *S. parvula*.

## **W403: Extremophile Genomes**

### **New Insights into Cryptobiotic Survival of Tardigrades**

**Nadja Mobjerg**, University of Copenhagen, Copenhagen, Denmark

Tardigrades are microscopic invertebrate animals found worldwide in permanent and temporal aquatic environments. They are renowned for their extreme stress tolerance and their ability to enter an ametabolic state of latent life, known as cryptobiosis. These minute animals generally depend on successfully entering cryptobiosis to survive extreme conditions. The latter is much unlike extremophile bacteria and archaea that are adapted to life in extreme environments with optimal growth depending on protein adaptations that match the given environment. In sharp contrast to e.g. thermophilic archaea, active state tardigrades are sensitive to high temperatures.

Tardigrades require to be surrounded by water to be in their active state. Species living on land in temporary aquatic microhabitats, such as moss cushions, endure periods of complete desiccation by entering the highly resilient cryptobiotic "tun" state. Maintaining structural integrity, while in this ametabolic tun state, is essential for resumption of life following cryptobiosis. Increased focus is currently on revealing the molecular mechanisms involved in sustaining the tun state and securing resumption of post-cryptobiotic life.

During dehydration tardigrades contract and reduce their body volume significantly, while reorganizing internal organs and withdrawing their legs, thereby entering the tun state. Muscle protein filaments seem to play a vital role not only for the structural reorganization during tun formation, but also for stabilizing the three-dimensional structure of the ametabolic and ATP depleted tun state. It has furthermore been suggested that tardigrade-specific intrinsically disordered proteins may provide structural stabilization during desiccation through vitrification. A highly effective antioxidant defense apparatus as well as high fidelity DNA repair systems are likely crucial for post-cryptobiotic survival. Accordingly, comparative genomics and transcriptomics have shown that all tardigrades seem to have a comprehensive number of genes encoding proteins involved in antioxidant defense. It was recently shown that a tardigrade damage suppressor protein (Dsup) binds to nucleosomes protecting chromosomal DNA from reactive hydroxyl radicals. Thus, while the secrets of latent life in many aspects still remain a mystery, recent research is beginning to unravel some of its underlying mechanisms.

#### **W404: Extremophile Genomes**

##### **The Genome of Broomcorn Millet, a Cereal with Extremely High Water Use Efficiency**

**Zhang Heng**, Shanghai Center for Plant Stress Biology, CAS, Shanghai, China

Broomcorn millet (*Panicum miliaceum* L.) is the most water-efficient cereal and one of the earliest domesticated plants. Here we report its high-quality, chromosome-scale genome assembly using a combination of short-read sequencing, single-molecule real-time sequencing, Hi-C, and a high-density genetic map. Phylogenetic analyses reveal two sets of homologous chromosomes that may have merged ~5.6 million years ago, both of which exhibit strong synteny with other grass species. Broomcorn millet contains 55,930 protein-coding genes and 339 microRNA genes. We find Paniceae-specific expansion in several subfamilies of the BTB (broad complex/tramtrack/bric-a-brac) subunit of ubiquitin E3 ligases, suggesting enhanced regulation of protein dynamics may have contributed to the evolution of broomcorn millet. In addition, we identify the coexistence of all three C<sub>4</sub> subtypes of carbon fixation candidate genes. The genome sequence is a valuable resource for breeders and will provide the foundation for studying the exceptional stress tolerance as well as C<sub>4</sub> biology.

#### **W405: Farm Animal Genome Editing**

##### **From Barn to Bedside - Genome Editing in Livestock for Food and Biomedicine**

**Daniel F. Carlson**, Recombinetics, Inc., Saint Paul, MN

#### **W406: Farm Animal Genome Editing**

##### **Opportunities for Gene Editing in Poultry**

**Rachel J. Hawken**, Cobb-Vantress, Siloam Springs, AR

Efforts are underway to investigate the utility of gene editing in poultry. If successful there are still many political, regulation and social acceptance obstacles to clear before any product will be available to consumers. This presentation will outline our efforts, and the efforts that our industry can make to ensure successful utilization of this technology to solve global production and welfare issues.

#### **W407: Farm Animal Genome Editing**

##### **Gene Editing for Healthier Animals and Sustainable Food System**

Elena Rice, Genus plc, DeForest, WI

#### **W408: Farm Animal Genome Editing**

##### **Modelling Feasibility of Eradicating PRRS with Genome Editing**

Jaap Buntjer<sup>1</sup>, Gertje Petersen<sup>2</sup>, Fiona Hely<sup>3</sup>, Tim Byrne<sup>3</sup>, Bruce Whitelaw<sup>1</sup> and **Andrea Doeschl-Wilson<sup>4</sup>**, (1)The Roslin Institute, Edinburgh, United Kingdom, (2)AbacusBio, Dunedin, New Zealand, (3)AbacusBio, New Zealand, (4)The Roslin Institute, University of Edinburgh, UK, Edinburgh, United Kingdom

Novel genomic technologies such as gene editing offer exciting new opportunities for infectious disease control, especially in situations where conventional approaches have failed. Recent breakthroughs in gene editing of pigs, in which a simple disruption of the CD163 gene confers complete resistance to infection with the Porcine Reproductive and Respiratory Syndrome Virus (PRRSV), exemplify this opportunity. Despite huge global efforts, PRRS continues to persist and remains one of the costliest infectious disease for the pig industry worldwide. Compared to all existing control methods, gene editing is the only method known to date that can fully protect pigs from becoming infected with PRRSV and thus also from transmitting the disease. This raises the question as to how the introduction of gene edited pigs in national pig herds would reduce PRRS prevalence, and in particular, whether the dispersion of gene edited pigs in commercial pig herds could eventually help to eradicate this devastating disease on a national level.

We have developed a genetic-epidemiological model to predict how gene editing may affect future risks of PRRS outbreaks in national pig herds. A gene flow model was used to simulate the propagation of genetically resistant pigs through the breeding pyramid into commercial pig herds. Herds were assumed to differ in size, the proportion of genetically resistant pigs, whether or not other control measures (e.g. vaccines) were applied, as well as in PRRS exposure probability. Different distribution scenarios were simulated, ranging from a centrally regulated distribution of genetically resistant pigs across herds to a voluntary uptake regime. The model predicts how the total number of genetically resistant pigs in the national population, their distribution across herds together with the above non-genetic factors, affect PRRS prevalence. The results suggest that PRRS eradication through gene editing alone would be extremely difficult as it would require large proportions of gene edited pigs. In contrast, when used complemented by conventional control option such as vaccination, introduction of genetically resistant pigs into commercial pig populations could achieve the hitherto unprecedented opportunity of PRRS eradication on a national scale within less than a decade. The results of this proof of concept modelling study provide some first insights for all stakeholders into the potential benefits, feasibility and cost-effectiveness of gene editing as disease control strategy in domestic livestock, and may contribute towards public acceptance of these highly contested innovative technologies.

#### **W409: Farm Animal Genome Editing**

##### **Genome Wide CRISPR Knockout Screen Identifies Host Factors Important for Bovine Herpes Virus Type 1 Replication**

**Wenfang Spring Tan<sup>1</sup>**, Enguang Rong<sup>2</sup>, Inga Dry<sup>1</sup>, Simon Lillico<sup>2</sup>, Andy Law<sup>2</sup> and Bruce Whitelaw<sup>3</sup>, (1)the Roslin Institute, Midlothian, United Kingdom, (2)the Roslin Institute, United Kingdom, (3)The Roslin Institute, Edinburgh, United Kingdom

CRISPR/Cas9 are molecular scissors that cut DNA in a site-specific manner; it relies on base pairing between the small CRISPR guide RNA (gRNA) and target DNA. This activity leads the gRNA-bound Cas9 protein to the target and exert cutting, creating double strand breaks in the DNA that are repaired by the cell, often resulting in gene inactivating *indels*. Since the specificity of CRISPR/Cas9 is largely determined by the base pairing, it is straightforward to introduce many guides with altered gRNA sequences to achieve knockout of many genes in parallel. To enable whole genome screening in cattle, we utilized this high throughput strategy and produced a CRISPR library containing a pool of 96,000 guides targeting 21,165 protein coding genes in the cow genome. We observed good performance after

transducing this library, one guide per cell, into our MDBK cell line that stably expresses Cas9 from the *rosa26* locus: guides targeting core essential genes are significantly depleted while those targeting non-essential genes or non-cutting control guides remain largely unchanged.

Bovine Herpes Virus Type 1 (BHV-1) causes infectious bovine rhinotracheitis, fatalities in calves and pregnancy abortions in cows, leading to huge economy loss to cattle farmers in Ireland and the UK. Unfortunately, little is known about how host cell factors intervene or facilitate BHV-1 infection, and this lack of knowledge impedes vaccine and drug developments. Thus, to study interactions between BHV-1 and the host, we infected the library transduced cells with a GFP tagged BHV-1 at high MOI, and FACS sorted live cells at 8 hours post infection into sub-populations with different intensities of GFP signals. We identified lists of genes with significantly depleted or enriched guides from these sub-populations, the GFP negative cells in particular. In GFP negative cells, genes targeted by enriched guides are candidates that facilitate or are essential for BHV-1 infection whereas genes targeted by depleted guides may inhibit this virus. We are currently validating candidate genes using a focused library that targets only the candidates from the genome wide screen. We are also using drug inhibitors and individual gRNAs that knockout, activate or repress gene expressions to study their detailed mechanisms.

## **W410: Farm Animal Genome Editing**

### **Type I Interferon Receptor Knockout Sheep: A Model for Viral Immunology and Reproductive Signaling**

**Christopher J. Davies**<sup>1,2</sup>, Kira P. Morgado<sup>1,2</sup>, Misha Regouski<sup>1</sup>, Zhiqiang Fan<sup>1</sup>, Ying Liu<sup>1</sup>, Aaron J. Thomas<sup>1,2</sup>, Sang-Im Yun<sup>1</sup>, Byung-Hak Song<sup>1</sup>, Arnaud Van Wettere<sup>1</sup>, Young-Min Lee<sup>1</sup> and Irina Polejaeva<sup>1</sup>, (1)Department of Animal, Dairy & Veterinary Sciences, Utah State University, (2)Center for Integrated BioSystems, Utah State University

Type I interferons are important for initiation of immune responses to viral infection. Their effects are mediated by the type I interferon (IFN) receptor, IFNAR, which is a heterodimer of two subunits: IFNAR1 and IFNAR2. One or both chains of the sheep IFNAR were knocked out using CRISPR/Cas9. Quantitative RT-PCR analysis of IFN-stimulated gene expression showed that fibroblasts from homozygous knockout sheep fail to respond to IFN- $\alpha$ . Furthermore, fibroblast cells from an *IFNAR2* knockout fetus supported a higher level of Zika virus (ZIKV) replication than wild-type fetal fibroblast cells. In a preliminary infection trial, an *IFNAR2* knockout ram infected with ZIKV at about 11 months of age died from bloat 16 days after infection, while a wild-type control sheep remained healthy. There is, however, no definitive evidence that ZIKV was involved in the pathogenesis of the bloat. In a second infection trial, two one-week old IFNAR knockout lambs and two wild-type lambs were infected with ZIKV. One wild-type lamb was euthanized on day 13 because of a catheter infection and one IFNAR knockout lamb was euthanized on day 15 because she was too weak to stand up. Both knockout lambs had viral abomasitis, enteritis and lymphadenitis at necropsy. Tissues from these lambs tested negative for bovine viral diarrhea virus, bluetongue virus, rotavirus and coronavirus. Testing for ZIKV is ongoing. While several lambs have died at various ages from viral infection, other IFNAR knockout sheep are currently over two-years old and in good health. Presumably these sheep are compensating for a lack of type I IFN signaling by the presence of intact type II (IFN- $\gamma$ ) and type III (IFN- $\lambda$ ) IFN pathways. These initial studies suggest that IFNAR knockout sheep will be an effective large animal model for studying the pathogenesis of viral infection. Furthermore, ruminants utilize a novel type I interferon, interferon tau (IFN- $\tau$ ), for pregnancy recognition. Consequently, IFNAR knockout sheep are a valuable model for studying pregnancy recognition. Early in ovine gestation (days 12-15) placental trophoblast cells secrete IFN- $\tau$ , which interacts with IFNAR on endometrial epithelial cells signaling the mother to maintain her corpus luteum and not return to estrus. If the established paradigm is correct, IFNAR knockout ewes should be infertile. Thus far, only one IFNAR knockout ewe has been bred and this ewe failed to become pregnant. A larger breeding trial will be conducted to test whether IFNAR knockout sheep are infertile. If IFNAR knockout sheep are infertile, these sheep will be a unique model for studying pregnancy establishment in ruminants. A breeding herd of heterozygous *IFNAR2* knockout sheep is being developed for production of homozygous *IFNAR2* knockout sheep for both infection and reproduction studies.

**Acknowledgements:** These studies were supported by grants from: the Utah Science, Technology and Research Initiative (USTAR); the Utah Agricultural Experiment Station, which receives funding from USDA, NIFA and the State of Utah; and the Utah State University Center for Integrated BioSystems.

#### **W411: Farm Animal Genome Editing**

##### **The Knockout of the HMG Domain of the Porcine SRY Gene by CRISPR/Cas RNP Microinjection Causes Sex Reversal in Gene-Edited Pigs**

**Stefanie Kurtz**<sup>1</sup>, Andrea Lucas-Hahn<sup>1</sup>, Heiner Niemann<sup>2</sup> and Björn Petersen<sup>1</sup>, (1)Institute for Farm Animal Genetics, Friedrich-Loeffler-Institute, Neustadt am Rübenberge, Germany, (2)Twincore, Hannover Medical School, Hannover, Germany

Sexing by Gene Editing in pigs is an alluring alternative to the surgical castration of piglets. Current alternatives like boar fattening, immunocastration and sperm sexing are often economically not well applicable or lack consumer acceptance.

In mice, the SRY-gene was first described as a genetic developmental switch for the male phenotype. The knockout of the murine SRY-gene by TALEN suppressed testis development in the fetal gonadal ridges and generated a female phenotype. In addition, the knockout of the 5' flanking region of rabbit SRY gene results in an equal phenotype as in mice. In our study, we aim to generate a knockout of the porcine SRY gene to investigate its function in pigs for generating solely female offspring.

For the first time, we successfully generated a knockout of the SRY gene in pigs using microinjection of two Clustered regularly interspaced short palindromic repeats (CRISPR) – associated protein - 9 nuclease (Cas9) ribonucleoprotein (RNP) complexes targeting the centrally located “high mobility group” domain (HMG) of the SRY gene. Mutation within the porcine HMG domain resulted in the development of complete external and internal female genitalia in genetically male piglets. Moreover, we could further confirm the potential function of the HMG box as the main functional domain for male sex development, by introducing a deletion within the 5' flanking region of the HMG domain, which was not associated with sex reversal in generated offspring. In summary, this study demonstrates the main role of the HMG box of the SRY gene triggering the male sex determination in pigs.

#### **W412: Farm Animal Genome Editing**

##### **Public Attitudes Towards Genome Edited Meat**

**Bruce Whitelaw**<sup>1</sup>, Tom Kirk<sup>2</sup>, John Crowley<sup>2</sup>, Daniel Collado<sup>3,4</sup> and Tim Byrne<sup>5</sup>, (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)AbacusBio International, Midlothian, United Kingdom, (3)Instituto Agroalimentario de Aragón, Zaragoza, Spain, (4)Departamento de Producción y Sanidad Animal, Zaragoza, Spain, (5)AbacusBio International, Midlothian, United Kingdom

Given the potential positive impact of gene-editing on food production, it is important that researchers and potential users are aware of public attitudes towards use of the technology in the human food chain. To increase awareness of consumers' perceptions and attitudes towards gene-edited food we performed an online survey with a focus on meat products. Consumption preferences were analysed based on consumers' willingness to pay (WTP) for gene-edited meat products relative to “normal” meat products. This WTP was also assessed in the context of added benefits – improvements in environmental impact, human health, and animal health/ welfare – relative to the non-edited meat products. Analysis showed that consumer attitudes towards gene-editing are consistently created towards the “whole package” of genetically modified (GM) foods, along with gene-editing and all its components (safety, environment and ethical), without distinguishing between them. Our analysis captured these attitudes in one component factor, which comprises each respondent's attitude towards both gene-edited and GM foods. Our analysis also found that WTP for gene-edited meat products was negative; consumers would pay less for gene-edited meat products than for normal meat products. However, when WTP for gene-edited meat

products was assessed in the context of added benefits relative to normal meat products, WTP became positive.

### **W413: Farmed Insects to Feed Future Populations**

#### **Workshop Introduction**

**Aaron T. Dossey**, All Things Bugs LLC, Oklahoma City, OK

### **W414: Farmed Insects to Feed Future Populations**

#### **Genome Assembly and Annotation of the Cricket *Gryllus bimaculatus***

**Guillem Ylla**<sup>1</sup>, Taro Nakamura<sup>2</sup>, Sumihare Noji<sup>3</sup>, Taro Mito<sup>3</sup> and Cassandra Extavour<sup>4</sup>, (1)Harvard University, Cambridge, MA, (2)National Institute for Basic Biology, Japan, (3)Tokushima University, Japan, (4)Harvard University

With more than a million species described, insects ecologically dominate most terrestrial environments and play essential roles in ecosystem stability. Additionally, insects are important models for genetic, ecological and developmental studies, as well as a growing source of food for livestock and human consumption. Nevertheless, most of our current understanding of insects comes from study of a limited number of holometabolan species such as flies, bees or butterflies. Hence, in order to expand our knowledge of insect biology and evolution, an increasing number of laboratories are adopting new model species for study, including hemimetabolan insects such as crickets, locusts, and cockroaches. The number of available tools for genetic manipulation of these non-traditional model organisms remain limited, especially when an annotated reference genome is unavailable. To that end, here we report sequencing and annotation of the genome of the two-spotted cricket *Gryllus bimaculatus* which is composed of 1.6 gigabases (Gb) and contains ~18,000 protein-coding genes. Additionally, we annotate the recently assembled genome of another cricket species, the Hawaiian cricket *Laupala kohalensis* which is of similar size (1.5Gb) and contains a predicted similar number of coding genes. We discuss the potential of these new annotations to contribute to establishing these crickets as model organisms for genetic studies, and to what extent the comparison of these three orthopteran genomes contributes to our understanding of insect evolution.

### **W415: Farmed Insects to Feed Future Populations**

#### **Reference Genomes for Two Cricket Species**

**Brenda Oppert**, USDA ARS CGAHR, Manhattan, KS

As the world approaches the population mark of 8.6 billion in 2030 that has been predicted by the United Nations, food shortages due to increasing political and social conflicts that are exacerbated by climate change are pressing for new technologies to feed the world. We need to have new models for sustainable protein production for both animal feed and humans. Farmed insects have many advantages over traditionally farmed animals, requiring up to 50% less water and land per pound of protein with higher reproductive rates. Most insect species studied to date have protein levels comparable to that obtained from beef and milk. Agrobiodiversity is becoming more limited, with  $\frac{3}{4}$  of our food obtained from just 12 plant and five animal species, but insects offer thousands of potential species to be explored as a potential food resource. As insect farms are increasing worldwide to meet these challenges, genetic tools are needed to improve insect traits for specific feed applications. To begin to address these needs, we developed reference genome assemblies for several of the more commonly farmed insects including: the house cricket, *Acheta domesticus*, and banded cricket, *Gryllodes sigillatus*. Assemblies were achieved using data generated from high quality long genomic DNA extracted from a single male pupa or adult and sequenced on PacBio (RSII or Sequel I) with Dovetail Chicago and Hi-C scaffolding. Both assemblies are 2.4 Gb in length with  $L_{50}$  of 5, and  $N_{50}$  values of 222 Mb for *A. domesticus* and 205 Mb for *G. sigillatus* with BUSCO scores of 95 and 98%, respectively. In addition, we have life stage data from a transcriptome study of *A. domesticus* embryos, one-day hatchlings, three nymphal stages, and male and female adults. In that study, we identified sequences from all life stages and sexes with identity to

*Gregarina niphandrodes* that suggest that these crickets carry the parasite. Data from the genome and transcriptome assemblies are being used in genetic transformation to improve the nutritional content, decrease allergenicity, and improve disease-resistance of farmed crickets. The genome assembly of *A. domesticus*, which is sensitive to the devastating densovirus AdDNV, is being compared to the genome assembly of the virus-resistant species, *G. sigillatus*, in an effort to identify species-specific sequences that may be involved in resistance to AdDNV. Using the knowledge gained in using long read and long range sequencing and scaffolding for reference quality genomes, sequencing will be expanded to other insect species to provide additional genetic resources for food production. This research is a collaboration between All Things Bugs LLC (ATB) and USDA ARS, funded by the Defense Advanced Research Projects Agency (DARPA) under Contract No. 140D6318C0055 to ATB.

#### **W416: Farmed Insects to Feed Future Populations**

##### **Developing Insects for Food, Feed and Bioproduction: White-Eye Marker for CRISPR/Cas9 in Crickets and Mealworms**

**Fu-Chyun Chu** and Aaron T. Dossey, All Things Bugs LLC, Oklahoma City, OK

The idea of insects as sustainable food ingredients has been the focus of a new emerging industry in recent years. As with any crop, insects can be engineered via gene editing and other means to provide more desirable crop phenotypes. Our research is focus on the two most commonly farmed edible insect species: house cricket (*Acheta domesticus*) and yellow mealworm (*Tenebrio molitor*). Our goal is to develop a genetic engineering system to increase the value of these insects as food crops, such as nutritional content, disease resistance or reduced allergenicity. In this work, we are successfully using CRISPR/Cas9 technology to create both Knock-in and Knock-out phenotypes in both insect species through embryo microinjections. To build this genetic engineering system, we tested two different marker systems in both species. The first marker is a recessive phenotype of the insect's eye color. We selected the eye color related genes *Ad vermillion* (*AdV*) and *Tm vermillion* (*TmV*), respectively, as the sgRNA targeting sites, and was able to generate successful eye color knocks outs by CRISPR resulting in white eye color phenotype in insects. The second marker used the CRISPR system to knock-in green fluorescent protein (EGFP) driven by species-specific promoters. We used the *Polyubiquitin* promoter from *A. domesticus* and the *muscle actin* promoter from *T. molitor*. Overall, our result shows 96% of the G<sub>0</sub> *A. domesticus* crosses produced white eye offspring and 73% of the G<sub>0</sub> crosses produced EGFP positive G<sub>1</sub> offspring. Around 39% of the G<sub>0</sub> *T. molitor* crosses provided white eye G<sub>1</sub> and G<sub>2</sub> offspring and 16% of the G<sub>0</sub> crosses produced EGFP positive G<sub>1</sub> offspring. The eye color Knock-out phenotype in *T. molitor* resulted as white color in all life stages. However, knock out *vermillion* gene don't totally knock out the eye color in *A. domesticus*. The eye color phenotype is vermillion color throughout the life stages in cricket. The EGFP marker gene in *A. domesticus* shows a central expression in its thorax in most EGFP positive G<sub>1</sub> individuals. Due to promoter differences, in *T. molitor*, EGFP expression is in all the muscle in all life stages. To further test this knock-in system in mealworm, we created a new construct containing EGFP and a proprietary protein separated by a T2A self-cleaving peptide driven by *T. molitor* muscle action promoter. We knocked-in construct in the *vermillion* location and found 17% of the G<sub>0</sub> crosses provide EGFP positive G<sub>1</sub>s. The EGFP expression is the same pattern as muscle expression, however, the brightness of EGFP is lower compared to EGFP only knock-ins. We plan to use mass spectrometry to verify the T2A peptide is functional in this species and system. By showing these two marker system in *A. domesticus* and *T. molitor*, we demonstrate the potential to engineer insects for beneficial phenotypes.

#### **W417: Farmed Insects to Feed Future Populations**

##### **Techniques for Quantitative and Functional Genetic Analyses in the Cricket *Gryllus bimaculatus***

**Arpita Kulkarni**, Harvard University, Cambridge, MA

#### **W418: Feline and Canine Genomics Workshop**

##### **Genetics of Severe Ringworm in Persian Cats**

**Alexandra N. Myers**, Aline Rodrigues Hoffmann and William J. Murphy, Texas A&M University, College Station, TX

Ringworm (also known as dermatophytosis) is a fungal disease affecting humans and animals worldwide. Persian cats exhibit extensive and severe forms of the disease, including an extreme form known as dermatophytic pseudomycetoma not seen in other long-haired breeds. The mode of inheritance in Persian cats has not been elucidated, but in humans, autosomal recessive point mutations in the *CARD9* gene cause extensive and deep dermatophytosis. Whole genome sequencing was performed to identify particular variants associated with extensive/deep dermatophytosis in Persian cats. A genome-wide association study (GWAS) was performed on a cohort of 34 Persians, including groups of resistant and severely affected cats. In addition, variant filtration using WGS data from 160 non-Persian cats and 8 breed-matched control cats was performed to identify variants private to affected cats. Lastly, structural variant analysis was performed with Delly and Manta. The GWAS identified two regions with SNPs reaching Bonferroni significance on chromosomes E1 and F1. The majority of affected cats share a run of homozygosity on F1 overlying the S100 family of genes. Six non-coding indels are found in the disease-associated haplotype and are clustered near *S100A15*. S100 proteins play a well-described role in fungal immunity, have been shown to inhibit dermatophyte growth *in vitro*, and are strongly expressed in the skin of cats, making this region a strong candidate for additional study.

#### **W419: Feline and Canine Genomics Workshop**

##### **Ultracontiguous Genome Assemblies from Felid F1 Interspecies Hybrids**

Kevin R Bredemeyer, Texas A&M University, COLLEGE STATION, TX, Gang Li, Shaanxi Normal University, China, Brian W. Davis, Texas A&M, College Station, TX, Leslie A. Lyons, University of Missouri-Columbia, Columbia, MO, Wesley Warren, Animal Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO and **William J. Murphy**, Texas A&M University, College Station, TX

Here we describe several new long-read felid genome assemblies derived from applying trio-binning to the parent species and F1 offspring of three divergent interspecies hybrids: Bengal cat (Asian leopard cat x domestic cat), Safari cat (Geoffroy's cat x domestic cat) and Liger (lion x tiger). Within these new assemblies, single contigs span the largest chromosome arms, entire metacentric chromosomes, and the centromeres of several chromosomes. Our assemblies for the first time resolve the structure of large multicopy gene families, massive tandem macrosatellite repeats involved in X chromosome inactivation, and reveal candidate copy number variation that may underpin disease resistance and reproductive isolation. These new genome assemblies will provide a powerful resource for felid disease and trait mapping and comparative genomics.

#### **W420: Feline and Canine Genomics Workshop**

##### **Human Gene Constraint Guided Disease Variant Identification in Cats**

**Reuben M. Buckley**, University of Missouri-Columbia, Columbia, MO

Distinguishing pathogenic variants from benign variants remains an important challenge in genomic medicine of companion animals. In humans, large aggregated genome variation datasets, such as gnomAD, have revealed human genes and gene regions with highly constrained genetic variation, suggesting functional variants in these genes are more likely to be pathogenic. Development of similar resources in companion animals is currently not feasible. Instead, shared ancestry between cats and humans is used to map human constraint metrics to the cat genome for identifying potential feline disease causing variants. Two available constraint metrics are gene probabilities for loss of function intolerance (pLI) and constrained coding region (CCR) percentile. Genes with a high pLI are depleted of loss of function (LoF) variants, indicating they are under selective constraint in humans and likely essential for healthy survival and development. Alternatively, CCR percentile ranks gene coding regions located between adjacent nonsynonymous variants according to their length and mutability. Constrained regions identified by either metric are enriched for human disease causing variants. Results showed feline nonsynonymous variants were significantly depleted from genes and gene regions under

constraint in humans, indicating human constrained regions are likely constrained in cats too. Additionally, an analysis of feline hydrocephalus using human constraint metrics showed the causative variant was easily identified. Finally, constraint metrics were also useful for correcting genome assembly and annotation issues in *Felis\_catus\_9.0*. Ultimately, successful implementation of constraint metrics for candidate variant identification will expand results beyond genes with known functions, increasing the potential for new disease gene discovery.

#### **W421: Feline and Canine Genomics Workshop**

##### **Dire Wolves were the Last of an Ancient New World Canid Lineage**

**Laurent Frantz**, Queen Mary University of London, London, United Kingdom

The dire wolf (*Canis dirus*) was a large (~68 kg) wolf-like canid and is considered to be the most common and widespread large carnivore in Pleistocene America. Dire wolf remains are present in the North American paleontological record from at least ~250,000 to around 13,000 years ago, at the end of the Pleistocene, particularly in the lower latitudes. Other canid species present in Late Pleistocene North America include the slightly smaller gray wolf (*C. lupus*), and much smaller coyote (*C. latrans*), though dire wolves were more common overall. The reason for the dire wolf extinction, while other canid species survived the transition between Pleistocene and Holocene, however, remains uncertain. To resolve the evolutionary history of the dire wolf, we examined 56 subfossil specimens for the presence of preserved genomic DNA and identified five samples ranging that possessed sufficient endogenous DNA to obtain both mitochondrial and nuclear genomes. Our results indicate that while morphologically very similar to extant gray wolf, dire wolves represent a highly divergent lineage that separated over 4 million years ago. In contrast to numerous examples of hybridization across Canidae, there is no evidence for gene flow between dire wolves and either North American gray wolves or coyotes, suggesting the dire wolf evolved in isolation from the Pleistocene ancestors of these species. Our results support an early New World origin of the dire wolf, while the ancestors of the gray wolf, coyote, and dhole evolved in Eurasia and only colonized North America relatively recently.

#### **W422: Feline and Canine Genomics Workshop**

##### **Breed Trends, Health Risk Variant Frequency, and Outcomes in over 100,000 Companion Dogs Genotyped on a Direct-to-Consumer High-Density SNP Microarray.**

**Adam R. Boyko**, Department of Biomedical Sciences Cornell University, Ithaca, NY

#### **W423: Feline and Canine Genomics Workshop**

##### **Breed Trends, Health Risk Variant Frequency, and Outcomes in over 100,000 Companion Dogs Genotyped on a Direct-to-Consumer High-Density SNP Microarray.**

**Adam R. Boyko**, Department of Biomedical Sciences Cornell University, Ithaca, NY

#### **W424: Feline and Canine Genomics Workshop**

##### **High Quality Reference Genomes from Roadkill Enable Species Delimitation in Aardwolf and Bat-Eared Fox**

Rémi Allio<sup>1</sup>, Marie-Ka Tilak<sup>1</sup>, Céline Scornavacca<sup>1</sup>, Nico L. Avenant<sup>2</sup>, Erwan Corre<sup>3</sup>, Benoit Nabholz<sup>1</sup> and **Frédéric Delsuc<sup>1</sup>**, (1)Institut des Sciences de l'Evolution de Montpellier (ISEM), CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France, (2)National Museum Bloemfontein, Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa, (3)CNRS, Sorbonne Université, FR2424, ABiMS, Station Biologique de Roscoff, Roscoff, France

With thousands of fatalities due to car collisions with wildlife reported each year, roadkill are an underexploited resource in genomics. Here we show that mammalian roadkill samples could be used as a suitable source of DNA for long-read sequencing using the Oxford Nanopore Technology MiniON

device for two carnivoran species frequently encountered along South African roads: the aardwolf (*Proteles cristatus*) and the bat-eared fox (*Otocyon megalotis*). For both species, hybrid assembly of short Illumina reads at ~85X coverage and MiniON long reads at ~12X coverage using the MaSuRCA assembler provided genomes with high contiguity (<10,000 contigs) and completeness (>90% of complete BUSCOs). We further demonstrate that >90% of the 14,509 single-copy orthologous genes of the OrthoMaM database could be successfully retrieved from these assemblies. These figures compare favorably with current mammalian genome assemblies and set our genomes among the best carnivore genomes currently available. This cost-effective strategy to obtain high quality reference mammalian genomes opens the way for large-scale population genomic studies of mammalian wildlife using resequencing of samples collected from roadkill. We illustrate the potential of the approach for genome-scale species delimitation in both species for which subspecies have been defined based on disjunct distributions and morphological differences.

#### **W425: Feline and Canine Genomics Workshop**

##### **Genomic Analysis of the Extinct American Cheetah Reveals a New Evolutionary History of Carnivores in Pleistocene Beringia**

**Molly R Cassatt-Johnstone**<sup>1</sup>, Nedda F Saremi<sup>1</sup>, Grant D. Zazula<sup>2</sup>, Richard E. Green<sup>1</sup> and Beth Shapiro<sup>1</sup>, (1)University of California, Santa Cruz, Santa Cruz, CA, (2)Government of Yukon, Department of Tourism and Culture, Whitehorse, YT, Canada

Analyses of DNA preserved in the remains of plants and animals can dramatically rewrite what we believe to be a species' evolutionary history. Recently, we recovered DNA from two felid bones from Pleistocene age fossil localities in Yukon Territory, northwest Canada, both of which were identified based on morphology as *Puma concolor* (mountain lion or puma). Radiocarbon dates on these bones estimate that the two felids lived around 28,000 years ago during a peak cold period of the last glacial interval. We assembled complete mitochondrial genomes from both and placed them into a phylogeny with other related felids. Surprisingly, this analyses revealed the animals instead belonged to the extinct American cheetah, *Miracinonyx trumani*, which is a sister lineage to the puma but was believed to be constrained ecologically to temperate ecosystems. To further explore the evolutionary relationships and adaptations of the American cheetah, we used a new and highly sensitive approach to isolating ancient DNA and generated sufficient data to assemble a complete nuclear genome for one specimen with an average 44x depth of coverage. Using this genome, we reconstructed the recent evolutionary history of the American cheetah, comparing its inferred demographic and selective history to that of other living felids, including the puma. Our data expand significantly the known range of the American cheetah and provide a new perspective on the ecology of carnivores in Pleistocene Beringia.

#### **W426: Feline and Canine Genomics Workshop**

##### **DNA Capture: Challenges and Utility for Non-Invasive Samples**

**Safia Janjua**, Wright State University, Dayton, OH

Our capacity to address important ecological, conservation, and evolutionary questions increased rapidly with the advances in DNA sequencing and genotyping technologies. Non-invasive samples are often the only available source of DNA for many wild species, but these samples yield low quantity and quality of DNA. Such samples present unique methodological challenges, which is the prime reason advanced genetic techniques are not used in wildlife and conservation studies. Researchers are tailoring laboratory and computational methods to better extract the information contained in non-invasive DNA. We obtained non-invasive fecal scats of snow leopards from seven different countries, and applied DNA capture, a widely used method for studying human populations, ancient DNA samples, eDNA, and forensics. We successfully sequenced endogenous DNA from samples. Additionally, we sequenced a significant amount of non-target DNA, mostly mitochondrial (mt) DNA. This non-target mtDNA originated from snow leopard and prey species. Here, we present how we can use this non-target data to harvest important genetic information about snow leopard populations, such as mtDNA-based analysis of population structure and diet.

## **W427: Flavor, Nutrition, and Post-Harvest Genomics**

### **Welcoming Remarks**

**Vincent Colantonio**, University of Florida, Gainesville, FL

## **W428: Flavor, Nutrition, and Post-Harvest Genomics**

### **Genetics of Carotenoids, Kernel Color, and Vitamin E in Maize Grain**

**Christine Diepenbrock**, University of California Davis, Davis, CA

Carotenoids (provitamin A) and tocochromanols (vitamin E) are lipid-soluble compounds that have important (and in some contexts, essential) functions in humans, animals, and plants. Vitamin A deficiency remains prevalent, including in geographic regions in which maize is consumed as a food staple. Improving the levels of certain nutrients through plant breeding and/or agronomy—a process termed biofortification—has been found to be a sustainable and cost-effective solution for amelioration of deficiency. To dissect the genetic basis of natural variation for eight carotenoid and ten tocochromanol traits in maize grain, we conducted joint linkage analyses, genome-wide association studies (using the ~29 million markers of HapMap v1 and 2), and RNA-seq expression analyses in the 5,000-line U.S. maize nested association mapping population. More than half of the identified quantitative trait loci (QTL) were found to be underlain by genes having a priori roles in carotenoid and tocochromanol biosynthesis and retention. Novel genes were also identified with roles in structure, transport, metabolism, and chlorophyll biosynthesis. More than half of the identified QTL were expression QTL, and most were also found to exhibit pleiotropy (in the form of highly correlated QTL allelic effect estimates across traits). We also dissected the genetic basis of kernel color—an important trait for acceptability of biofortified, provitamin A-dense maize—in 1,651 genotypes from the Ames maize inbred panel, in a genome-wide association study. Certain genes identified for kernel color have been implicated in both the content (overall abundance) of carotenoids and, importantly, the composition (relative abundance) of individual carotenoids. The latter finding suggests that breeding for deeply orange, provitamin A-dense maize must be conducted with careful attention to the interplay between kernel color and priority nutritional traits within the carotenoid pathway. Together, these studies have increased our understanding of the accumulation and retention of carotenoids and tocochromanols and the formation of kernel color in maize grain, within an analytical framework that is readily integrated with genomics-assisted breeding strategies.

## **W429: Flavor, Nutrition, and Post-Harvest Genomics**

### **Metabolomic Selection for Enhanced Fruit Flavor**

**Marcio F. R. Resende**, University of Florida, Gainesville, FL

Although a staple in cuisines globally, commercial fruit varieties have become considerably less flavorful. Due to the cost and difficulty associated with flavor phenotyping, many breeding programs have long been barred from selecting for this complex trait. To address this issue, we leveraged targeted metabolomics of diverse tomato and blueberry accessions and their corresponding consumer flavor panel ratings to create statistical and machine learning models that can predict sensory perceptions of fruit flavor.

## **W430: Flavor, Nutrition, and Post-Harvest Genomics**

### **Sugar Metabolism Reprogramming during Ripening of a Climacteric Fruit and Its Non-Climacteric Mutant**

**Macarena Farcuh**, University of Maryland, College Park, College Park, MD, Bosheng Li, University of California, Davis, Rosa Rivero, CEBAS, CSIC, Spain, Avi Sadka, Department of Fruit Trees Sciences, Institute of Plant Sciences, Bet Dagan, Israel and Eduardo Blumwald, Dept. of Plant Sciences, University of California, Davis, CA

Sugar metabolism, which determines fruit sugar content and composition, is of crucial importance to develop cultivars that can meet consumer expectations in terms of flavor and nutritional value. We investigated sugar metabolism in leaves and fruits of two Japanese plum (*Prunus salicina* Lindl.) cultivars, the climacteric Santa Rosa and its bud sport mutant the non-climacteric Sweet Miriam, during ripening on the tree. We used a systems biology approach to show a reprogramming of metabolism of major and minor sugars occurring in fruits and leaves of both cultivars. Pearson coefficient correlations of metabolomics and transcriptomic data and weighted gene co-expression network analysis (WGCNA) of RNA sequencing (RNA-Seq) data allowed the identification of eleven key sugar metabolism associated genes. The expression patterns of these genes were further validated based on transcript levels, and the functions of gene products were assessed enzymatically and by metabolite analyses. Our results demonstrated the reprogramming of sugar metabolism in both leaves and fruits in the non-climacteric plum, which displayed a shift towards increased sorbitol synthesis, and lower amounts of sucrose, glucose, and fructose as compared to climacteric plums. Climacteric and non-climacteric fruits showed differences in their UDP-galactose metabolism towards the production of galactose and raffinose, respectively. Overall, our results contribute to the understanding of mechanisms that allow fruits to 'switch' to a sorbitol-based metabolism, which could have important impacts, since sorbitol is a healthier natural sweetener to sucrose.

#### **W431: Flavor, Nutrition, and Post-Harvest Genomics**

##### **Pedigree-Informed Genome-to-Phenome Mapping of Morphometric Traits in Strawberry**

**Mitchell J. Feldmann**<sup>1</sup>, Michael A. Hardigan<sup>1</sup>, Randi A. Famula<sup>1</sup>, Amy Tabb<sup>2</sup>, Glenn S. Cole<sup>1</sup> and Steven J. Knapp<sup>1</sup>, (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)USDA-ARS-AFRS, KEARNEYSVILLE, WV

Over the last 300 years, domestication and breeding have profoundly altered the development and morphology of strawberry, greatly distancing modern cultivars from their wild progenitors. One result has been the emergence of high yielding cultivars with large, firm, visually appealing, long shelf-life fruit that can withstand the rigors of harvest, handling, storage, and long-distance shipping. At the other extreme, wild ecotypes and heirloom cultivars are low yielding and fast-ripening with exceedingly short shelf-lives and significantly greater percentages of deformed and flawed fruit. The genes targeted by selection for this phenotypic complex have not been identified, although hypotheses have emerged from genome-wide scans for selective sweeps that large-effect genes could be involved. Genome-to-phenome mapping of previously identified signatures of selection, however, have not yet been undertaken. Our study partly focused on the problem of translating geometric phenotypic variables into objective, human-understandable phenotypic variables for application in forward genetic studies in strawberry. We phenotyped 7,000 fruit harvested from 575 hybrid individuals developed with a factorial mating design in a coastal California environment. These individuals were genotyped with a 49,000-single nucleotide polymorphism (SNP) array anchored to the octoploid genome. We applied Generalized Procrustes, Elliptical Fourier, and several latent variable methods to the problem of holistically describing fruit shape and other fruit attributes. These analyses yielded continuous phenotypic variables for genome-wide association and genomic prediction studies. Our initial results will be reported.

#### **W432: Flavor, Nutrition, and Post-Harvest Genomics**

##### **Use of Genomics Tools to Study Postharvest Senescence in *Brassica* vegetables**

**Tie Liu**, University of Florida, Gainesville, FL

The facts of postharvest food loss and waste and the resulting consequences affect us in many ways, ranging from important economic and social issues to lasting and detrimental environmental problems. We are using genomic tools to understand preharvest and postharvest senescence and deterioration in brassica vegetables. This is particularly interesting because little is known about the molecular mechanism of postharvest senescence in *Brassica* species including broccoli, cabbage, cauliflower. To carefully characterize the effects of environmental stresses on postharvest broccoli, we are working on identifying candidate transcripts, related proteins and metabolic compounds that accurately reflect the physiological age or freshness of broccoli and the other *Brassica* species during postharvest storage

using transcriptomics, proteomics and metabolomics approaches. Identifying those 'fresh-indicators' that has the potential to mediate senescence and to generate germplasm for breeding new varieties with stress tolerance and postharvest color retention in Brassica vegetables. The long-term goal of the proposed research is to aid development of an innovative tool to accurately estimate the freshness of produce. Such a tool would allow a new level of postharvest logistics, supporting availability of high-quality, nutritious, fresh vegetables and fruits.

#### **W433: Flax Genomics**

##### **Comparative Transcriptomics of Root Responses to Pathogenic (*Fusarium oxysporum* f. sp. *lini*) and Non-Pathogenic (*Rhizoglyphus irregulare*) Fungi**

**Michael K. Deyholos**, University of British Columbia, Kelowna, BC, Canada

*Fusarium oxysporum* f. sp. *lini* is a hemibiotrophic pathogen, and the cause of wilt in flax (*Linum usitatissimum*). In contrast, *Rhizoglyphus irregulare* is an arbuscular mycorrhizal fungus that generally forms mutualisms with flax. Both *F. oxysporum* and *R. irregulare* colonize roots at the beginning of their symbioses. We were therefore interested to compare how gene expression patterns in roots of flax inoculated with *F. oxysporum* differed from roots inoculated with *R. irregulare*. We used RNASeq to compare flax at two time points that had been inoculated with four treatments: a mock control; *F. oxysporum*; *R. irregulare*; or both fungi simultaneously. Results show clear evidence of an active response to the pathogen, with a more passive response to AMF. We discuss these results within the context of concepts such as priming, SAR (systemic acquired resistance), and MIR (mycorrhizal induced resistance).

#### **W434: Flax Genomics**

##### **Can Flax 'edit' its Genome**

**Christopher Cullis**, Case Western Reserve University, Cleveland, OH

The rapid variation of the genome in flax in response to the environment provides a window into possible mechanisms by which new sequences, perhaps containing new genes, can arise *de novo*. These regions are not just random assemblies of nucleotides, but are specific fragments that are present elsewhere in the germplasm. The data are consistent with there being a mechanism that can result in specific new sequence information arising, perhaps in response to environmental stimuli.

#### **W435: Flax Genomics**

##### **Gene Expression Analysis Associated with Salinity Stress in Flax/Linseed (*Linum usitatissimum*)**

**Manisha Sharma**<sup>1</sup>, **Rhithu Rai**<sup>1</sup>, **Rohini Sreevathsa**<sup>2</sup>, **Raju Datla**<sup>3</sup> and **Prasanta K Dash**<sup>4</sup>, (1)ICAR-National Institute of Plant Biotechnology, New Delhi, India, (2)ICAR-NRC on Plant Biotechnology, New Delhi, India, (3)Global Institute for Food Security, Saskatoon, SK, Canada, (4)ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

Salinity is one of the most important environmental factors that limit crop growth, development and productivity. Globally, almost 20% of the world's cultivated land and nearly 50% of irrigated land suffer from saline stress. Saline stress has extensively studied in model plant Arabidopsis. However, in the field crop flax (*Linum usitatissimum* L.) the molecular and physiological mechanisms of salt tolerance are yet to be elucidated. Our study involved comprehensive analysis of flax seedlings to analyze expression perturbation during salinity stress. We carried out physiological and whole genome transcriptomic study of flax roots and shoots exposed to a 6 h and 12 h of salinity treatment. A total of 6134 differentially expressed genes (DEGs) were identified in shoot and root tissues together. Further, in-depth analysis of tissue specific DEGs revealed that gene groups involved in ion transport such as sodium proton antiporter (NHX), osmotic stress, ROS homeostasis and calcium signaling are involved in ionic homeostasis in flax seedlings. Further based on sequence homology search, a key family of Na antiporters in flax-LusNHX was identified and their expression was studied. In conclusion, our results

suggest that the differential expression of stress and transport related genes might be contributing to fine-tune the ionic and ROS homeostasis in flax tissues.

#### **W436: Flax Genomics**

##### **Genomic Improvement of Flax using Natural Variation and Induced Mutations**

**Bradley J. Till**<sup>1</sup>, Gabriela A. Aravena Abarzua<sup>1</sup>, Rachel Howard-Till<sup>1</sup>, Rocío Quián<sup>1</sup>, Humberto Gajardo<sup>1</sup>, Sylvie Cloutier<sup>2</sup>, Bourlaye Fofana<sup>3</sup> and Braulio J. Soto-Cerda<sup>1</sup>, (1)Centro de Genómica Nutricional Agroacuicola (CGNA), Temuco, Chile, (2)Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (3)Charlottetown Research and Development Centre, Agriculture and Agri-Food Canada, Charlottetown, PE, Canada

Modern technologies have enabled more precise use of nucleotide variation for functional genomics and breeding. Alleles in germplasm collections can be efficiently identified and combined, and limitations of a narrow genetic base in many crop species can be overcome with mutagenesis. We have developed flax as an oilseed crop for southern Chile. High omega-3 and low mucilage varieties have been released. We are currently using multiple GWAS methods for the genetic dissection of agronomic, seed quality and resilience-related traits. More than 200 stable quantitative trait nucleotides (QTNs) have been identified. For example, fifty-six QTNs for drought-related traits along with accessions carrying favorable alleles were identified and used to pyramid superior drought-adaptive alleles. The current wealth of QTN data will be used for the development of new “smart” flax cultivars using marker-assisted breeding and genomic selection. Mutant populations are also being developed with existing EMS-mutagenized seed and newly gamma-irradiated material. Pooled amplicon sequencing methods have been established for recovery of point mutations in chemically mutagenized material for TILLING reverse-genetic screens. Gamma irradiation produces a broader spectrum of induced alleles and has been successfully used in breeding programs for over 80 years. To facilitate the use of gamma irradiation, we are adapting low-coverage whole genome sequencing for routine recovery of induced copy number variation. Our aim is to create a platform for rapid deployment of natural and induced alleles to ensure sustainable productivity in a changing environment.

#### **W437: Flax Genomics**

##### **Insights into Genetic Structure and Genomic Prediction of Powdery Mildew Resistance in Flax (*Linum usitatissimum* L.)**

**Sylvie Cloutier**<sup>1</sup>, Frank You<sup>1</sup>, Khalid Rashid<sup>2</sup>, Chunfang Zheng<sup>3</sup>, Samuel Lam<sup>3</sup>, Nadeem Khan<sup>3</sup>, Liqiang He<sup>3</sup>, Jin Xiao<sup>4</sup>, Zhen Yao<sup>2</sup> and Pingchuan Li<sup>2</sup>, (1)Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)Agriculture and Agri-Food Canada, Morden, MB, Canada, (3)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (4)Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, Ottawa, ON, Canada

In flax, powdery mildew (PM) is caused by the fungus *Oidium lini* which can cause defoliation and reduce seed yield and quality. Until now, only one major dominant gene (*Pm1*) and three QTL on chromosome 1, 7 and 9 were reported for PM resistance. To fully dissect genetic structure and identify quantitative trait loci for PM resistance, a diverse flax core collection of 372 accessions and an additional 75 breeding lines were re-sequenced and field resistance evaluation was performed for six to eight years (2010-2017) in Morden, Manitoba, Canada. Genome-wide association studies were performed using two single-locus and seven multi-locus statistical models, 247,160 single nucleotide polymorphisms (SNPs) and the phenotypes of the 447 individuals for each individual year and the means over years. A total of 355 QTL based on haplotype blocks were identified, of which, 43 large effect QTL ( $\geq 10\%$  of  $R^2$ ) were highly stable over years. All QTL explained 91% of the phenotypic variation in the genetic panel. The total number of favorable alleles per accession was significantly correlated with PM resistance ( $r = 0.74$ ) and genomic selection (GS) models using all identified QTL generated significantly higher prediction accuracy ( $r = 0.93$ ) than those using all genome-wide SNPs ( $r = 0.42$ ), validating the overall reliability of the QTL. The QTL were found on all 15 chromosomes but the 43 major QTL were clustered in 16

genomic regions of 12 chromosomes, especially on chromosome 5 (0.4-5.6 Mb and 9.4-16.9 Mb) and 13 (4.7-5.2 Mb). Of the 355 QTL, 236 harbored disease resistance related genes (390) within 200 Kb. In particular, 27 major QTL co-located with NBS-related genes, receptor-like kinase, protein-like kinase, TM-CC, RPW8, WRKY, and mildew locus O (MLO) genes. In conclusion, QTL for PM resistance are additive and GS models based on QTL data are effective for predicting resistance of breeding lines for use in flax molecular breeding.

#### **W438: Forage, Feedstocks & Turf**

##### **An Empirical Test of Genomic Selection to Improve Alfalfa Forage Yield**

**E. Charles Brummer**, University of California, Davis, Davis, CA, Donald Viands, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, Julie Hansen, Cornell Univ., Ithaca, NY, Xuehui Li, North Dakota State University, Fargo, ND and Yanling Wei, Aldevron

Alfalfa, *Medicago sativa* L., is widely grown in temperate and dry subtropical regions of the world for highly nutritious livestock forage. Somewhat inexplicably, biomass yield increases observed during much of the last century have stalled over the past 25-30 years in most areas of the USA. As part of a larger breeding program, we have begun to investigate the use of genomic selection as a possible avenue to improve forage biomass yield. Several years ago we developed a genomic selection model based on a recurrent selection breeding program. We then applied that model to indirectly select for higher forage yield through two cycles of selection. Selection for both higher and lower yield was accomplished and a randomly chosen control population was included. Seed was increased of all populations and yield trials were planted at Ithaca, NY and Tulelake, CA. We have completed two full years of yield data after the establishment year. Preliminary results suggest that genomic selection worked to improve or to decrease yield, depending on the intended direction, at least in the first year post-establishment. The talk will discuss the results and suggest ways to improve on the methodology. Given that field-based yield estimates require multiple harvests over multiple years, even small improvements in yield from a genomics-based program could be beneficial.

#### **W439: Forage, Feedstocks & Turf**

##### **Greenhouse Gas Removal and Climate Resilient Grassland Crops**

**Iain Donnison**, Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth University, United Kingdom

Energy grasses such as *Miscanthus* in addition to providing a renewable source of energy, materials and chemicals, when combined with carbon capture utilisation and storage can be used to facilitate greenhouse gas removal and enable economies to become net zero by providing negative emissions to compensate for hard to decarbonise industries. Plant breeding and agronomic targets for sustainable bioenergy requires crops that can achieve high yields from low inputs, deliver ecosystem services including carbon sequestration and produce a harvestable biomass that can be converted with maximum efficiency. A further challenge for energy crops is how do we produce sufficient feedstock without negatively impacting food production and whilst maximising environmental benefits. Means by which this may be achieved include growing energy crops on marginal and even contaminated land which requires crops which exhibit greater stress tolerance and resilience. The presentation will consider how *Miscanthus* germplasm and subsequent varieties can be used to deliver productive agronomic and environmental benefits, and what traits and associated agronomies represent targets for crop improvement and upscaling.

#### **W440: Forage, Feedstocks & Turf**

##### **Harnessing Genetic Diversity to Improve Key Traits in Red Clover**

**Roland Kölliker**<sup>1</sup>, Lea A. Frey<sup>1</sup>, Franz Xaver Schubiger<sup>2</sup>, Christoph Grieder<sup>2</sup>, Leif Skot<sup>3</sup>, Tom Ruttink<sup>4</sup>, Steven Yates<sup>1</sup>, Bernadette Julier<sup>5</sup> and Bruno Studer<sup>1</sup>, (1)Molecular Plant Breeding, Institute of

Agricultural Sciences, ETH Zurich, Zurich, Switzerland, (2)Agroscope, Switzerland, (3)IBERS, Aberystwyth University, Aberystwyth, United Kingdom, (4)ILVO Plant Sciences, Melle, Belgium, (5)INRAE, P3F, France

Red clover (*Trifolium pratense* L.) is an important forage legume of temperate regions, particularly valued for its high yield potential and its high forage quality. Despite the significant progress achieved through systematic breeding in the last decades, continuous improvement of cultivars is crucial to ensure yield stability in view of newly emerging diseases or changing climatic conditions. The large amount of genetic diversity present in red clover ecotypes, landraces and cultivars provides an invaluable, but often untapped resource for the improvement of key traits such as yield, quality and resistance to biotic and abiotic stresses. In the framework of the EU funded Horizon 2020 project "EUCLEG" ([www.eucleg.eu](http://www.eucleg.eu)), a collection of 395 red clover accessions representing cultivars, breeding material, landraces and ecotypes of 25 countries from five continents has been established. All accessions have been genotyped using a pooled GBS (genotyping by sequencing) approach with 200 plants per accession. Field experiments have been established at five locations (United Kingdom, Czech Republic, Switzerland, Norway and Republic of Serbia) in 2018 and analysis of first full growing season phenotypic data (2019) revealed large variation for traits such as juvenile development, dry matter yield, vigour or flowering time. In addition, resistance to biotic and abiotic stress was evaluated in separate experiments. For example, resistance to southern anthracnose, caused by *Colletotrichum trifolii*, was assessed in the greenhouse by spray inoculation using a single spore isolate, followed by an additional inoculation of surviving plants with a mixture of seven additional isolates. Repeatability across three replicates ranged from  $r=0.62$  to  $r=0.67$  for both inoculations and a significant differences across accessions was observed. The mean survival rate for single and mixed isolate inoculation was 26.7% and 16.5%, respectively. Although some cultivars with considerable resistance to the disease were observed (survival rates >50%), this highlights the urgent need to improve resistance to southern anthracnose in red clover. However, these phenotypic results provide a valuable basis for GWAS and the identification of candidate resistance genes. Resistance screening in four unrelated bi-parental  $F_1$  populations indicated resistance to be controlled by one or few resistance loci. GBS sequencing using resistant and susceptible pools from these populations revealed a number of candidate genes, which could valuably complement the resistance sources identified in the EUCLEG accessions.

**Acknowledgments:** This research has received funding from the European Union's Horizon 2020 project EUCLEG under grant agreement n°727312.

#### **W441: Forage, Feedstocks & Turf**

##### **The Roles of *Spl* Genes in Flowering Regulation and Axillary Bud Information in Switchgrass**

Zengyu Wang, College of Grassland Science, Qingdao, China

#### **W442: Forage, Feedstocks & Turf**

##### **Translational Genomics of Forage Legumes to Advance Molecular Breeding of Seed Traits in Cover Crop Legumes**

Maria J. Monteros, Tim D. Hernandez, Yanina Alarcon, Shahjahan Ali and Nick Krom, Noble Research Institute, Ardmore, OK

Cover crop legumes provide ground cover to reduce soil erosion, can improve soil structure and are capable of symbiotic nitrogen fixation. Although seed shattering for dispersion is essential for the distribution of wild plants, shattering in many crop legumes including hairy vetch (*Vicia villosa* Roth) is undesirable because it results in seed yield losses. Despite differences in genome size, chromosome number and ploidy level, the genetic relatedness and genome conservation in multiple forage and cover crop legumes enables translational and comparative genomics for trait development. The objectives of this project are to identify genomic regions targeting pod indehiscence to use for genomics-based breeding and selection approaches to develop enhanced cover crop cultivars. Complementary approaches include leveraging the genome sequences of alfalfa, soybean and other legumes to mine

conserved regions of candidate genes and transcription factors associated with pod indehiscence traits, genome-wide association studies and mining transcriptomic datasets to identify differentially expressed genes between high and low seed shattering vetch genotypes. Hairy vetch accessions contrasting for seed shattering traits under field evaluations at three locations were genotyped using genotyping-by-sequencing (GBS). Detection of SNPs and indels between seed shattering and non-seed shattering hairy vetch plants enabled the discovery of candidate gene targets associated with pod retention. Genes differentially expressed between seeds and pods vs. other plant tissues provide further insight to prioritize their utilization as functional markers for efficient trait integration. Integration of cover crop legume cultivars with shatter resistance can enhance the resiliency of agricultural systems.

#### **W443: Forage, Feedstocks & Turf**

##### **Multi-Omics Prediction of Deep Root Development in Perennial Ryegrass**

**Torben Asp**, Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark

In the future, crops will likely face more severe drought periods during the growing season. It is therefore necessary to identify germplasm resilient to drought for sustainable intensification of agriculture. A key aspect of this is the development of crops with deeper roots. Roots are vital organs for plants, and the effective use of resources from the soil is important for yield stability. However, phenotypic variation in root traits among crops is mostly unknown and field screening of root development is costly and labour demanding. As a consequence, new methods are needed to investigate root traits of crops under field conditions, particularly roots in the deeper soil horizons. We have developed a large-scale phenotyping facility (RadiMax) for the study of root growth and soil resource acquisition under semi-field conditions. Roots are observed through mini-rhizotrons using a multispectral imaging system, and images are analyzed using deep neural network. Plants are grown in rows perpendicular to a water stress gradient created by a multi-depth sub-irrigation system and movable rainout shelters. The water stress gradient allows for a direct link between root observations and the development of stress response in the canopy.

A total of 300 perennial ryegrass populations were tested in the RadiMax facility. The transcriptomic and epigenetic responses to drought stress and deep root development were studied. Gene co-expression network analysis allowed for the identification of regulatory networks involved in deep root development and drought response. Multi-omics prediction models integrating SNP markers, transcriptomic and epigenetic data were developed, and used to predict deep root development and above-ground traits.

#### **W444: Forest Tree**

##### **The Early Bud-Break Regulon in Poplar**

**Abdul Azeez**, School of Forest Resources and Environmental Science, MTU, Houghton, MI

Bud dormancy is an adaptive strategy in perennials from temperate and boreal climates to survive unfavorable conditions during winter months. Bud dormancy is developed in fall, released after prolonged cold during winter and followed by reactivation of growth (aka bud-break) in response to warm temperatures in spring. Molecular regulation of dormancy release and bud-break are largely unknown. We have identified through activation tagging early bud-break dominant (*ebbD*) poplar mutants and accordingly the underpinning genes are called EARLY BUD-BREAK (EBB). Previously we have reported the isolation and characterization of the EBB1 gene from one of these mutants, encoding an AP2/ERF transcription factor. We have now found that EBB1 directly represses the poplar SVL (SHORT VEGETATIVE PHASE-LIKE), a MADS-box protein which was recently found to negatively regulate bud-break in *Populus*. We also report the identification and characterization of the *ebb3D* mutant. The corresponding gene EARLY BUD-BREAK 3 (EBB3), encodes another AP2/ERF domain-containing transcription factor. EBB3 overexpressing lines showed early bud-break whereas, RNAi plants showed significantly delayed bud-break as compared to wild type control. We show that EBB3 is downstream of EBB1 and SVL. EBB1 positively, while SVL negatively regulates *EBB3* expression. Further, we show that EBB3's effect on bud-break is mediated by the regulation of the cell cycle. EBB3 directly and positively regulates the *CYCD3.1* gene, encoding an important checkpoint in the progression of the cell

cycle. In summary, our results outline the backbone of a novel regulatory module that controls dormancy release and reactivation of growth in poplar.

#### **W445: Forest Tree**

##### **Genome-Wide Association Study of Wood Anatomical and Morphological Traits in *Populus trichocarpa* Torr. & Gray**

**Hari Chhetri**, West Virginia University, Morgantown, WV

To understand the genetic mechanisms underlying wood anatomical and morphological traits in *Populus trichocarpa*, we used 869 unrelated genotypes from a common garden in Clatskanie, OR that were previously collected from across the distribution range in western North America. Using GEMMA mixed model analysis, we tested for the association of 25 phenotypic traits and 9 multitrait combinations with 6.741 million SNPs covering the entire genome. Broad-sense trait heritabilities ranged from 0.117 to 0.477. As expected, traits were correlated within the trait types - anatomical traits were correlated to each other and morphological traits were correlated to each other. Most traits were significantly correlated with geoclimatic variables suggesting the role of climate and geography in shaping the variation of this species. We identified a total of 17 and 16 gene models from single and multitrait GWAS, respectively. Two SNPs from single trait GWAS and six SNPs from multitrait GWAS passed a Bonferroni threshold of  $7.417 \times 10^{-9}$ , leading to the identification of two and five nearby candidate genes, respectively. We have presented here one of the most comprehensive GWAS analyses for *P. trichocarpa* to date including the first GWAS for wood anatomical traits for this species. Candidate genes for wood anatomical traits were related to defense mechanisms and abiotic stress tolerance. Genes associated with morphological traits were involved in light and hormone signaling pathways. The identified genes have great potential for enhancing *Populus* stress tolerance, and for optimizing lignocellulosic biofuel production.

#### **W446: Forest Tree**

##### **Evo-Devo of Secondary Growth Traits in the Seed Plant Lineage**

**Danielle Roodt**, University of Pretoria, Pretoria, South Africa

Wood (secondary xylem from vascular cambium) is found in the gymnosperm and angiosperm lineages but has been lost in the monocotyledons. Although wood formation is ancestral to the seed plants and was likely lost at the base of the monocots, differences between the eudicots and most gymnosperm lineages are apparent at the anatomical, cell wall and molecular levels. One gymnosperm lineage, the Gnetales, shares similarities with the eudicots but not other gymnosperms. We aimed to gain insight into the cause of this gain and loss of woody characteristics in the Gnetales and monocots, respectively. We analysed orthologs of genes preferentially expressed in developing xylem of eudicot trees across the angiosperms, constructed a regulatory model of early vascular cambial cell identity and differentiation and investigated its conservation in the sequenced genomes of 26 eudicot and seven monocot species and the early diverging angiosperm *Amborella trichopoda*. We also constructed high quality gene catalogues for representative species from the three non-coniferous gymnosperm lineages from transcriptome data of xylogenetic and contrasting tissues, allowing for comparative transcriptomics and investigations of the gene networks underlying xylogenesis. Furthermore, we performed wood morphology and wood chemistry comparisons between the seed plant lineages and identified candidate genes for further elucidation of convergently evolved or ancestrally retained traits. These results broaden our understanding of the evolution of the highly divergent seed plant lineage and provide candidate genes that may have led to the loss of wood formation in the monocot lineage, and the development of “angiosperm-like” traits in the Gnetales.

#### **W447: Forest Tree**

##### **Genome-Wide Scans Reveal How Gene Dosage Affects Quantitative Trait Variation in *Populus***

Héloïse Bastiaanse<sup>1</sup>, Luca Comai<sup>2</sup>, Isabelle M. Henry<sup>3</sup> and **Andrew T. Groover**<sup>1</sup>, (1)US Forest Service, Davis, CA, (2)Plant Biology and Genome Center, UC Davis, Davis, CA, (3)UC Davis, Davis, CA

Recent studies are uncovering gene copy number and structural variation within and among *Populus* species, but the significance of such gene dosage variation on phenotypic traits is largely unknown. Here, we used a large pedigree of *Populus nigra* x *deltoides* hybrids carrying deletions and additions to assay the effect of gene dosage variation on leaf morphology and related gene expression. Gamma irradiation of a *P. nigra* male was used to induce chromosomal deletions and additions, and used in a controlled cross to *P. deltoides* females. The DNA from several hundred resulting F1 individuals was sequenced to identify the exact locations in indels in each individual, and after clonal propagation individual genotypes were introduced into field trials. For two consecutive years, the first fully mature leaf was sampled from the apex of each ramet and used for leaf shape analysis. The outline of each leaf was automatically retrieved from individual images and used in a global morphometric analysis, including descriptors of leaf shape and leaf size. Dosage-dependent QTL (dQTL) were identified for multiple traits, showing a prevalence of dosage-sensitive genes influencing leaf morphology traits across the genome. RNA was extracted from newly initiated leaves of each ramet for gene expression analysis. Gene expression was assayed in leaves for lines associated with trait extremes and dQTL using RNAseq. The effect of dosage on expression levels was evaluated, and eQTL were identified that correlated with morphological traits and phenotypic extremes. Expression data were also used in a co-expression framework to identify gene modules associated with leaf morphology traits or enriched with genes of specific functions associated with leaf development. Overall this study provides a detailed dataset that illustrates linkages between gene dosage, gene expression and phenotypic trait variation.

#### **W448: Forest Tree**

##### **Development of a High-Throughput Genome-Wide Genotyping Array for Tropical and Subtropical Pine Species**

**Colin Jackson**<sup>1</sup>, Nanette Christie<sup>2</sup>, Madison Caballero<sup>3</sup>, Melissa Reynolds<sup>2</sup>, Christopher Marais<sup>2</sup>, Erik A. Visser<sup>2</sup>, Sanushka Naidoo<sup>2</sup>, Gary Hodge<sup>1</sup>, Ross Whetten<sup>1</sup>, Jill L. Wegrzyn<sup>3</sup>, Fikret Isik<sup>1</sup>, Juan José Acosta<sup>1</sup> and Alexander A. Myburg<sup>2</sup>, (1)Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC, (2)Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa, (3)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

We performed gene and genome targeted SNP discovery towards the development of a high-throughput, genome-wide genotyping array for tropical pines. Pooled RNA-seq data from shoots of seedlings from five species, *Pinus patula*, *Pinus tecunumanii*, *Pinus oocarpa*, *Pinus greggii* and *Pinus maximinoi* was used to identify between 426K and 1.3M transcriptome based SNPs per species. Using Affymetrix SNP probe design criteria we identified 175K to 301K candidate probes per species aggregating to a total of 1.3M probes. In addition, capture-seq data for the five species plus *Pinus caribaea* was generated using a custom 40K capture probe set (RAPiD Genomics, Gainesville, FL) of which ~30K were from single copy regions of the v2.01 *Pinus taeda* genome assembly and ~10K were from *P. patula* and *P. tecunumanii* transcriptome assemblies. Across the six species, pooled DNA from 81 provenances representing their natural ranges was analyzed using capture-seq. We identified 1.02 million SNPs that were detected in at least two of the 81 provenances, of which 415K are shared among two or more species translating into 562K candidate SNP probes. We will present SNP discovery and probe design based on the RNA-seq and capture-seq data (non-redundant set of 1.7 million SNP probes) as well as our strategy for selecting 420K markers for a genome-wide SNP screening array. Based on the screening array results, 50K SNPs will be selected for a final production array with utility in six pine species that together with their hybrids represent the majority of tropical and subtropical pine plantations globally.

#### **W449: Forest Tree**

##### **Can Poplar Plants Use Mobile Protein Signals to Influence Mycorrhizal Fungi?**

**Xiaohan Yang**<sup>1</sup>, Hua Cheng<sup>2</sup>, Linling Li<sup>2</sup>, Xiaoli Hu<sup>2</sup>, Md Mahmudul Hassan<sup>2</sup>, Zhenzhen Qiao<sup>2</sup>, Haiwei Lu<sup>2</sup>, Guoliang Yuan<sup>2</sup>, Jin Zhang<sup>2</sup>, Degao Liu<sup>2</sup>, Jennifer Morrell-Falvey<sup>2</sup>, Jessy Labbe<sup>2</sup>, Wei Shen<sup>3</sup>, Shi-You Ding<sup>3</sup>, Julie Mitchell<sup>2</sup>, Jay Chen<sup>2</sup>, Wellington Muchero<sup>2</sup> and Gerald Tuskan<sup>1</sup>, (1)Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, (2)Oak Ridge National Laboratory, Oak Ridge, TN, (3)Michigan State University, East Lansing, MI

Symbiosis between plants and fungi in the rhizosphere represents a very important type of mutual beneficial plant-microbe interactions. It is widely accepted that many fungal species can influence their plant hosts using secreted proteins. However, it has not been well established whether plants can use secreted protein signals to affect their fungal partners. To address this issue, we have initiated an effort to investigate the role of small secreted proteins (SSPs) in *Populus* species (poplar), which are fast-growing trees important for biofuel production and ecosystems service, in response to the colonization of ectomycorrhizal fungus *Laccaria bicolor*, which forms symbiosis with poplar roots to improve tree health and growth. Through computational analysis of the *P. trichocarpa* transcripts responsive to *L. bicolor* inoculation, we predicted more than 10 *P. trichocarpa* SSPs (PtSSPs) having DNA-binding capability. Then we used various approaches to characterize these DNA-binding PtSSPs. Specifically, transgenic *Arabidopsis thaliana* and poplar plants were created to overexpress PtSSPs fused to green fluorescent protein (GFP). The transgenic plants were co-cultured with *L. bicolor* to study the movement of the PtSSP-GFP fusion proteins from the plant roots into the fungal hyphae. Also, transgenic yeast strains overexpressing the PtSSPs fused to GFP were created to examine the PtSSP secretion from yeast and the entry of PtSSPs into fungal hyphae. Furthermore, the genetic diversity and evolution of the PtSSPs were investigated using the rich genomic resources available for poplar. In addition, the protein domains and 3D structures of these PtSSPs were predicted using computational approaches.

#### **W450: Forest Tree**

##### **Genomics of Quantitative Resistance to White Pine Blister Rust in Sugar Pine**

**Matthew Weiss**<sup>1</sup>, Daniela Puiu<sup>2</sup>, Richard Sniezko<sup>3</sup>, Marc Crepeau<sup>4</sup>, Steven L. Salzberg<sup>2</sup>, David B. Neale<sup>5</sup> and Amanda de la Torre<sup>6</sup>, (1)Northern Arizona University, Flagstaff, AZ, (2)Johns Hopkins University, School of Medicine, Baltimore, MD, (3)USFS Dorena Genetic Resource Center, Portland, OR, (4)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (5)Department of Plant Sciences, University of California, Davis, Davis, CA, (6)Northern Arizona University, Flagstaff, AZ

Five-needle pines in North America are currently threatened by an invasive fungal disease, White Pine Blister Rust (WPBR). Efforts to mitigate effects of WPBR in sugar pine (*Pinus lambertiana*) have focused on breeding resistant trees. This project aims to identify genes conferring quantitative resistance to WPBR, which may have potential for genomic-based breeding in sugar pine. We tested for genome-wide associations (GWAS) between a set of 130k SNPs obtained from whole-genome resequencing, and WPBR disease phenotypes and presence of the major gene for resistance in individuals distributed across the species' range. Population structure was determined using principal component analysis with Aegenet. A Bayesian cluster analysis using fastSTRUCTURE was conducted using 10 independent runs of K=2-10. The most frequent  $\Delta K$  from these 10 runs was used to determine number of clusters within the population (K=4). A total of 154 SNPs were significantly associated with the measured traits. Of the significant SNPs, 27 were from coding regions of the sugar pine genome with complete annotations found for 19 of these SNPs. An average sex map constructed from two parental maps combined using LPmerge. The resulting map was 1943.116 cM in length and consisted of 8702 SNPs at 5527 unique loci.

#### **W451: Forest Tree**

##### **Landscape Genomics, Ancestry Mapping and Nested Association Analysis in *Eucalyptus grandis* and its Interspecific Hybrids**

Marja M. O'Neill<sup>1</sup>, Julia Candotti<sup>1</sup>, J. Stephanus Henning<sup>1</sup>, Stephanus Engelbrecht<sup>1</sup>, S. Melissa Reynolds<sup>1</sup>, Nanette Christie<sup>1</sup>, Justin O. Borevitz<sup>2</sup> and **Alexander A. Myburg**<sup>1</sup>, (1)University of Pretoria, Pretoria, South Africa, (2)The Australian National University, Canberra, ACT, Australia

Interspecific hybrids of *Eucalyptus grandis*, particularly with *E. urophylla* and *E. nitens*, are widely planted in subtropical and temperate regions, respectively. *E. grandis* has been cultivated and bred in South Africa for over 100 years for divergent growing regions and end products (timber vs pulp and paper) creating an interesting opportunity to study the genomic consequences of artificial selection. Towards that, we first assessed neutral and adaptive genetic diversity in 596 individuals from 32 provenances along the natural range of *E. grandis* using a landscape genomics approach. Genome-wide SNP analysis with the EUChip60K SNP chip allowed identification of putative adaptive genetic variation and, in some provenances, ancestry mapping revealed extensive interspecific introgression that may provide novel genetic variation for adaptation to local climate change. Next, we performed genome-wide SNP analysis of over 600 advanced generation *E. grandis* individuals from three different breeding programs and uncovered genome-wide evidence of interspecific introgression in some breeding material. Finally, large commercial F1 hybrid breeding trials create opportunities to study genomic patterns underlying hybrid combining ability, while dissecting the genetic basis of growth and resilience. Towards this, we have embarked on a large nested association mapping study in F1 hybrid breeding trials of *E. urophylla* x *E. grandis*, where we have identified trait-linked SNP markers for growth and wood properties, as well as genome-wide profiles of pre- and postzygotic barriers underlying hybrid combining ability.

#### **W452: Forest Tree**

##### **Sweet Genomes: Sequencing, Assembling, and Annotating Three Maples**

**Susan McEvoy**, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

The reference genomes for *Acer negundo* (box elder), *Acer saccharum* (sugar maple), and *Acer griseum* (paperbark maple) represent the most comprehensive characterization of the 150 member *Acer* genus to date. Box elder is widely spread across North America and is commonly used as a fast growing ornamental in urban areas, while sugar maple, valuable for both its wood and sap, represents a narrower range in Northern portions of the Eastern and Central U.S. Sugar maple populations are declining and moving North in response to abiotic stressors, while the phylogenetically distant box elder has proved resilient to similar impacts. Paperbark is a panthocarpic species used as a landscaping tree in the U.S., though native to the Yunnan province in China where populations are endangered. These genomic resources will contribute to the relatively small collection of hardwood genomes sequenced to date, and provide a basis for investigations of their adaptive potential. The sequencing design consists of deep long read coverage (90x) from Nanopore and Pacific Biosciences SEQUEL, short reads from Illumina HiSeq (150bp PE), and Hi-C data (negundo, 100x; saccharum, 65x). These diploid, highly heterozygous trees have moderate genomes, estimated at 440Mbp 590Mbp, and 452Mbp respectively. Gene annotation combined existing and novel approaches to evaluate gene prediction methods, leveraging RNA-Seq data generated for all species. Genomic comparisons among the four existing maples and other annotated land plants was used to identify putative expansions and contractions of gene families underlying the unique and shared biology of these species.

#### **W453: Forest Tree**

##### **Convergence of Polygenic Climate Adaptation in Conifers**

**Brandon Lind**<sup>1</sup>, Pooja Singh<sup>2</sup>, Sam Yeaman<sup>2</sup> and Sally N. Aitken<sup>1</sup>, (1)University of British Columbia, Vancouver, BC, Canada, (2)University of Calgary, Calgary, AB, Canada

Instances of evolutionary convergence provide a window into the genetic constraints that shape adaptation in different species. Such knowledge can be gained by examining the overlap of genetic components underlying local adaptation to climate across diverged species with environmentally heterogeneous ranges and independent evolutionary trajectories. Here we present an ambitious attempt to study local adaptation to climate across 40 range-wide populations in each of four economically

important species of North American conifers, spanning > 150 million years of independent evolution (*Pseudotsuga menziesii*, *Pinus banksiana*, *Pinus contorta*, and *Larix occidentalis*). Using genome-wide association methods (single-locus genotype-environment association and redundancy analysis) across millions of exome-targeted pool-seq SNPs in each species, we are identifying the suite of genes underlying local adaptation to climate and characterizing the architecture found within and among species. We then use these suites of climate-associated genes to quantify the observed degree of repeatability across species relative to expectations under a null model. The degree of similarity among species for the genetic components targeted by independent selection events across species provides an understanding of the genotypic redundancy underlying climate adaptation in these conifers. In addition to evolutionary processes (selection mutation, drift, migration), redundancy is an inherent property of variation in the available routes to local adaptation. Furthermore, the sequencing, assembly and annotations efforts, including our publicly available bioinformatics code, will provide an important resource for future tree research.

#### **W454: Forest Tree**

##### **The Open Chromatin Regulation of Complex Traits**

**Nathaniel W. Ellis**, Annette M. Fahrenkrog, Kelly Balmant, Daniel Conde, Christopher Dervinis, Jerald D. Noble, Marcio F. R. Resende, Jason Smith, Brad Barbazuk and Matias Kirst, University of Florida, Gainesville, FL

*Populus deltoides* (Eastern cottonwood) is a short-rotation woody crop with strong potential for bioenergy production because of its fast growth and wide adaptation to the Midwest and Southern United States. *P. deltoides* is cultivated worldwide for lumber and biomass, but development of new cultivars is hindered by lengthy breeding cycles and difficulties in phenotyping traits such as disease resistance and yield. The long-term goal of this research is to uncover and apply genomic information to guide and accelerate improvement of poplar cultivars. In order to uncover genes regulating critical traits we are characterizing a genetically unrelated population of 425 *P. deltoides* individuals. Genome-wide association studies (GWAS) have been now expanded to include growth properties measured under field conditions and disease resistance to *Sphaerulina musiva* in this population. In addition, we are using the Assay for Transposase-Accessible Chromatin combined with next-generation DNA sequencing (ATAC-seq) to uncover open chromatin regions across the *P. deltoides* genome and variants contained within them. Our hypothesis is polymorphisms within these accessible regions of the *P. deltoides* genome contribute a larger proportion of the phenotypic variance. Evidence indicates that, contrary to results obtained in species with highly methylated genomes, regions of open chromatin are not enriched for loci that control complex traits in *Populus*.

#### **W455: Forest Tree**

##### **The Hardwood Genomics Project: An Online Database for Tree Genetic and Genomic Data**

**Matthew Huff**, University of Tennessee, Knoxville, TN

As genetic and genomic data for woody tree species increases, the need for this data to be easily accessible for fellow researchers also increases. At present, much of this data is only available in raw or unannotated formats, if it is available at all. The Hardwood Genomics Website (HWG) is dedicated to addressing the growing need for a central database for woody tree genomes. Our site houses genomes and transcriptomes of trees unavailable on any other platform and also provides searchable functional annotation for genes and transcripts. To further characterize gene sequences, we house gene expression data from high throughput RNASeq experiments, allowing users to determine how their genes are affected by environmental conditions or if their genes are differentially expressed between tissues. We have 10 species with expression datasets currently represented and biocurators actively adding more. The site allows users to see genes that were statistically significant in the expression experiment and to generate heatmaps for genes of interest. We have also integrated a number of tools for researchers to easily access our data, including a powerful search engine, BLAST sequence similarity searching, and JBrowse for genome browsing. This system allows users to use data available on HWG, as well as data uploaded from their computer, as input in a workflow. Following the successful

integration of high-throughput transcriptome data, we continue to welcome new data submissions, suggestions, and partnerships to continue development. HWG is supported by NSF Awards #1444573.

#### **W456: Forest Tree**

##### **RNAi-Mediated Disease Resistance against *Septoria* (*Sphaerulina musiva*) in *Populus***

**Michael Gordon**<sup>1</sup>, Cathleen Ma<sup>1</sup>, Posy Busby<sup>1</sup>, Steven H. Strauss<sup>1</sup> and Jared M. LeBoldus<sup>2</sup>, (1)Oregon State University, Corvallis, OR, (2)Department of Plant Pathology, North Dakota State University, Fargo, ND

Productivity in hybrid poplar is threatened by the fungal pathogen *Sphaerulina musiva*—a major cause of leaf spot and stem canker (formerly *Septoria musiva*). Host induced gene silencing (HIGS) is a transgenic method whereby a host makes its own RNA-based “pesticides” that confer protection against specific insects and pathogens. It provides a new means to develop specific heritable resistance when natural sources are not available or difficult to breed for. To explore the feasibility of HIGS in the *Populus*–*S. musiva* pathosystem, we first demonstrated that *S. musiva* takes up dsRNAs from its environment using fluorescein-labeled dsRNAs and confocal microscopy. We then designed dsRNAs that would target the *S. musiva* housekeeping genes *cyp51* and *dcl*, both of which are essential for pathogen growth and/or virulence, and are homologous to effective HIGS targets in published studies. We show that *S. musiva* growth was reduced in broth culture when dsRNAs targeting *cyp51* and *dcl* was added. I will first review the state of HIGS design and effectiveness in plants, project design and goals, and then describe our results to date.

We thank the USDA Biotechnology Risk Assessment Grants program (2019-33522-30199) and an OSU Provost Fellowship for support.

#### **W457: Forest Tree**

##### **Genetic Diversity in *Populus trichocarpa* for Rare Variant Genetic Associations Querying 1,000+ Genomes**

**Anthony Piot**, Université Laval, Québec, QC, Canada

Rare small genetic variants are found in very large number across genomes and can have important effects on coding sequences. As such, these rare variants hold great potential to explain a significant part of the missing heritability observed in most genetic association studies. Given their low frequency in populations (MAF < 0.05), however, rare variant identification requires dataset of large numbers of individuals. Moreover, rare variants need to be identified with confidence, as they can be confounded with sequencing errors. In this work, we use a filtered dataset of 1,014 pure *Populus trichocarpa* to identify rare and common small genetic variants across individual nuclear genomes. We compare variant calls between two software types and applied strict quality filters for improved genetic variant identification. Finally, we retain genetic variants that were identified by both variant callers, thus increasing calling confidence. We found a high genomic diversity in *P. trichocarpa*, with 7.4 million small genetic variants. Importantly, 358k non-synonymous and 25k nonsense variants were uncovered. GO enrichment analysis showed that genes with nonsense variant were enriched in functions related to wood formation. Using RNA-seq data, we further analyze the non-synonymous variants at the transcriptomic level in order to pinpoint genetic variants located in pseudogenes. We highlight the importance of genomic diversity and rare nonsense variants in explaining more of *P. trichocarpa*'s phenotypic variability in association genetics. The goal is to associate both rare and common alleles with poplar's wood quality traits to support selective breeding for an improved bioenergy feedstock.

#### **W458: Forest Tree**

##### **Genome-Wide Association of Cold-Related Traits in Coastal Douglas-fir**

**Benjamin Wilhite**<sup>1</sup>, Daniela Puiu<sup>2</sup>, Brad St Clair<sup>3</sup>, Marc Crepeau<sup>4</sup>, Steven L. Salzberg<sup>2</sup>, Charles H. Langley<sup>4</sup>, Brian J Allen<sup>5</sup>, David B. Neale<sup>6</sup> and Amanda de la Torre<sup>7</sup>, (1)Northern Arizona University,

Flagstaff, AZ, (2)Johns Hopkins University, School of Medicine, Baltimore, MD, (3)USDA Forest Service Pacific Northwest Research Station, Corvallis, OR, (4)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (5)Department of Plant Science, University of California, Davis, Davis, CA, (6)Department of Plant Sciences, University of California, Davis, Davis, CA, (7)Northern Arizona University, Flagstaff, AZ

This study aims to combine phenotypic, genotypic, and environmental data to gain insights into the genomic basis of cold adapted traits in coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*). Genome-wide SNPs obtained from whole-genome re-sequencing of Douglas fir individuals were used to test for associations between 22 phenotypic traits related to cold damage (e.g. electrolytic leakage), growth and resource partitioning (e.g. root to shoot ratio), and phenology (e.g. budburst). Individuals were also used to test for associations between 54 environmental variables (e.g. temperature and precipitation). Population structure analysis showed that two distinct groups exist within our study zone; one existing in Southern Oregon (type 2), with the other more dominant type throughout the rest of the sampled range, up to the Canadian border (type 1). The two types hybridize, and we found a small number of hybrids in Southern Oregon. The hybrids are all more closely related to type 1 than type 2, suggesting asymmetric gene flow from type 1 to type 2. The hybrids cluster separately from type 1 in ordination space, suggesting functional differences between the genetic clusters. Most of the significant phenotype associations (163 in total) were with traits related to growth and resource partitioning such as diameter, height after year two, and root length. We did not find any associations with cold related traits. Most of the GEA results (723 in total) were associated with July maximum temperature, sun exposure and distance to sea. There were no associations in common between the GWAS and GEA results.

#### **W459: Fruit/Nuts**

##### **The Bracteatus Pineapple Genome and Domestication of Clonally Propagated Crops**

Liyu Chen, Fujian Agriculture and Forestry University, Fuzhou, China, Robert Van Buren, Department of Horticulture, Michigan State University, East Lansing, MI, Margot Paris, University of Fribourg, Fribourg, Switzerland, Hongye Zhou, University of Georgia, Athens, GA, Christian Lexer, University of Vienna, Vienna, Austria, Jeffrey L. Bennetzen, Dept. of Genetics, University of Georgia, Athens, GA and **Ray Ming**, University of Illinois at Urbana-Champaign, Urbana, IL

Domestication of clonally propagated crops such as pineapple from South America was hypothesized to be a “one-step operation”. We sequenced the genome of *Ananas comosus* var. *bracteatus* CB5 and assembled 513 Mb into 25 chromosomes with 29,412 genes. Comparison of the genomes of CB5, ‘F153’, and MD2 elucidated the genomic basis of fiber production, color formation, sugar accumulation, and fruit maturation. We also re-sequenced 89 *Ananas* genomes. Cultivars ‘Smooth Cayenne’ and ‘Queen’ exhibited ancient and recent admixture, while ‘Singapore Spanish’ supported “one-step operation” for domestication. We identified 24 selective sweeps, including a strong sweep containing a pair of tandemly duplicated bromelain inhibitors. Four candidate genes for self-incompatibility were linked in ‘F153’, but were not functional in self-compatible CB5. Our findings support co-existence of sexual recombination and “one-step operation” in domestication of clonally propagated crops. This work helps exploration of new sexual and asexual domestication trajectories in other clonally propagated crops.

#### **W460: Fruit/Nuts**

##### **Genome Assembly and Early Stage Fruit Transcriptome Analysis of Red Raspberry *Rubus idaeus***

Muzi Li<sup>1</sup>, Junhui Zhou<sup>1</sup>, Yongping Li<sup>2</sup>, Norah Hilger<sup>3</sup>, Yuwei Xiao<sup>1</sup>, Christopher Dardick<sup>4</sup>, Ann Callahan<sup>4</sup>, Chunying Kang<sup>2</sup>, Winston Timp<sup>3</sup>, Stephen M. Mount<sup>1</sup> and **Zhongchi Liu**<sup>1</sup>, (1)Dept. of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, (2)Huazhong Agricultural University, Wuhan, China, (3)Johns Hopkins University, Baltimore, MD, (4)USDA-ARS Appalachian Fruit Research Station, Kearneysville, WV

The red raspberry, *Rubus idaeus*, is a major commercial fruit valued for its taste, high antioxidant and vitamin content. However, *Rubus* breeding is a long and slow process hampered by limited genomic and

molecular resources. Genomic resources such as a complete genome sequencing and transcriptome will be of exceptional value to improve research and breeding of this high value crop. Using a hybrid sequence assembly approach with sequencing data from Nanopore, PacBio, and Illumina, we will present the first high quality *Rubus idaeus* genome (Joan J. variety). The *de novo* assembled genome has a genome size of 297,546,867 bp with a scaffold N50 of 44,297,384 and Busco genome completeness of 96.5%. Leveraging a linkage map, we anchored 80.1% of the genome onto seven chromosomes. In addition, we staged early fruit development and generated paired-end RNAseq reads from three dissected fruit tissues at six early stages of fruit development. Comparative transcriptome with strawberry fruit RNA-seq data will lead to novel insights into fleshy fruit evolution and diversity.

#### **W461: Fruit/Nuts**

##### **Deciphering the Blue Colour in Blueberry**

**Cecilia H. Deng**<sup>1</sup>, Catrin Guenther<sup>1</sup>, Declan Lafferty<sup>1</sup>, Andrew Dare<sup>1</sup>, Blue Plunkett<sup>1</sup>, Tony McGhie<sup>2</sup>, Ella Grierson<sup>2</sup>, Nick Albert<sup>2</sup> and Richard Espley<sup>1</sup>, (1)The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, (2)The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Fresh fruits and vegetables have been increasingly considered as part of a healthy diet. Blueberry fruit are small in size but packed with abundant phytochemicals, antioxidants and vitamins, contributing to their 'superfood' status. These phytochemicals include anthocyanins which provide the characteristic purple/blue pigmentation to the skin of blueberries and have been the focus of numerous human and animal health studies. The two major commercially grown types are the early-season Northern Highbush blueberry (*Vaccinium corymbosum*) and the late-season Rabbiteye blueberry (*V. virgatum*). Previous reports from highbush blueberry cultivars have shown that while there is similarity in the types of anthocyanins present, the relative concentrations of the pigments were quite different. There are relatively few reports on the anthocyanin profile from Rabbiteye blueberries, but these are increasingly being cultivated. Blueberry has deeply pigmented skin but colourless or light green flesh. To study the molecular mechanism that drives the anthocyanin accumulation and composition in blueberry, we have adopted a multi-layer '-omics approach integrating genomics, transcriptomics, comparative genomics and metabolomics. We have investigated the gene expression and compound variations in skin, flesh and whole berries across five fruit development stages of two types of blueberry, Highbush 'Nui' and Rabbiteye 'Velluto Blue'. Metabolomics data revealed distinct patterns of anthocyanin regulation in the two types of blueberry. RNA-Seq analysis identified differentially expressed genes (DEGs) involved in modifying anthocyanin profiles. Gene co-expression networks were constructed incorporating DEGs and anthocyanin compounds. Comparative genomics analysis showed more clusters of orthologues genes shared between blueberry and cranberry, compared with other fruits with colour, including raspberry, strawberry, tomato, kiwifruit, grape, apple, European and Asian pears. In addition, our research extends to genomics and transcriptomics analysis of bilberry (*V. myrtillus*), a close relative to blueberry with smaller fruit, dark purple/black skin and red/purple flesh. Our study builds knowledge in understanding the mechanisms of anthocyanin composition in different fruit tissues. We hope this resource will help the breeding of hybrid berries producing more health benefits, such as higher volumes of anthocyanins in red flesh.

#### **W462: Fruit/Nuts**

##### **Mutation of a bHLH Transcription Factor Allowed Almond Domestication**

**Raquel Sánchez-Pérez**<sup>1</sup>, S. Pavan<sup>2</sup>, R. Mazzeo<sup>2</sup>, C. Moldovan<sup>3</sup>, Riccardo Aiese Cigliano<sup>4</sup>, J. Del Cueto<sup>5</sup>, F. Ricciardi<sup>3</sup>, Federico Dicenta<sup>6</sup>, R.L. López-Marqués<sup>3</sup> and B. Lindberg Møller<sup>3</sup>, (1)CEBAS-CSIC, ESPINARDO (MURCIA), Spain, (2)University of Bari, Italy, (3)University of Copenhagen, Denmark, (4)Sequentia Biotech, Bellaterra, Spain, (5)CEBAS-CSIC, Campus Universitario de Espinardo, Spain, (6)Centro de Edafología y Biología Aplicada del Segura Consejo Superior de Investigaciones, Espinardo (Murcia), Spain

Wild almond species are toxic and bitter because of the accumulation of amygdalin, a cyanogenic glucoside present in more than 3,000 plant species. The first two steps in the biosynthesis of amygdalin is due to the expression of two P450s (*PdCYP79D16* and *PdCYP71AN24*). Almond domestication was enabled by the selection of sweet almonds lacking of the expression of these two P450s. With the sequencing of the almond genome we have identified a cluster of five *bHLH* genes, where the *Sweet kernel* locus had been previously localized. Functional characterization in *Nicotiana benthamiana* and in *Saccharomyces cerevisiae* revealed that bHLH2 controls the expression of the two P450s. A non-synonymous point mutation (Leu to Phe), in one of the fourteen amino acids implicated in the dimerization of bHLH2, blocks the transcription of *PdCYP79D16* and *PdCYP71AN24*. The immediate consequence is the absence of prunasin, the precursor of amygdalin, in the seed coat or tegument and therefore the lack of amygdalin in the cotyledon, the edible part, which now will be non-bitter and deliciously sweet.

#### **W463: Fruit/Nuts**

##### **Sequence Analysis of the Sakura Genome Toward Cherry Blossom Forecast**

**Kenta Shirasawa**<sup>1</sup>, Tomoya Esumi<sup>2</sup>, Hideki Hirakawa<sup>1</sup>, Hideyuki Tanaka<sup>2</sup>, Akihiro Itai<sup>3</sup>, Andrea Ghelfi<sup>1</sup>, Hideki Nagasaki<sup>1</sup> and Sachiko Isobe<sup>1</sup>, (1)Kazusa DNA Research Institute, Kisarazu, Japan, (2)Shimane University, Matsue, Japan, (3)Kyoto Prefectural University, Kyoto, Japan

Flowering cherry, called sakura, is Japan's unofficial national flower and is a popular ornamental tree in Japan and elsewhere. Cherry blossoms are symbols of spring, when blooming typically occurs. Accordingly, since flowering cherries are important resources for the tourism industry in the spring season in Japan, cherry blossom forecast is important for both the industry and tourists. We hypothesized that the flowering date could be predicted using the gene expression profiles in floral bud development stages. So, we determined the phased genome sequence of an interspecific hybrid, the flowering cherry 'Somei-Yoshino' (*Cerasus* × *yedoensis*) which is the most popular cultivar in Japan. The sequence data was obtained by single-molecule real-time sequencing technology and assembled using a trio-binning strategy. The resultant assembly consisting of two haplotype genomes spanned 690.1 Mb with 4,552 contigs and an N50 length of 1.0 Mb, in which totals of 95,076 high-confidence genes were predicted. A time-series transcriptome data in floral bud development stages revealed comprehensive changes in gene expression toward flowering. The information would be helpful for cherry blossom forecast.

#### **W464: Fruit/Nuts**

##### **On the Road to Molecular Breeding for Flavor: Linking Metabolomics and Genomics for Blueberry Improvement**

**Luis Felipe V. Ferrão**<sup>1</sup>, Timothy S Johnson<sup>2</sup>, Juliana Benevenuto<sup>1</sup>, Patrick Edger<sup>3</sup>, Thomas Colquhoun<sup>2</sup> and Patricio R. Munoz<sup>1</sup>, (1)Horticultural Sciences Department, University of Florida, Gainesville, FL, (2)Environmental Horticulture, University of Florida, Gainesville, FL, (3)Department of Horticulture, Michigan State University, East Lansing, MI

Fruits are important dietary sources of micronutrients for human health. Increased fruit consumption has been systematically associated with positive eating experiences, which makes flavor an essential component for consumer acceptance. Flavor is a multifactorial and subjective trait, involving a combination of multiple human senses. More specifically, it is the interaction between our olfactory system and the volatile organic compounds (VOCs) released by the food that provides distinct flavor experiences. Aiming to implement molecular breeding for blueberry flavor improvement, we integrated genomics, metabolomics, and sensory panel approaches. In this study, we demonstrate that VOCs in blueberry: i) are controlled by few major genomic regions, some of which harboring biosynthetic enzyme-coding genes; ii) can be accurately predicted using molecular markers; and iii) can enhance or deter consumer's overall liking.

Altogether, our results demonstrate how the understanding of the genetic basis and the role of VOCs on consumer preference can assist breeders in selecting more flavorful genotypes at a more inexpensively and quickly pace. We anticipate our assay to be a starting point to implement molecular breeding, opening an important venue to bring flavor improvements to the front-line of practical breeding programs.

#### **W465: Fruit/Nuts**

##### **High-Throughput Genomics in *Armeniaca* Species Reveal a Genome-Wide Impact of Domestication on the Apricot Chromosomal Organization and Gene Architecture**

**Veronique Decroocq**<sup>1</sup>, Shuo Liu<sup>2</sup>, Alexis Groppi<sup>3</sup>, Aurelien Barre<sup>4</sup>, Quynh Trang Bui<sup>5</sup>, Stephane Decroocq<sup>5</sup>, David Tricon<sup>5</sup>, Nathalie Rodde<sup>6</sup>, Sandrine Arribat<sup>6</sup>, William Marande<sup>6</sup>, Olivier Bouchez<sup>7</sup>, Céline Lopez-Roques<sup>8</sup>, Jean-Marc Aury<sup>9</sup>, Corinne Cruaud<sup>10</sup>, Caroline Belser<sup>10</sup>, Benjamin Istace<sup>10</sup>, Macha Nikolski<sup>4</sup>, Amandine Cornille<sup>11</sup> and Tatiana Giraud<sup>12</sup>, (1)UMR BFP 1332 - INRA-Université de Bordeaux, Villenave d'Ornon Cedex, France, (2)INRA, BORDEAUX, France, (3)Center of bioinformatic of Bordeaux (CBiB), Bordeaux, France, (4)CBiB Bordeaux Bioinformatics Center, Bordeaux, France, (5)INRA UMR BFP, Villenave d'Ornon, France, (6)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (7)INRAE GeT Genomics Facility, Castanet-Tolosan, France, (8)INRA Plateforme Génomique GetPlaGe, Castanet-Tolosan, France, (9)CEA - Genoscope, Evry, France, (10)Genoscope, Institut de biologie François Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France, (11)INRA - Université Paris-Sud - CNRS - AgroParisTech GQE, Gif-sur-Yvette, France, (12)Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France

Evolutionary processes differ significantly between perennial and annual species, most probably because of different life traits, i.e. the juvenility phase, extended lifespan, overlapping generations, vegetative propagation etc... In addition, under the long-term process of selection, perennial tree plants have experienced complex historical events, such as hybridization or bottleneck, together with domestication and adaptation to new environments and climatic conditions. This is expected to have impacted the perennial population dynamics but also gene and genome evolution. However, the extent of this impact is still under question.

Among the perennial species, our model plant is *Prunus armeniaca* L. (apricot). Apricot is a temperate stone fruit tree, it belongs to the family *Rosaceae*, to the genus *Prunus* and is one of the five species of the section *Armeniaca*. The species *P. armeniaca* refers to both the crop species (also called 'common apricot') and its wild progenitor which still grows in the forests of Tian Shan and Pamir Mountains in Central Asia. Taking advantage of the small size of the apricot genome (< 220 Mbp for 2n=16), we assembled *de novo* a high-quality reference sequence Through multiple local and national fundings<sup>[1]</sup>, we resequenced the complete botanical *Armeniaca* collection held at INRA Bordeaux (~400 Individuals) and part of the cultivated germplasm collection (National repository, INRA GAFL Avignon) (~200 individuals).

From this extensive resequencing data, we first focused on how selection has influenced the genomic architecture in apricot. We inferred the demographic history of *P. armeniaca* and its wild related species. To test for common or distinct signatures of selection, we took advantage of the parallel history of domestication in the European and Chinese apricots and compared with their wild, Central Asian progenitor (Liu et al, 2019, Molecular Ecology). We then detected evidence for artificial selection at a genome-wide scale, both for European and Chinese apricots, with a significant number of homologous genomic signatures of domestication, thus indicating convergent yet independent selection of a common set of genes during two geographically and culturally distinct domestication processes. We also identified signatures of selection which could be associated with local adaptation in either wild or cultivated apricots. We are currently assessing the correlation between allelic and phenotypic variation for the most interesting candidate genes. We also performed comparative genomic analysis of the *Armeniaca*

genomes assembled to date as well as of domesticated and wild apricot genomes and thus question the impact of domestication and of inter- and intra-specific (wild-to-crop) gene flow into diversification and adaptation of this long-lived perennial species.

[1] We thank the University of Bordeaux (ATT G2P “SWAGMAN” and AAP interdisciplinaire “ABXING”) and , the ANR CHEX “ABRIWG” and France Génomique “SWAG” project

#### **W466: Fruit/Nuts**

##### **Distinctive Gene Expression Patterns Define Endodormancy to Ecodormancy Transition in Apricot and Peach**

**Jiali Yu**<sup>1</sup>, Anna Conrad<sup>2</sup>, Veronique Decroocq<sup>3</sup>, Tetyana Zhebentyayeva<sup>4</sup>, Daniel E Williams<sup>5</sup>, Dennis Bennett<sup>6</sup>, Guillaume Roch<sup>7</sup>, Jean-Marc Audergon<sup>8</sup>, Christopher Dardick<sup>6</sup>, Zongrang Liu<sup>6</sup>, Albert G. Abbott<sup>9</sup> and Margaret Staton<sup>10</sup>, (1)University of Tennessee, Knoxville, Knoxville, TN, (2)The Ohio State University, Columbus, OH, (3)UMR BFP 1332 - INRA-Universite de Bordeaux, Villenave d'Ornon Cedex, France, (4)The Pennsylvania State University, University Park, PA, (5)University of Tennessee, Knoxville, TN, (6)USDA-ARS, AFRS, Kearneysville, WV, (7)CEP Innovation, Lyon, France, (8)UR1052 GAFL, INRA, Montfavet, France, (9)Forest Health Research and Education Center, University of Kentucky, Lexington, KY, (10)University of Tennessee, Institute of Agriculture, Knoxville, TN

Dormancy is a physiological state that plants enter for winter hardiness. Environmental-induced dormancy onset and release in temperate perennials coordinate growth cessation and resumption, but how the entire process, especially chilling-dependent dormancy release and flowering, is regulated remains largely unclear. We utilized the transcriptome profiles of floral buds from fall to spring in apricot (*Prunus armeniaca*) genotypes with contrasting bloom dates and peach (*Prunus persica*) genotypes with contrasting chilling requirements to explore the genetic regulation of bud dormancy. We identified distinct gene expression programming patterns in endodormancy and ecodormancy that reproducibly occur between different genotypes and species. During the transition from endo- to eco-dormancy, 1,367 and 2,102 genes changed in expression in apricot and peach, respectively. Over 600 differentially expressed genes were shared in peach and apricot, including three DORMANCY ASSOCIATED MADS-box (*DAM*) genes (*DAM4*, *DAM5* and *DAM6*). Of the shared genes, 99 are located within peach CR quantitative trait loci, suggesting these genes as candidates for dormancy regulation. Co-expression and functional analyses revealed that distinctive metabolic processes distinguish dormancy stages, with genes expressed during endodormancy involved in chromatin remodeling and reproduction, while the genes induced at ecodormancy were mainly related to pollen development and cell wall biosynthesis. Gene expression analyses between two *Prunus* species highlighted the conserved transcriptional control of physiological activities in endodormancy and ecodormancy and revealed genes that may be involved in the transition between the two stages.

#### **W467: Fruit/Nuts**

##### **Functional Genomics of Postharvest Physiology in Pome Fruits**

**Loren A Honaas**, USDA-ARS Tree Fruit Research Lab, Wenatchee, WA

Functional genomics is an emerging technological frontier in pome fruit production systems. This is driven in part by the increasing accessibility of genomics resources for specialty crops, which include genomes and transcriptomes. How pome fruit respond in the postharvest period to storage conditions, and how the at-harvest state of the fruit dictates those responses, is of particular interest for the development of risk assessment tools. Using functional genomics to understand gene activity while fruit is on the tree, through harvest, and into the often protracted storage period will help us understand postharvest fruit physiology and may reveal signatures that can be used as biomarkers. Furthermore, with a greater understanding of physiological responses to the postharvest environment, we can optimize current technology and even explore new approaches to storing pome fruit.

## W468: Fruit/Nuts

### Comparative Genomics of Six *Juglans* Species Reveals Disease-Associated Gene Family Contractions

**Alexander Trouern-Trend**, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

*Juglans*, the most speciose genus in the walnut family (Juglandaceae) represents most of the family's commercially valuable fruit and wood-producing trees and includes several species used as rootstock for their resistance to various physical and biological stressors.

We present the full structural and functional genome annotations of six members of *Juglans* and one outgroup within Juglandaceae (*Juglans regia*, *J. cathayensis*, *J. hindsii*, *J. microcarpa*, *J. nigra*, *J. sigillata* and *Pterocarya stenoptera*) using the BRAKER2 semi-supervised gene prediction pipeline and EnTAP. The sizes of these assemblies range between 641 Mb (*J. nigra*) and 992 Mb (*P. stenoptera*). For each of the 7 annotations, gene predictors were trained using 19 tissue-specific *J. regia* transcriptomes, which offered a gradient of genome-evidence phylogenetic distance. Additional evidence (EnTAP pipeline) and filters (gFACs) were applied to multiexonic and monoexonic putative genes predicted by BRAKER2 to yield between 27,000 and 44,000 high-confidence gene models per species. Comparison of gene models to the BUSCO embryophyta dataset suggested that, on average, genome completeness was 85.6%.

We utilized these annotations to assess gene family evolution within *Juglans* and between *Juglans* and selected species from across the breadth of Embryophyta. Finally, an ancient whole genome duplication that took place in a common ancestor of Juglandaceae was dated using site substitution comparative analysis.

## W469: Fruit/Nuts

### Association and Linkage Mapping of Walnut (*Juglans regia* L.) Phenological Traits

**Anthony Bernard**<sup>1,2</sup>, Annarita Marrano<sup>3</sup>, Armel Donkpegan<sup>4</sup>, Patrick J. Brown<sup>3</sup>, Charles A. Leslie<sup>3</sup>, David B. Neale<sup>3</sup>, Fabrice Lheureux<sup>2</sup> and Elisabeth Dirlewanger<sup>1</sup>, (1)INRA Bordeaux, UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon, France, (2)Ctifl, centre opérationnel de Lanxade, Prignonrieux, France, (3)Department of Plant Sciences, University of California, Davis, Davis, CA, (4)UMR 1332 BFP, INRA, Université de Bordeaux, Villenave d'Ornon, France

The walnut species cultivated for nut production, *Juglans regia* L. (Persian or English walnut), is grown worldwide in temperate areas and is one of the oldest food sources known. The walnut tree is monoecious with imperfect flowers and is diploid with  $2n=2x=32$  chromosomes. France, with a production of 40 kt in 2017, is the 2<sup>nd</sup> leading European producer behind Romania. Unraveling the genetic bases of the main agronomical traits is one of the key points for the new French walnut breeding program, and traits related to phenology are of the utmost importance in the context of climate change. To reach this goal, our study combines the two commonly used approaches: a QTL analysis based on a F<sub>1</sub> mapping progeny from a cross between two parents having contrasted phenological characteristics, and association mapping using a panel of 170 accessions selected from previous diversity and structure analyses. Phenotypic data were recorded for two years regarding budbreak, heterodichogamy, and male and female flowering. Both plant materials were genotyped using the high-density Axiom™ *J. regia* 700K SNP array recently available. Linkage map and QTL detection were performed using JoinMap® v5 and MultiQTL® v2.6, and Genome-Wide Association Study implemented with GAPIT was conducted using Best Linear Unbiased Predictors. Our results highlight significant marker-trait associations with budbreak date on Chr 1 confirmed by QTLs detected in the same genome region, and one SNP associated with heterodichogamy on Chr 11. Candidate genes are also found regarding male and female flowering processes. Then, the first KASP marker for walnut phenology has been developed and validated. These results are a breakthrough for the understanding of the genetic architecture underlying the phenology in

walnut. As far as breeding of perennial trees is a long process, these tools will be able to decrease both cost and duration in order to meet grower's need.

#### **W470: Fruit/Nuts**

##### **Flowering Networks in Hardy Pecan (*Carya illinoensis*)**

**Jennifer J. Randall**, New Mexico State University, Las Cruces, NM

Pecan [*Carya illinoensis* (Wangenh.) K. Koch] is a large deciduous tree within the Juglandaceae family. Pecan is native to North America and grows discontinuously from Oaxaca Mexico to Northern Iowa, USA. It is an economically important and popular horticultural tree crop due to its nutritious and delicious nutmeats and is planted beyond its native region within the US and the world. The largest horticultural constraint on consistent pecan production is due to alternate bearing which results in extreme interannual variability in fruiting and nutmeat quality. In pecan trees, alternate bearing is driven by fluctuations in the production of pistillate flowers which is influenced by environment x pecan genetics-physiology. In order to effectively address alternate bearing, it is essential to understand the genetic mechanisms that control flowering. Pecan trees are monoecious and heterodichogamas. Utilizing a number of horticultural studies such as the use of plant growth regulators and impacts of watering requirements on return bloom combined with several genetic tools we have been dissecting the molecular mechanisms of floral initiation in pecan trees. These genetic tools include chromosome level resolution genome sequencing/annotation of protandrous and protogynous pecan trees, GBS/GWAS studies, in-silico analyses of promoter and regulatory regions for flowering genes, RNA-Seq of flowers and buds, and quantitative reverse-transcriptase PCR of developing buds through time courses. Analyses of these data have allowed for the development of a two-step model for flowering, discovery of putative genes involved specifically in pistillate and staminate flower development, and the preliminary gene networks for flowering in pecan.

#### **W471: Fruit/Nuts**

##### **Loci Determining Trunk Caliper of Interspecific Pistachio Rootstock Identified using Phased, Chromosome-Scale Genome Assemblies**

**William J. Palmer**<sup>1</sup>, Ewelina Jaczygrad<sup>2</sup>, Salih Kafkas<sup>3</sup>, Deborah Golino<sup>2</sup>, John E. Preece<sup>4</sup> and Richard Michelmore<sup>5</sup>, (1)Genome Center, University of California, Davis, Davis, CA, (2)Foundation Plant Services, Davis, CA, (3)University of Cukurova, Adana, Turkey, (4)Nat'l Clonal Germplasm Repository, Davis, CA, (5)The Genome Center, University of California, Davis, CA

We have developed phased, chromosome-scale genome assemblies for both parents of UCB-1, the most common rootstock used in the \$1.6 billion American pistachio industry. UCB-1 is an interspecific F1 hybrid of *Pistacia atlantica* and *P. integerrima* that was initially bred to confer both Verticillium and frost tolerance. Because these species are divergent, outbreeding, and heterozygous - the resultant F1 seedling rootstocks exhibit considerable variation in both size and form that results in variation in the clonal, grafted scion. In commercial orchards, this size variation often necessitates the costly removal of small individuals and replanting. As the nut-producing scion is typically clonal *P. vera*, the genetic basis of this variation in the rootstock is of great interest for marker-development and selection of uniformly-sized trees prior to planting in orchards.

We have utilized our chromosome-scale assemblies, a mapping population of over 3,000 UCB-1 trees, phenotypic data from multiple years, and a linkage, phasing, and trait association workflow to identify loci strongly associated with trunk caliper in both commercial (grafted) and experimental (ungrafted) orchards. We leveraged short and long-read technologies including 10X Chromium, Dovetail Hi-C, and Oxford Nanopore Promethlon as well as GBS and Skim-seq sequencing strategies. The combination of phased, chromosome-scale assemblies of the parental trees, and the very large F1 mapping population from the same two trees allowed us to dissect the basis of trunk caliper and its genotype-by-environment interaction in extremely fine detail. This revealed a major effect locus associated with variation in trunk

caliper, together with several others interacting epistatically. Together these loci explain and describe size variation observed in commercial orchards.

#### **W472: Functional Annotations of Animal Genomes (FAANG)**

##### **Decrypting the Noncoding Genome**

**Bing Ren**, University of California, San Diego, San Diego, CA

#### **W473: Functional Annotations of Animal Genomes (FAANG)**

##### **Perspective from NIFA on FAANG Effort**

**Lakshmi Matukumalli**, NIFA-USDA, Kansas City, MO

#### **W474: Functional Annotations of Animal Genomes (FAANG)**

##### **Functional Annotation of the Bovine Genome**

**Pablo J. Ross**<sup>1</sup>, Hamid Beiki<sup>2</sup>, Anegla Canovas<sup>3</sup>, Sarah Corum<sup>4</sup>, Clare A. Gill<sup>5</sup>, Rui Hu<sup>6</sup>, Honglin Jiang<sup>7</sup>, Zhihua Jiang<sup>8</sup>, Chandlar Kern<sup>9</sup>, Colin Kern<sup>1</sup>, Wansheng Liu<sup>10</sup>, Pengcheng Lyu<sup>6</sup>, Wenzhi Ma<sup>9</sup>, Stephanie McKay<sup>11</sup>, Juan F. Medrano<sup>1</sup>, Jennifer J. Michal<sup>8</sup>, Brenda M. Murdoch<sup>12</sup>, James M. Reecy<sup>13</sup>, Monique Rijnkels<sup>14</sup>, Gonzalo Rincon<sup>15</sup>, Timothy P.L. Smith<sup>16</sup>, Milt G. Thomas<sup>17</sup>, Guosong Wang<sup>14</sup>, Hongyang Wang<sup>18</sup>, Yue Xing<sup>14</sup>, Xiaoqin Xu<sup>19</sup>, Xiaohui Zhang<sup>18</sup>, Yunqi Zhang<sup>18</sup> and Huaijun Zhou<sup>1</sup>, (1)Department of Animal Science, University of California, Davis, CA, (2)Department of Animal Science, Iowa State University, (3)Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (4)Zoetis, (5)Department of Animal Science, Texas A&M University, College Station, TX, (6)Department of Animal and Poultry Sciences, Virginia Tech, (7)Department of Animal and Poultry Sciences, VirginiaTech, Blacksburg, VA, (8)Washington State University, Pullman, WA, (9)Department of Animal Science, Penn State University, (10)Department of Animal Science, Penn State University, University Park, PA, (11)University of Vermont, Burlington, VT, (12)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (13)Department of Animal Science, Iowa State University, Ames, IA, (14)Texas A&M University, College Station, TX, (15)Zoetis, Kalamazoo, MI, (16)USDA, ARS, USMARC, Clay Center, NE, (17)Colorado State University, Fort Collins, CO, (18)Washington State University, (19)Animal Science, University of California, Davis, CA

The functional annotation of the cattle genome project seeks to generate high quality transcript and chromatin status data sets from a comprehensive set tissues and cells collected from Hereford cattle closely related to Dominette L1, the individual from which the reference bovine genome was sequenced, and from Holstein cows and Angus cattle. So far, transcriptomic data, including RNA-seq, smallRNA-seq, and ATTS-seq has been generated from two biological replicates of 28 adult tissues, 10 fetal tissues, and 4 cell types. Current data sets include 64 ATTS libraries with a total of 388M sequenced reads, 123 RNA-seq libraries with 4.6 billion reads, and 123 smallRNA-seq libraries with 1.9 billion reads. Analysis of these data will provide a comprehensive characterization of the expressed regions of the genome as well as accurate comparisons of differential gene expression across multiple tissues and cell types that will be harnessed for the identification of regulatory elements active in the bovine genome. Chromatin state data sets generated so far include profiling DNA methylation using WGBS in 68 adult tissues, representing 12.0 billion clean sequenced reads. Also, ChIP-seq assays are underway for profiling seven different chromatin marks including H3K27me3, H3K4me3, H3K27ac, H3K4me1, H3K36me3, H3K9me3, and CTCF. So far, 235 Libraries have been sequenced representing 20 different tissues and 3.7 billion mapped reads. Finally, conditions for generating ATAC-seq data from frozen tissues are being optimized for the identification of open chromatin profiles in tissue samples. The experimental data generated by this USDA funded project will be systematically analyzed to discover and annotate the functional elements in both coding and non-coding regions of the bovine genome, including enhancers, promoters, insulators, and small and large RNA transcripts. Data sets and findings will be made publicly available through database and genome information centers. Funded by USDA-NIFA-AFRI-2018-67015-27500 .

## **W475: Functional Annotations of Animal Genomes (FAANG)**

### **Epigenomic Annotation of Candidate *cis*-Regulatory Elements in the Chicken Genome**

**R. David Hawkins**<sup>1</sup>, Stephanie L. Battle<sup>2</sup>, Naresh Doni Jayavelu<sup>2</sup>, Lisa Griggs<sup>3</sup> and Yvonne Drechsler<sup>4</sup>, (1)University of Washington, Seattle, WA, (2)University of Washington, (3)College of Veterinary Medicine, Western University of Health Sciences, (4)College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA

*Cis*-regulatory elements (CREs) control temporal, spatial, and cell-type specific gene expression, and therefore consequently phenotype. Epigenomic annotation of genomes has greatly improved our understanding of gene regulation through the annotation of CREs and dramatically advanced our understanding of non-coding genetic variants. Applying this approach to annotate the chicken genome will catapult research in agricultural animals. Using sequencing-base approaches to map epigenetic modifications such as histone tail modifications and DNA methylation as well as regions of open chromatin will aid in uncovering the location of regulatory elements in the chicken genome. Coupling this information with the transcriptome will provide a better understanding of how expression is controlled in a tissue- or cell-specific manner. We will provide an update on our efforts to annotate CREs and the transcriptome from cells and tissues in the Michigan 6x7 F1 line.

## **W476: Functional Annotations of Animal Genomes (FAANG)**

### **Functional Annotation of the Porcine Genome- 2019 Update**

**Christopher K. Tuggle**<sup>1</sup>, Huaijun Zhou<sup>2</sup>, Catherine W. Ernst<sup>3</sup>, Crystal Loving<sup>4</sup>, James E. Koltes<sup>5</sup>, James M. Reecy<sup>6</sup>, Pablo J. Ross<sup>2</sup>, Dan Nonneman<sup>7</sup>, Timothy P.L. Smith<sup>8</sup>, Juan P. Steibel<sup>3</sup>, Wen Huang<sup>9</sup>, Haibo Liu<sup>6</sup>, Juber L. Herrera-Uribe<sup>6</sup>, Kristen Byrne<sup>10</sup>, Zhangyuan Pan<sup>11</sup>, Colin Kern<sup>12</sup> and Ryan J. Corbett<sup>13</sup>, (1)Animal Science, Iowa State University, Ames, IA, (2)Department of Animal Science, University of California, Davis, CA, (3)Department of Animal Science, Michigan State University, East Lansing, MI, (4)USDA-ARS-NADC, (5)Iowa State University, Ames, IA, (6)Department of Animal Science, Iowa State University, Ames, IA, (7)USDA, (8)U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, (9)Michigan State University, East Lansing, MI, (10)USDA-ARS-NADC, Ames, IA, (11)Animal Science, University of California, Davis, CA, (12)Animal Science, University of California Davis, Davis, CA, (13)Genetics & Genome Sciences Graduate Program, Michigan State University, East Lansing, MI

We will update the progress made since project initiation (2018). Across objectives, we are using RNA assays (RNAseq, the 5' end mapping technique RAMPAGE, and PacBio Iso-seq) and epigenetics assays (histone ChIP-seq, ATAC-seq and DNA methylation) to deeply annotate the pig genome. In Objective 1 (adult tissues), we have added six tissues to those currently being annotated using the above assays. In Objective 2 (fetal tissues and allele-biased expression), we have used RNAseq of tissues from gestation day 30 and gd70 along with fetal and parental genome sequencing from reciprocal crosses between Meishan and White cross pigs to identify allele-biased expression. We are also using histone-ChIP seq for 2 tissues and plan to test linkage between allele-biased chromatin modifications and allele-biased RNA expression. In Objective 3 (immune cells), we have isolated and performed RNAseq expression analysis of nine circulating white blood cell populations in healthy pigs, as well as macrophage responses to bacterial and viral mimics. We have also profiled gene expression of single cells in peripheral blood monocytes as well as immune tissues. We have performed whole genome bisulfite sequencing (WGBS) analysis in 8 samples to assess differential and allele-specific methylation patterns in fetal liver, and initiated WGBS analysis in isolated immune cells. In Aim 4 (integration), we have found a significant association between changes in H3K27ac modification at active promoters and differential expression upon stimulation of alveolar macrophages. Overall, we have completed 43% datasets out of the total of 725 proposed. Supported by USDA-NIFA-AFRI-2018-67015-27501.

## **W477: Functional Annotations of Animal Genomes (FAANG)**

## **Genome-Wide Identification and Annotation of Functional Regulatory Regions in the Chicken, Cattle, and Pig Genomes**

**Colin Kern**, Department of Animal Science, University of California, Davis, CA

Only a small fraction of animal genomes is known to be transcribed, however the non-transcribed regions play key roles in gene regulation, and therefore significantly impact phenotypic traits. Epigenetic marks are a primary factor in identifying these regulatory regions, and while consortia such as ENCODE have made great progress in generating epigenetic data in human, mouse, and other model organisms, very little has been done in farm animals where researchers would benefit greatly from such data by improving the understanding of economically important traits. The Functional Annotation of ANimal Genomes (FAANG) consortium was created to coordinate epigenomics research in domesticated animals. As one of the FAANG pilot projects, we have completed the first comprehensive identification of regulatory elements in farm animals which included eight assays profiling the transcriptome, four histone modifications, CTCF, DNA methylation (RRBS) and open chromatin across eight tissues in the chicken, cattle, and pig genomes. Integration of these data produced genome-wide chromatin state predictions resulting in catalogs of promoters, enhancers, insulators, and polycomb repressed regions for each tissue in each species. These data and results are publicly available on the FAANG data portal and viewable on genome browsers via a UCSC track hub. The computational pipeline has also been made available on GitHub for use by the research community. These datasets, which are being expanded by ongoing FAANG-related projects, provide a resource for further studies that will improve the understanding of complex traits and the evolution of regulatory elements.

### **W478: Functional Annotations of Animal Genomes (FAANG)**

#### **Ovine FAANG Update**

**Brenda M. Murdoch**<sup>1</sup>, Stephen N. White<sup>2</sup>, Michelle R. Mousel<sup>2</sup>, Alisha Massa<sup>3</sup>, Kimberly M. Davenport<sup>1</sup>, Kim C. Worley<sup>4</sup>, Alan L. Archibald<sup>5</sup>, Emily Clark<sup>5</sup>, Brian Dalrymple<sup>6</sup>, James W. Kijas<sup>7</sup>, Shannon M. Clarke<sup>8</sup>, Rudiger Brauning<sup>9</sup>, Timothy P.L. Smith<sup>10</sup>, Tracy Hadfield<sup>11</sup> and Noelle Cockett<sup>11</sup>, (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (3)Washington State University, (4)Baylor College of Medicine, Houston, TX, (5)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (6)Institute of Agriculture, The University of Western Australia, CRAWLEY, Australia, (7)CSIRO Animal, Health and Food Science, St Lucia, Australia, (8)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand, (9)AgResearch Limited, Invermay Agricultural Centre, Mosgiel, New Zealand, (10)USDA, ARS, USMARC, Clay Center, NE, (11)Utah State University, Logan, UT

The goal of the Ovine FAANG project is to facilitate a greater understanding of the complex nature of gene regulation within this globally important food and fiber species. Compilations of functional assays have been used to characterize and define the multifaceted biological mechanisms that contribute to gene regulation. Specifically, we examined coding and non-coding transcript isoforms, alternative splicing, promoters and cis-acting regulatory elements, open chromatin, histone modifications, and DNA methylation across a wide range of sheep tissues. To date we have mRNA short read sequencing data on 60 tissues, long read data on 8 tissues, and micro RNA on 30 tissues. To complement this gene expression data, cap analysis of gene expression (CAGE) data for 56 of these same tissues was used to confirm transcription start sites across the ovine genome. Histone modification sequencing results using H3K4me3, H3K27ac, H3K4me1 and H3K27me3 marks are currently being analyzed for a subset of 47 of these tissues. Other assays currently underway include ATAC-Seq assays to assess chromatin accessibility, and whole genome and reduced representation bisulfite sequencing to determine DNA methylation status. Overall, this project will provide one of the highest resolution annotations of the reference genome of a livestock species. The resources we have generated are foundational. More specifically, this expands our understanding of how gene-regulation controls phenotypic plasticity in this economically important livestock species.

## **W479: Functional Annotations of Animal Genomes (FAANG)**

### **Equine FAANG: Off to the Races**

**Carrie Finno**, University of California Davis, Davis, CA, Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, Ted Kalbfleisch, University of Louisville, Louisville, KY and Jessica L. Petersen, University of Nebraska-Lincoln, Lincoln, NE

Our overall objective is to create a tissue- and sex-specific atlas of gene expression and regulation in the horse, as part of the Functional Annotation of ANimal Genomes (FAANG) initiative. To date, we have completed a set of functional annotation data in two healthy adult mares. Data includes whole-genome sequencing, transcriptome profiling of 50 tissues, characterization of four histone marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) and CCCTC-insulator marks, and reduced representation bisulfite sequencing on eight prioritized tissues. Two tissues, liver and laminae, were used to compare ATAC-seq preparations between nuclei extracted from frozen tissue versus nuclei prepped at the time of tissue collection. These newly annotated genomic elements are being used in studies of equine-specific phenotypes as well as in cross-species studies. Ongoing work is extending data collection and genome annotation to two Thoroughbred stallions, which will enable comparisons between sexes.

## **W480: Functional Annotations of Animal Genomes (FAANG)**

### ***Functional Annotation of the Rainbow Trout Genome***

**Mohamed (Moh) Salem**, University of Maryland, College Park

With sustained high rates of human population growth and depletion of natural fisheries resources, the US seafood imports are expected to continue to increase sharply. Aquaculture is the world's fastest-growing agri-food business. Rainbow trout is the most cultivated cold-water fish in the US. The US aquaculture industry has recently begun to implement applications of genomics in breeding programs to define genetic and phenotypic parameters that control complex traits such as disease resistance and feed efficiency. The FAASG international consortium "Functional Analysis of All Salmonid Genomes" has been established to study the functional genomic basis of phenotypic variation in all salmonids including rainbow trout (<https://www.faasg.org/>). A trout reference genome sequence is now available (GenBank Accession GCA\_002163495). However, major improvements in genome reference and annotation can improve the accuracy and efficiency of the genomic applications such as GWAS, identification of causative variations, marker assisted selection and genomic selection in rainbow trout. Therefore, the objectives of our current USDA-supported project are: 1) Closing the Swanson reference genome assembly by utilizing the long-read PacBio sequencing platform in combination with the BioNano whole genome optical map. 2) Annotate the reference genome for the coding and non-coding transcript isoforms and alternative splicing, by full-length single-molecule sequencing. 3) Annotate the genome for chromatin histone modifications and chromatin accessibility by integrating data from RNA-seq, DNase-seq, and ChIP-seq across a wide range of rainbow trout tissues. The improved assembly and annotation of the reference genome will accelerate the genetic selection efforts, particularly through GWAS and genomic selections, for improving important production traits in rainbow trout.

## **W481: Functional Annotations of Animal Genomes (FAANG)**

### **The EU-H2020 Project GENE-Switch (The regulatory GENome of SWine and CHicken: functional annotation during development)**

**Elisabetta Giuffra**, INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France and GENE-SWitCH Consortium

The H2020 project GENE-SWitCH started in July 2019 and will span 4 years. It aims to deliver new underpinning knowledge on the functional genomes of two main monogastric farm species (pig and chicken) and to enable immediate translation to the pig and poultry sectors. The activation status of functional genome sequences varies across time and space, and in response to environmental perturbations. In full coordination and synergy with global effort and ongoing projects of the FAANG community, we will characterize the dynamics ("switches") of the functional genome from embryo

(chicken) and fetus (pig) to adult life by targeting a panel of tissues relevant to sustainable production. New expression QTL data in pigs and existing high-resolution QTL data in chicken will be used for developing innovative genomic predictive models that integrate functional annotations, and these models will be validated in commercial pig and poultry populations. In addition, nutritional epigenetic data will allow evaluation of the influence of maternal diet on the epigenome of the pig fetus and whether such effects persist until post-weaning. These open-shared datasets will conform fully with FAANG standards and add valuable knowledge on genetic and epigenetic variation of functional elements to FAANG. A comprehensive plan of dissemination and outreach activities to a large audience of stakeholders will be implemented. The GENE-SWitCH consortium brings together partners representing pan-European excellence (including the academic institutions which pioneered FAANG) and world-leading animal breeding and biotech industry in a true co-creation effort. Overall, GENE-SWitCH will contribute to the global FAANG effort considerably, demonstrate how functional annotation of genomes can foster the advancement of genomic selection for immediate benefit to the breeding industry, and produce cutting-edge research paving the way to new studies and strategies for sustainable productions. This talk will provide an update of activities carried out in the first 6 months of the project.

## **W482: Functional Annotations of Animal Genomes (FAANG)**

### **BovReg (Cattle) Update**

**Christa Kuehn**<sup>1</sup>, Jens Vanselow<sup>2</sup>, Cedric Notredame<sup>3</sup>, Dominique Rocha<sup>4</sup>, Didier Boichard<sup>4</sup>, Didier Allaer<sup>5</sup>, Carole Charlier<sup>6</sup>, Hubert Pausch<sup>7</sup>, Yvette deHaas<sup>8</sup>, Mogens Sandø Lund<sup>9</sup>, Johanna Vilkki<sup>10</sup>, Hiroaki Taniguchi<sup>11</sup>, Franck Meijboom<sup>12</sup>, Daniel R. Zerbino<sup>13</sup>, Andrea Rosati<sup>14</sup>, Graham S. Plastow<sup>15</sup>, Emily L. Clark<sup>16</sup>, James Prendergast<sup>17</sup>, Ann Bruce<sup>17</sup>, Marion Schmicke<sup>18</sup>, Amanda J. Chamberlain<sup>19</sup>, Hans D. Daetwyler<sup>19</sup>, Veronique Blanchet<sup>20</sup>, Andreia Fonseca<sup>21</sup> and Donald Bruce<sup>22</sup>, (1)Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, (2)Leibniz Institute for Farm Animal Biology (FBN), Germany, (3)Center for Genomic Regulation, Barcelona, Spain, (4)INRA UMR1313 Animal Genetics and Integrative Biology, Jouy-en-Josas, France, (5)Diagenode, Belgium, (6)5GIGA, Université de Liège, Belgium, (7)ETH Zurich, Zurich, Switzerland, (8)Wageningen Livestock Research, Netherlands, (9)Aarhus University, Tjele, Denmark, (10)Natural Resources Institute Finland, Jokioinen, Finland, (11)Institute of Genetics and Breeding, Poland, (12)University Utrecht, Netherlands, (13)European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, (14)European Federation of Animal Science, Italy, (15)Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (16)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (17)The University of Edinburgh, United Kingdom, (18)Faculty of Natural Sciences III, Martin-Luther Universität Halle-Wittenberg, Halle, Germany, (19)Agriculture Victoria, Bundoora, Australia, (20)University of Limoges, France, (21)FMVUL, Lisboa, Portugal, (22)Edinethics Ltd, United Kingdom

Despite the revolution in functional genome analysis a wide gap in understanding associations between the (epi)genome and complex phenotypes of interest currently remains and impedes efficient use of annotated genomes for precision breeding. As part of the global FAANG initiative and funded by a recent EU H2020 project call, the BovReg consortium is set up to provide a comprehensive map of functionally active genomic features in cattle and how their (epi)genetic variation in beef and dairy breeds translates into phenotypes. This constitutes key knowledge for biology-driven genomic prediction needed by scientific and industry livestock communities. BovReg brings together a critical mass of experts in ruminant research and beyond encompassing bioinformatics, molecular and quantitative genetics, animal breeding, reproductive physiology, ethics and social science. A total of 20 partners from the EU, UK, Switzerland, Canada and Australia form a global interdisciplinary team, which builds upon previous and ongoing national and EU-funded research and collaborations with industry partners. BovReg is set up to generate functional genome data based on FAANG core assays from representative bovine tissues and newly established cell lines covering different ontological stages and phenotypes applying novel bioinformatic pipelines in a standardized and reproducible fashion. Key traits for BovReg are phenotypes related to robustness, health and biological efficiency. Data, metadata, knowledge and protocols will be deposited in European biological archives, aiming to set up and maintain a knowledge hub and establish

gold standards. Long-term availability of data, methods, targeted dissemination and communication activities are guaranteed by EMBL-EBI, FAANG and EAAP and will follow ELIXIR guidelines. BovReg will develop biology-driven genomic prediction tools by integrating biological knowledge of regulatory genomic variation and genomic selection methods for local and global cattle populations. The results will enable more environmentally sustainable cattle production, while respecting animal-welfare.

### **W483: Functional Annotations of Animal Genomes (FAANG)**

#### **AQUA-FAANG (Fish species) Update**

**Ross D. Houston**, The Roslin Institute, Edinburgh, United Kingdom

Genomic tools have been underpinned many recent developments in research and commercial innovation in European aquaculture. However, our understanding of the functional genomic basis for commercially important phenotypic variation (i.e. growth rate, disease resistance, etc.) remains limited, limiting our ability to exploit the predictive ability of genetic information. "AQUA-FAANG", funded by the European Commission H2020 program, is tackling this knowledge gap, and aims to deliver a step improvement in understanding of genome function and the exploitation of genotype-to-phenotype prediction in the six most important European farmed fish species, Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*), European seabass (*Dicentrarchus labrax*), gilthead seabream (*Sparus aurata*), turbot (*Scophthalmus maximus*) and common carp (*Cyprinus carpio*), which together account for >90 % commercial finfish production in Europe.

The consortium comprises an interdisciplinary team of academics and industry partners from eight European countries: Norway, the United Kingdom, France, the Netherlands, Spain, Italy, Greece and Poland. AQUA-FAANG is one of three projects funded under the same H2020 Research and Innovation Action, with the others focussed on terrestrial livestock. The AQUA-FAANG project will functionally annotate the genomes of the six targeted species, employing standardized assays and pipelines defined by FAANG. The goal is to document genome-wide functional and regulatory features under distinct biological conditions, including chromatin accessibility (by ATAC-Seq), enhancer and promoter activity (by ChIP-Seq), and protein and non-coding gene expression (by RNA-Seq). This will be done across distinct tissues and development stages in healthy animals, in addition to immune-activated and disease-challenged states, addressing the need to tackle the huge threat posed to aquaculture by infectious disease. Data from the project will be made publically available in an easy-to-visualize format via the Ensembl genome browser.

The project includes applied research activities that aim to establish functional mechanisms underpinning resistance to infectious disease, and enhance the accuracy of genomic prediction of disease resistance. For example, marker panels enriched for prioritized causative genetic variants and novel statistical methods will be developed to improve the accuracy of genomic selection using functional annotation data.

### **W484: Functional Annotations of Animal Genomes (FAANG)**

#### **The FAANG Data Coordination Centre: New European Perspectives for Our Continued Global Effort**

**Peter W Harrison**, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom

The Functional Annotation of Animal Genomes (FAANG) Project is a coordinated international effort to produce and collate high quality functional annotation of livestock genomes. The FAANG Data Coordination Centre (DCC) at EMBL-EBI has been developing the core infrastructure to support the global community to create this rich genome to phenome resource over a number of years. Three new European FAANG Horizon2020 projects started in the last year focussing upon cattle (BovReg), chicken and pigs (GENE-SWitCH) and Salmonids (AQUA-FAANG), that collectively provide a focal point for European FAANG research and are each supporting continued development of the FAANG DCC at EMBL-EBI. This is driving new requirements and development of FAANG metadata standards, validation

and submission software and the FAANG data portal, of importance to the global FAANG community. The FAANG DCC also continue to be guided and active in the global FAANG steering committee and working groups.

FAANG has a particular focus on ensuring high quality and rich supporting metadata to describe the project's samples and experimental assays. The DCC is launching a new improved metadata validation service that is hosted within the FAANG data portal (<http://data.faang.org/home>), rather than existing as a separate site, that will continue to ensure rich data submissions to the public archives. The documentation is also being shifted to the data portal so that all FAANG submission and retrieval data services are accessible from a single site. The data portal acts as a single access point for the wealth of livestock functional annotation data available from FAANG contributors combined with existing data available from public archives imported under legacy standards. The portal provides direct links for downloading data files direct from the public archives, with web-based bulk download support recently added, and supports programmatic API access.

The DCC will continue to support the community with a dedicated helpdesk ([dcc-faang@ebi.ac.uk](mailto:dcc-faang@ebi.ac.uk)) and is currently developing new software to assist the community in improving the suitability of ontologies for use in livestock context. Through effective standard driven metadata validation, a powerful search driven data portal and promotion of best practice in metadata implementation, the FAANG DCC aims to maximise effectiveness and inter-comparability of assay data, supporting the community to create a rich genome to phenome resource.

## **W485: Functional Annotations of Animal Genomes (FAANG)**

### **A Vision for Bioinformatics within the Global FAANG Project**

**Peter W Harrison**, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom and **Michael Watson**, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom

The global FAANG project now consists of hundreds of researchers and multiple, separately funded projects, loosely co-ordinated by committees, mailing lists and informal communications. It has been an incredible success so far.

One of the challenges of such a loosely co-ordinated project is to ensure reproducible research across multiple projects and species. Luckily, a suite of technologies now exist that enable reproducible research in bioinformatics and computational biology, at any scale. Workflow managers, software environments, software containers and cloud technologies ensure that we can all access the exact same pipelines, and ensure maximum reproducibility.

Here we propose a vision for shared development of bioinformatics pipelines across the entire set of FAANG (and associated) projects, based on the principles of open science, open source code and reproducible workflows and environments

## **W486: Functional Annotations of Animal Genomes (FAANG)**

### **Annotation of Sequence Variants in the Bovine Genome with the Functional-and-Evolutionary Trait Heritability (FAETH) Score**

**Ruidong Xiang**<sup>1</sup>, Irene van den Berg<sup>2</sup>, Iona M MacLeod<sup>3</sup>, Ben J. Hayes<sup>4</sup>, Claire P. Prowse-Wilkins<sup>5</sup>, Min Wang<sup>6</sup>, Bolormaa Sunduimijid<sup>3</sup>, Zhiqian Liu<sup>2</sup>, Simone Rochfort<sup>7</sup>, Coralie M. Reich<sup>3</sup>, Brett A. Mason<sup>3</sup>, Christy J Vander Jagt<sup>3</sup>, Hans D. Daetwyler<sup>3</sup>, Mogens Sandø Lund<sup>8</sup>, Amanda J. Chamberlain<sup>3</sup> and Michael E. Goddard<sup>5</sup>, (1)University of Melbourne, Bundoora, VIC, Australia, (2)Agribio, Bundoora, VIC, Australia, (3)Agriculture Victoria, Bundoora, Australia, (4)University, Bundoora, VIC, Australia,

(5)The University of Melbourne, Parkville, Australia, (6)La Trobe University, Bundoora, Australia, (7)Agriculture Victoria, Bundoora, VIC, Australia, (8)Aarhus University, Tjele, Denmark

The Functional Annotation of ANimal Genomes consortium provides comprehensive functional genomic information on farm animals. While many functional datasets are generated with different assays, it is unclear which types help identify causal genes. Moreover, it is unclear how to effectively analyse diverse functional datasets to locate causal mutations that will benefit genomic selection and precision animal breeding. We propose a framework which ranks millions of cattle sequence variants based on their 'Functional-And-Evolutionary Trait Heritability' (FAETH) score. We first collected 30 categories of functional and evolutionary data in cattle and other species, including metabolic quantitative trait loci (QTLs), splicing QTLs, ChIP-seq peaks, selection signatures and conserved sites across 100 vertebrate species. Then, we partitioned genetic variance of 34 cattle traits into these functional categories with genome-wide restricted maximum likelihood analysis in over 44,000 Australian dairy bulls and cows with 17.7 million imputed sequence variants. Based on the per-variant trait heritability and the variant memberships to different functional categories, we rank both the importance of different functional datasets and individual variants. Validated in over 7,700 Danish cattle, the high FAETH-ranking variants showed significantly increased genetic variances and genomic prediction accuracies. Our study provides 1) proof-of-concept evidence that functional genomics can improve genomic selection in global cattle breeds; 2) novel and effective methods for integrating functional data into genome-wide analysis of sequence variants and 3) the publicly available FAETH score of 17.7 million cattle variants which can be used as biological prior for genomic selection or new functional annotation resources.

#### **W487: Functional Genomics**

##### **Molecular Cross Talk between Resistant and Defense Genes Promote Resistance to *Heterodera Glycines***

**Naoufal Lakhssassi**, Southern Illinois University at Carbondale, Carbondale, IL and Sarbottam Piya, Sadia Bekal, Shiming Liu, Zhou Zhou, Catherine Bergounioux, Long Miao, Jonas Meksem, Aicha Lakhssassi, Karen Jones, My Abdelmajid Kassem, Moussa Benhamed, Abdelhafid Bendahmane, Kris Lambert, Adnane Boualem, Tarek Hewezi, and Khalid Meksem

The serine hydroxymethyltransferase (SHMT, E.C. 2.1.2.1) catalyzes the interconversion of serine/glycine as well as tetrahydrofolate (THF) and 5,10-methyleneTHF. Soluble NSF attachment proteins ( $\alpha$ -SNAP) are conserved across animal and plant kingdoms and are involved in vesicular trafficking, cytokinesis, and plasma membrane repair and stability. Both GmSNAP18 and GmSHMT08 protein members have been described previously to play a major role in SCN resistance. Expression analysis of *GmSHMT08* and *GmSNAP18* identified the need of a minimum expression requirement to trigger the SCN resistance reaction. In the current study, we discovered the presence of a molecular interaction between the GmSHMT08 and GmSNAP18 proteins that is potentiated by the presence of a pathogenesis related protein GmPR10-08. GmPR10-08 was mapped to a novel quantitative trait locus for broad resistance to SCN using two different mapping populations. GmPR10-08 was induced in response to SCN infections and, like GmSHMT08 and GmSNAP18, PR10-08 was found to localize in the cytosol. Overexpression of GmPR10-08 decreased the number of SCN cysts by nearly 65% in transgenic soybean roots. Pathogenesis related proteins (PRs) are widely present in plants being induced following pathogen attack, elicitors, wounding, or stress, and are toxic to invading fungal pathogens. Immunostaining and in-situ assays demonstrated that GmSNAP18 expression and localization hyper-accumulated at the plasma membrane and was specific to the root cells surrounding the nematode in SCN resistant soybean, but not in the SCN susceptible line, indicating the involvement of GmSNAP18 in molecular trafficking. Genes encoding key components of the salicylic acid signaling pathway were induced under SCN infection. *GmSNAP18* and *GmPR08-Bet VI* were also induced under salicylic acid and cytokinin exogenous treatments, while *GmSHMT08* was induced only when the resistant *GmSNAP18* was present. These data point to the presence of a molecular crosstalk between SCN resistant genes and plant defense genes. Due to the involvement of these three proteins in human diseases including cancers, the finding revealed in this manuscript may have widespread implications within the field of biology and pharmacogenomics, paving the way for novel therapeutics.

## **W488: Functional Genomics**

### **NCBI Refseq Resources for Plant Genomics**

**Anjana Raina Vatsan**, Francoise Thibaud-Nissen and Terence D. Murphy, NCBI/NLM/NIH, Bethesda, MD

National Center for Biotechnology Information (NCBI) generates genome annotation data for a wide variety of plants. It is a non-redundant dataset consisting of coding and non-coding transcripts, associated with a Gene record. There are currently 107 plant species annotated using the NCBI's Eukaryotic Genome Annotation Pipeline. Few of these annotated genomes are in scope for manual curation. Manual curation ensures accurate and full-length representation of nucleotide and protein sequences and helps resolve data conflicts and ambiguities. Gene and protein names are assigned, and publications added, when available. It provides for a more accurate and enriched data set.

Data is available through our Gene resource, RefSeq database and a powerful genome browser called the Genome Data Viewer (GDV). It is also accessible via BLAST and Entrez. The full list of plants that NCBI has annotated is available at: [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/all/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/).

Data can be downloaded from our FTP directory at <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/> or accessed from the RefSeq home page at <https://www.ncbi.nlm.nih.gov/refseq/>.

## **W489: Functional Genomics**

### **High Resolution Genome Wide Methylation Profiling of Olive using Methylated DNA Immunoprecipitation followed by Sequencing**

**Oussama Badad**, University Mohamed the Fifth, Rabat, Morocco, Rabat, Morocco

DNA methylation is an epigenetic print involved in the regulation of multiple biological processes. To elucidate this involvement, especially in important pathways such as oil biosynthesis, ripening and alternate bearing, a genome wide epigenomic profiling was embraced, using a fast forward and Low input DNA and cost effective technique called Methylated DNA immunoprecipitation-sequencing (MeDIP-seq). Genomic DNA was fragmented to 100-500bp fragments by sonication. DNA-end was repaired to overhang a 3' -dA, then adapters were ligated to the end of DNA fragments. Double-stranded DNA was denatured, Then DNA fragments were immunoprecipitated by 5mC antibody, Real-time PCR was used to validate the quality of immunoprecipitation. DNA fragments with size 200-300bp were selected after PCR amplification. The sequencing was performed by an Illumina HiSeq 2000 platform, generating 63 to 65M read of Raw data per sample. The raw reads were subject to quality control, trimming and removal of low quality read then mapped to Oleaster reference genome using BWA. The resulting BAM files undergo a MeDIP-seq quality control using MEDIPS R package showing that the set of reads is sufficient to saturate the entire genome. Differential peak calling was performed using MACS tool and the region associated with peaks were annotated using HOMER. The genome wide epigenomic profiling uncovered the involvement of DNA methylation in

many biological processes in the olive genome specially the most important one “The oil biosynthesis pathway”

**W490: Functional Genomics of Complex Traits in Diverse Organisms: the Next Leading EDGE (Enabling Discovery through GENomic Tools)**

**TBD**

**Edward S. Buckler**, Institute for Genomic Diversity, Cornell University, Ithaca, NY

**W491: Functional Genomics of Complex Traits in Diverse Organisms: the Next Leading EDGE (Enabling Discovery through GENomic Tools)**

**Unraveling the Complex Pathways Underpinning Honey Bee Social Behavior and Health using Genomics**

**Christina Grozinger**, Penn State University/College of Agricultural Sciences/Huck Institutes of the Life Sciences, University Park, PA

Honey bees are an outstanding model system in which to study the proximate and ultimate mechanisms mediating complex social behaviors. They are also critical pollinators of flowering plants, and reports of global declines in wild and managed bees have catalyzed efforts to understand the factors which both undermine and support bee health. Genomic approaches have helped us characterize the molecular, physiological, and ecological factors that underpin social behavior and health in honey bees and related species. Genomic tools, including RNAi and gene editing technologies, also provide opportunities to directly modify these processes to better characterize these mechanisms and potentially improve bee health outcomes. However, there are many technical, biological and social constraints which reduce the tractability of these applications.

**W492: Functional Genomics of Complex Traits in Diverse Organisms: the Next Leading EDGE (Enabling Discovery through GENomic Tools)**

**Juggling Genes in *Prochlorococcus*: Discoveries, Challenges, and Opportunities**

**Giovanna Capovilla**, MIT, Cambridge, MA and Raphaël Laurenceau, Zev Cariani, Christina Bliem, Thomas Hackl, Markus Ankenbrand, Sallie W. Chisholm

The marine cyanobacterium *Prochlorococcus* is the smallest and most abundant primary producer in the oceans. It has a very small genome – on average 2000 genes - but the pangenome of the global collective is estimated to be roughly 80,000 genes. Progress in understanding the functions embedded in this vast diversity has been hampered by the lack of genetics. Efforts to demonstrate efficient conjugation or transformation have historically been unsuccessful. Recently, some success has been achieved using electroporation to introduce DNA into cells but efforts to optimize the DNA delivery failed to generate transformants; stable integration of foreign DNA into the chromosome was not achieved.

While prophage and other common mobile genetic elements are not found in *Prochlorococcus*, we have recently discovered self-replicating integrative elements that contain putative integrase genes and appear to insert at specific sites within the *Prochlorococcus* genome – affording opportunities for the introduction of genetic tools. To this end we have synthesized modified elements that incorporate an antibiotic resistance gene regulated by a strong, *Prochlorococcus*-specific promoter. We plan to introduce the element into the cell using our recently optimized electroporation technique. If successful, we will engineer the synthetic elements for use in knock-in or knock-out experiments using transposons or a CRISPR-Cas system. Questions we are interested in addressing with this system include knocking out the production of specific secondary compounds, exploring the countless genes of unknown function and determining the absolute minimal photosynthetic cell.

## **W493: Functional Genomics of Complex Traits in Diverse Organisms: the Next Leading EDGE (Enabling Discovery through GENomic Tools)**

### **Expanding Functional Genomics to the *Daphnia pulex* Model System**

**Michael Lynch**, Biodesign Center for Mechanisms of Evolution/School of Life Sciences/ASU, Tempe, AZ

The microcrustacean *D. pulex* has long been exploited for studies in limnology and physiological ecology. With its short generation time; an ability to reproduce clonally, by selfing, and by outcrossing; a well-understood ecology; and thousands of genomes now sequenced; this system provides an excellent platform for integrating functional biology into a genomic, evolutionary, and ecological context. To make this possible, we are embarked on: 1) the development of methods for gene transformation; 2) the establishment of clonal constructs with chromosomal landing pads for controlled insertions; 3) the production of reagents for carrying out these tasks; and 4) the production of an atlas of tissue-specific gene expression. An overview will be provided on where this work is headed.

## **W494: Functional Genomics of C<sub>4</sub> and CAM photosynthesis**

### **Comparative Evolutionary Genomics of C<sub>4</sub> Evolution: A Wealth of Possibilities**

**Rowan F. Sage**, Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, Canada, **Matt Stata**, University of Toronto, Toronto, ON, Canada, **Tammy L. Sage**, Department of Ecology and Evolutionary Biology University of Toronto, Toronto, ON, Canada, **Xinguang Zhu**, Center of Excellence of Plant Molecular Sciences, Institute of Plant Physiology and Ecology, CAS, Shanghai, China and **Ming-Ju Amy Lyu**, National Key Laboratory for Molecular Plant Genetics, Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, China

With over 65 independent origins, mother nature provides many cases of C<sub>4</sub> evolution that researchers can exploit to address when, where and how the C<sub>4</sub> photosynthetic pathway evolved. We have assembled a diverse collection of C<sub>4</sub> species representing over twenty independent C<sub>4</sub> lineages, including 12 lineages with species that exhibit intermediate traits between fully C<sub>3</sub> and fully C<sub>4</sub> character states. Many of these intermediates have been newly characterized in species from our collection, providing additional examples to understand the steps involved in C<sub>4</sub> evolution. Newly-identified intermediates in the genus *Blepharis* (Acanthaceae) are particularly informative as they exhibit a highly stepwise phylogenetic and phenotypic progression from a near-C<sub>3</sub> to near-C<sub>4</sub> condition. With these living resources, we have great potential to develop genomic platforms to further address how C<sub>4</sub> evolved. Through phylogenomics, we can clarify species relationships and the evolutionary histories of genes important to C<sub>4</sub> function. Transcriptomes now in hand identify patterns of gene expression in relation to phenotype, and full genome assemblies in progress will allow us to examine evolutionary diversification of genes, chromosomes and patterns of gene regulation during C<sub>4</sub> evolution. Using distinct C<sub>4</sub> lineages within *Blepharis*, we will present recent results from phylogenomic and genome analyses that allow us to address parallel evolution of genes that have been separately co-opted into the C<sub>4</sub> pathway. By doing so, we can exploit the multiple C<sub>4</sub> origins to address the broader issue of how convergent is evolutionary convergence at the genome level.

## **W495: Functional Genomics of C<sub>4</sub> and CAM photosynthesis**

### **Unraveling Crassulacean Acid Metabolism (CAM) Induction in Facultative CAM Species *Talinum triangulare*: A Transcriptome Analysis**

**Eva Maleckova**, Heinrich Heine University, Duesseldorf, Germany, **Dominik Brilhaus**, Heinrich Heine University Duesseldorf, Germany and **Andreas P. M. Weber**, Heinrich Heine University, Düsseldorf, Germany

Crassulacean acid metabolism (CAM) has evolved as one of the carbon concentrating mechanisms and at the same time is the most water-use efficient of the three modes of photosynthetic carbon assimilation. CAM is thus not only an important ecophysiological adaptation of plants inhabiting arid

regions but also offers an alternative strategy for crop improvement, e.g. enabling food and/or biomass production on marginal agricultural land. While the CAM pathway is understood well biochemically, the genetic blueprints for a functional CAM pathway are just being unravelled and that is where we aim to contribute.

In our research, we employ the facultative CAM species *Talinum triangulare*, in which CAM pathway is induced reversibly as an adaptation to drought periods. Our work revealed that CAM induction can also be achieved by exogenous abscisic acid (ABA) in a controlled and very rapid manner. A time-course mRNA sequencing experiment enabled us to describe changes at the level of transcript abundances during the initial 21 hours of the induction processes. We observed not only transcript accumulation of key CAM enzymes, such as phosphoenolpyruvate carboxylase, but also identified a number of genes with altered temporal patterns in response to ABA.

We thus propose *Talinum triangulare* as a suitable model system for further investigation of the CAM pathway and identification of the minimal set of genes required for a fully functional CAM pathway. With growing genomic resources for both C<sub>3</sub> and CAM species of different evolutionary origins and operating in a variety of CAM types, comparative studies are possible. At the moment, we focus on the coding part of *Talinum triangulare* genome and performed full-length mRNA sequencing (isoform sequencing; Iso-Seq). Samples originating from several tissue types and sampling conditions yielded 2,867,487 reads and a total of 1,462,827 full-length non-chimeric reads (FLNC). Further processing included long read clustering (isONclust), contig reconstruction and detection of unique isoforms (Cogent), and finally their functional annotation (OrthoFinder).

In this way we obtained the most complete representation of *Talinum triangulare* transcriptome, enabling both repeated analysis of existing data sets as well as new comparative studies. Coding sequences of species employing various carbon fixation strategies will be compared with the aim to shed more light on genetic basis for a functional CAM pathway, including but not limited to “CAM switches” transcriptionally responsive during CAM induction and CAM-specific gene isoforms or their regulatory domains.

## **W496: Functional Genomics of C<sub>4</sub> and CAM photosynthesis**

### **Genomics, Physiology, and Breeding Implications of Leaf Stable Carbon Isotope Composition in Maize**

**Anthony J. Studer**, University of Illinois at Urbana-Champaign, Urbana, IL

Soil water deficit is an increasing threat to crop production across the globe. In the US, drought is the leading cause of yield loss in maize and has recently become a target for commercial breeding. Selecting for increased abiotic stress tolerance has been one successful approach to enhance plant performance under water limiting conditions. However, a more proactive approach is to select for increased water-use efficiency (WUE) by reducing the amount of water required per unit of grain. However, selecting directly for water use is difficult. Leaf stable carbon isotope composition,  $\delta^{13}\text{C}$ , is a proxy for transpiration efficiency and a possible tool for breeders to select plants with increased WUE. To better understand the genetic architecture and variability of  $\delta^{13}\text{C}$  in maize, we sampled both diversity panels and elite breeding materials. These populations have allowed us to begin mapping genes influencing  $\delta^{13}\text{C}$ . We have identified several significant quantitative trait loci, which we are fine-mapping to identify the causative polymorphism. A subset of lines is also being used to dissect the physiology underlying the observed variation. Our data indicate that improvements in transpiration efficiency would be complimentary to modifications of morphological traits to improve whole-plant WUE. Experiments aimed at better understand  $\delta^{13}\text{C}$  are being translated into germplasm improvement for hybrid production in combination with other WUE traits. Our next step is to measure whole-plant water-use efficiency and test a model that combines  $\delta^{13}\text{C}$  with morphological traits and genomic data to assess the predictive power for large scale improvement of WUE.

## **W497: Functional Genomics of C<sub>4</sub> and CAM photosynthesis**

## **Functional Genomics of CAM in Kalanchoe**

**James Hartwell**, University of Liverpool, Liverpool, United Kingdom

## **W498: Functional Genomics of C<sub>4</sub> and CAM photosynthesis**

### **Genome Sequencing of *Portulaca amilis*, a C<sub>4</sub>+CAM Plant**

**Ian S. Gilman**, Yale University, New Haven, CT and **Erika J. Edwards**, Yale University

C<sub>4</sub> and CAM represent extremes on a spectrum of photosynthetic phenotypes, with many recognized intermediates that can not or do not exclusively fix CO<sub>2</sub> through a carbon concentrating mechanism (CCM). Numerous studies have shown that there are many physiological, morphological, and genetic traits shared between C<sub>4</sub> and CAM taxa, yet, only two lineages have been found to harbor plants capable of both C<sub>4</sub> and CAM photosynthesis: *Portulaca* (Portulacaceae) and *Ottelia* (Hydrocharitaceae). It has been proposed that the categorical similarities between C<sub>4</sub> and CAM render them mutually exclusive in the same plant. In order to understand the genomic mechanisms underlying this unique photosynthetic pathway, we sequenced, assembled, and annotated the genome of *P. amilis*, a C<sub>4</sub> South American herb with a facultative CAM cycle.

## **W499: Functional Genomics of C<sub>4</sub> and CAM photosynthesis**

### **CRISPR/Cas9-Mediated Genome Editing for Photosynthesis Research**

**Degao Liu**<sup>1</sup>, **Colby Starker**<sup>1</sup>, **Gerald A. Tuskan**<sup>2,3</sup>, **Xiaohan Yang**<sup>2,3</sup> and **Daniel Voytas**<sup>1</sup>, (1)Department of Genetics, Cell Biology and Development, Center for Precision Plant Genomics, and Center for Genome Engineering, University of Minnesota, Saint Paul, MN, (2)DOE-Center for Bioenergy Innovation (CBI), Oak Ridge National Laboratory, (3)Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Human population growth, global climate change and reductions in natural resources could present great challenges for achieving sustainable food and energy supplies. One approach to meet these challenges is to increase the efficiency and productivity of photosynthesis in crop plants. Genome editing using CRISPR/Cas9 has the potential to advance our understanding of photosynthesis in plants to allow informed and guided improvements via genetic engineering. Here we explored the potential of CRISPR/Cas9 systems to install the enablers of C<sub>4</sub> photosynthesis into C<sub>3</sub> plant species to reduce the CO<sub>2</sub> compensation point and induce a primordial C<sub>4</sub>-like carbon cycle through two strategies: 1) heterologous expression of the C<sub>4</sub> evolution enablers from C<sub>3</sub>-C<sub>4</sub> intermediates and simultaneously knock-out of endogenous genes in C<sub>3</sub> plant species; 2) perturbation of the spatiotemporal expression pattern of endogenous genes in C<sub>3</sub> plant species. Also, we developed and implemented a toolbox for CRISPR/Cas9 application in the model crassulacean acid metabolism (CAM) photosynthesis species *Kalanchoë fedtschenkoi* and tested the capability of the CRISPR/Cas9 system to characterize the function of CAM-related genes and to link gene function with CAM physiology. Our results demonstrated that the CRISPR/Cas9 system could open a new door to functional genomics research of plant photosynthesis and facilitate the redesign of photosynthesis to increase plant productivity for food and bioenergy production in a changing world.

## **W500: Fungal Genomics**

### **Next-Generation Screening for Pathogen-Effector, Host-Target Interactions**

**Roger P Wise**<sup>1</sup>, **Valeria Velásquez-Zapata**<sup>2</sup>, **Sagnik Banerjee**<sup>2</sup>, **Priyanka Surana**<sup>3</sup>, **Greg Fuerst**<sup>1</sup> and **J. Mitch Elmore**<sup>1</sup>, (1)Corn Insects and Crop Genetics Research, USDA-Agricultural Research Service, Ames, IA, (2)Iowa State University, Ames, IA, (3)The University of Sydney - Plant Breeding Institute, Sydney, NSW, Australia

Pathogen effectors are excellent tools to explore the dynamic regulation of plant resistance and susceptibility. To discover novel mechanisms of effector action, we exploited the biotrophic powdery

mildew fungus, *Blumeria graminis* f. sp. *hordei* (*Bgh*), and its host, barley (*Hordeum vulgare* L.). We used next-generation sequencing to identify interacting partners from high-throughput yeast two-hybrid (Y2H) assays, using *Bgh* effectors as baits, and as preys, a time-course cDNA library from infected barley and isogenic immune mutants. We evaluated selected vs. non-selected conditions for positive interactors using a robust informatics and statistics pipeline, including mapping reads to barley and *Bgh* genomes, reconstruction of prey fragments and fusions with GAL4-AD, and processing of count data. This information was used to develop a ranking system for the preys, comprising **1)** significant enrichment under selection for positive interactions, **2)** in-frame with GAL4-AD, and **3)** degree of enrichment in pairwise comparisons of baits under selection. Outputs from this pipeline facilitated sorting and validation by binary Y2H.

To position the newly discovered effector targets within cereal signaling pathways, we generated a predicted barley protein-protein interaction network where we can model signaling effects of the barley-*Bgh* interaction. The predicted barley interactome was constructed from evidence-based interlog determination between barley and several plant species, including *Arabidopsis*, rice and maize. Additionally, we included data from the most complete available eukaryotic interactome (*Saccharomyces cerevisiae*), which we expect rescues highly conserved interactions in eukaryotic cells. To obtain interlogs, we used the Ensembl database and InParanoid8 software and collected experimentally validated interactions from BioGRID version 3.5.171, the Protein-Protein Interaction database for Maize (PPIM) and the Predicted Rice Interactome Network (PRIN) databases, as well as literature review. Results from these analysis revealed immune modules enriched for genes associated with transcription, phosphorylation and intracellular transport.

Research supported in part by Fulbright-Colciencias & Schlumberger Faculty for the Future fellowships to VVZ, USDA-NIFA-ELI postdoctoral fellowship 2017-67012-26086 to JME, and NSF-PGRP 1339348 to RPW.

## **W501: Fungal Genomics**

### **Enabling High-Throughput Pan-Genome Analysis on Mycocosm**

**Richard D Hayes**<sup>1</sup>, Robin A. Ohm<sup>2</sup> and Igor V. Grigoriev<sup>1</sup>, (1)US Department of Energy Joint Genome Institute, Berkeley, CA, (2)Utrecht University, Netherlands

Following the production of a high-quality reference annotation at the initial stages of a genome project for a newly sequenced species, it is common to expand sequencing and analysis to additional samples of the same species or related strains to build a "pan-genome", a representation of the collective gene content and structural diversity of that sample population. Initially developed for analysis of prokaryotic populations, this approach increasingly has been transferred to eukaryotic systems, with published results for model land plant and small fungal datasets. We present the results of a survey of several pan-genome analysis methods applied to a pilot study of eight strains of *Schizophyllum commune*, all assembled and annotated at the DOE Joint Genome Institute (JGI). The eventual goal is the development of a pan-genome pipeline that is fully integrated with our existing annotation and functional analysis tools, easily updated as new genomes become available, providing increased value for our Community Science Program scientific partners.

## **W502: Fungal Genomics**

### **Evidence of Local Adaptation and Regional Effector Repertoires in *Parastagonospora nodorum***

**Jonathan K. Richards**<sup>1</sup>, Eva H. Stukenbrock<sup>2</sup>, Zhaohui Liu<sup>3</sup>, Christina Cowger<sup>4</sup>, Justin D. Faris<sup>5</sup> and Timothy L. Friesen<sup>5</sup>, (1)Department of Plant Pathology and Crop Physiology, Louisiana State University, Baton Rouge, LA, (2)Christian-Albrechts Universität & Max Planck Institute for Evolutionary Biology, Kiel, Germany, (3)North Dakota State University, Fargo, ND, (4)USDA-ARS, (5)USDA-ARS, Fargo, ND

Septoria nodorum blotch (SNB) is caused by the necrotrophic fungal pathogen *Parastagonospora nodorum* and continues to threaten wheat production throughout the world. Within the continental United States, *P. nodorum* has adapted to diverse environmental conditions due to the wide geographical range of wheat acreage planted. *P. nodorum* utilizes specialized effector proteins to elicit a programmed cell death response to facilitate disease development. Currently, little is known about differences in effector diversity or genomic regions of positive selection between discrete *P. nodorum* populations. Whole-genome sequences of 197 *P. nodorum* isolates collected from spring, durum, and winter wheat production regions of the United States were developed to test the hypothesis that regional selection pressures, including environmental and host factors, have driven the selection of beneficial genes including potential effectors. A total of 1,026,859 polymorphisms were identified within the natural population. Population structure analyses classified two major populations corresponding to the Upper Midwest and the Southeastern United States. Selective sweep analyses identified 10 and 19 non-overlapping regions of positive selection specific to the Upper Midwest and Southeastern populations, respectively, and co-localized with 92 genes. The number and size of sweep regions in the Southeastern population was greater than the Upper Midwest population, likely indicating more recent selection events. A selective sweep was detected at the *SnToxA* locus specific to the Upper Midwestern population. Further analysis of individual isolates indicated that 95.7% of isolates from that population harbored a functional *SnToxA* compared to only 6.2% of isolates in the Southeastern population. This correlates with the maintenance or elimination of host sensitivity gene *Tsn1* from local germplasm. Additionally, as evidenced by pN/pS ratios, different suites of effectors were diversifying in a population specific manner. Overall, effector genes were under greater diversifying selection when compared to genes encoding non-secreted proteins or secreted non-effectors. Genes confined to a small accessory chromosome were also observed to be more rapidly evolving. Combined, this study sheds light onto local adaptation and genomic mechanisms of diversification in an economically important pathogen of wheat

### **W503: Fungal Genomics**

#### **Linking Genome Organization to Function and Evolution - the Impact of Transposons on a Plant Pathogenic Fungus**

**Michael F. Seidl**<sup>1</sup>, David E. Cook<sup>2</sup>, Jasper RL Depotter<sup>3</sup>, Martin Kramer<sup>4</sup>, David Torres Sanchez<sup>4</sup> and Bart Thomma<sup>5</sup>, (1)Utrecht University, Utrecht, Netherlands, (2)Kansas State University, Manhattan, KS, (3)University of Cologne, Cologne, Germany, (4)Wageningen University & Research, Wageningen, Netherlands, (5)Wageningen University, Wageningen, Netherlands

Eukaryotic genomes are highly organized, and the linear arrangement of genes and transposable elements (TEs) as well as the three-dimensional organization of the chromatin within the nucleus are thought to contribute to genome function and evolution. Here we will discuss our ongoing efforts to study the impact of TEs on *Verticillium dahliae*, an asexual soil-borne fungus that causes vascular wilt disease on hundreds of plant species. We have recently revealed that the genome of *V. dahliae* contains regions with extensive presence/absence polymorphisms, so-called lineage-specific (LS) regions, that are enriched for *in planta*-induced effector genes contributing to host infection. These LS regions evolve by segmental duplications and by genomic rearrangements mediated by erroneous double-strand break repair pathways, often at TEs. In contrast to the core genome, LS regions are also enriched in 'active' TEs that contribute to genome variability. By interrogating a collection of *V. dahliae* strains, we identified a subset of TE families with frequent presence/absence polymorphisms. Interestingly, TEs impact the expression of genes in their proximity, and in particular in LS regions TE and effector gene expression is correlated. We furthermore observed that LS regions display markedly higher sequence conservation in coding and noncoding regions than the core genome. We consider that differences in chromatin organization explain the specific phenomena linked to LS regions. By combining a range of modern sequencing approaches, we demonstrated that LS regions are associated with histone 3 lysine 27 methylated histones (H3K27me3) and have increased DNA accessibility than similar TE-rich regions in the core genome that are typically inaccessible and associated with DNA methylation and histone 3 lysine 9 methylated histones (H3K9me3). Importantly, we determined the three-dimensional organization of the chromatin within the nucleus, and revealed that LS regions co-localize, which provides evidence

linking chromatin organization to genomic variability. Collectively, our results highlight the significance of TEs and the emerging links between genome organization, function, and evolution in an important fungal plant pathogen.

#### **W504: Fungal Genomics**

##### **Chromosomal Rearrangements Drive the Evolution of Secondary Metabolic Gene Clusters within the Entomopathogenic Fungus *Tolypocladium inflatum***

Rodrigo Olarte, University of Minnesota, Jon Menke, Rebiotix, Ying Zhang, Minnesota Supercomputing Institute, Shawn Sullivan, Phase Genomics, Seattle, WA, Jason C. Slot, Ohio State University, Yinyin Huang, University of Minnesota College of Pharmacy, Jonathan P. Badalamenti, University of Minnesota Genomics Center, Alisha C. Quandt, University of Colorado Boulder, Joseph Spatafora, Oregon State University and **Kathryn Bushley**, University of Minnesota, St. Paul, MN

Entomopathogenic fungi are producers of diverse secondary metabolites ranging from lifesaving antibiotics and biopesticides to potent toxins involved in plant and animal pathogenesis, yet few studies have systematically investigated the role of genome architecture and chromosomal evolution in generating diversity of secondary metabolite biosynthetic gene clusters (SMBGCs) within a single fungal species. The entomopathogenic fungus *Tolypocladium inflatum*, while best known as the source of the multi-billion dollar lifesaving immunosuppressant drug cyclosporin, harbors at least 45 additional SMBGCs that may play important roles in pathogenesis of insect pests or colonization of plant hosts. In order to reveal how variation in chromosome structure contributes to the diversification of metabolite clusters in fungi, we used PacBio and Hi-C chromosome conformation capture technologies to assemble chromosomal scale assemblies of the strain from which cyclosporin was originally isolated (NRRL8044) and five additional isolates of *T. inflatum*. We demonstrate that movement of core biosynthetic enzymes between clusters on different chromosomes is an important mechanism driving diversification of metabolite clusters. We also uncover a polymorphic metabolite clusters present in only two strains that contains a homolog of the polyketide synthase and several accessory genes involved in production of the mycotoxin aflatoxin, a carcinogen of major concern to public health. In contrast, the metabolite cluster responsible for production of cyclosporin, an important virulence factor for infection of insects, was conserved across all strains. Our findings clearly demonstrate that within species sampling will be required to adequately capture the pangenome of secondary metabolites produced by a fungal species.

#### **W505: Fungal Genomics**

##### **Evolutionary Pan-Genomics of the Charcoal Rot Fungus, *Macrophomina phaseolina***

**Emile Gluck-Thaler**<sup>1</sup>, Horacio Lopez-Nicora<sup>2</sup>, Timothy Ralston<sup>3</sup>, Veena Devi Ganeshan<sup>3</sup>, Cristhian Grabowski Ocampos<sup>4</sup>, Anne E. Dorrance<sup>5</sup>, Terry L. Niblack<sup>3</sup> and Jason C. Slot<sup>3</sup>, (1)Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, (2)Universidad San Carlos, Paraguay, (3)Ohio State University, (4)Universidad Nacional de Asunción, Paraguay, (5)The Ohio State University/OARDC, Wooster, OH

*Macrophomina phaseolina* (Botryosphaerales) is a globally distributed necrotrophic fungus capable of infecting over 500 plant species, including soybean. Yet despite its broad host range, *M. phaseolina* reproduces exclusively through asexual means. How this fungus overcomes constraints imposed by asexuality in order to generate the genomic variation required for adaptation remains an open question, with important consequences for predicting and managing disease. Here, we present results of an on-going project whose objective is to characterize the processes that generate variation in genomic content among *M. phaseolina* isolates (i.e., the pan-genomic variation of this species), and the genetic features that facilitate these diversity-generating processes. We first sampled *M. phaseolina* from soybean fields in North and South America and captured 200 isolates from Ohio and 300 isolates from Paraguay. We phenotyped virulence on soybean and growth at various temperatures, and sequenced the genomes of 12 isolates with variable phenotypes using a combination of short and long read technologies. We developed a combination of best-practices techniques, which will be discussed in detail, in order to generate a high-confidence pan-genome. We then quantified rates of gain and loss across all genes in

the pan-genome to address the questions: (1) Is gene gain or loss more important for generating differences in genome content within asexual fungal populations? (2) What are the predominant mechanisms of gene gain and loss? and (3) Are certain genomic regions associated with increased rates of gene gain and loss?

## **W506: Galaxy: An Open Platform for Data Analysis and Integration**

### **Introduction to Galaxy and the Galaxy Ecosystem**

**Dave Clements**, Johns Hopkins University, Eugene, OR, Mo Heydarian, Johns Hopkins University, Baltimore, MD and Galaxy Community

Galaxy is a free, open source, and web-based data integration and analysis platform for life science research ([galaxyproject.org](http://galaxyproject.org)). This talk will introduce the Galaxy platform and ecosystem and discuss how it can help you and your collaborators integrate and analyze your research data without requiring you to learn Linux systems administration and command line interfaces. Galaxy is deployed at hundreds of organizations around the world and is supported by a global community of researchers, trainers, software engineers, and tool developers. Thousands of tools and data sources are available in the Galaxy ecosystem. You will gain insight into how Galaxy can be applied to your own research to make it more accessible, reproducible, and transparent. If you are looking for a better way to manage and perform your analyses, or if you are new to high-throughput sequencing data analysis, then this workshop will be of interest to you.

## **W507: Galaxy: An Open Platform for Data Analysis and Integration**

### **Galaxy for Excellence in Breeding**

Alexis Dereeper<sup>1</sup>, Angel Villahoz-Baleta<sup>2</sup>, Venice Margarette Juanillas<sup>3</sup>, Sivasubramani Selvanayagam<sup>4</sup>, **Star Yanxin Gao**<sup>5</sup>, Umesh Rosyara<sup>6</sup>, Victor Jun M. Ulat<sup>7</sup>, Paulino Perez-Rodriguez<sup>8</sup>, Yaw A. Nti-Addae<sup>9</sup>, Elizabeth Jones<sup>5</sup>, Ramil Mauleon<sup>10</sup>, Kelly Robbins<sup>5</sup>, Michael Quinn<sup>11</sup> and Kenneth McNally<sup>10</sup>, (1)IRD, UMR IPME, Montpellier cedex 5, France, (2)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (3)International Rice Research Institute, Metro Manila, Philippines, (4)ICRISAT, Hyderabad, India, (5)Cornell University, Ithaca, NY, (6)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (7)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Edo. de Mexico, Mexico, (8)Colegio de Postgraduado, Texcoco, Mexico, (9)Genomic Open-Source Breeding Informatics Initiative, Ithaca, NY, (10)International Rice Research Institute, Los Baños, Philippines, (11)Excellence in Breeding, CIMMYT, Texcoco, Mexico

Galaxy is a free and open source web-based data integration and analysis platform for life science research ([galaxyproject.org](http://galaxyproject.org)). It is used by thousands of scientists across the world. Excellence in Breeding (EiB) Platform <https://excellenceinbreeding.org/> aims to modernize breeding programs in the developing world by providing access to cutting-edge tools, services, best practices, training, and advice. As part of EiB, the mission of Genomic Open-source Breeding informatics initiative (GOBii) project is to enable routine implementation of marker-assisted and genomic selection by building genomic data management and marker application tools. Working with Galaxy developers and statisticians, we have put existing R-packages and tools in transparent, reproducible, and routine workflows in Galaxy to manage routine genomic selection and genome-wide association study (GWAS) analysis. In this presentation, GOBii genomic selection-Galaxy project, deployment of the EiB Galaxy server <http://galaxy-demo.excellenceinbreeding.org/>, and capabilities will be introduced; Sustainable support and adoption through community of practice will be discussed.

## **W508: Galaxy: An Open Platform for Data Analysis and Integration**

### **Galaxy Demo: Genomic Selection and Genome Wide Association Study (GWAS) Analysis with Excellence in Breeding Tools**

Star Yanxin Gao<sup>1</sup>, **Dave Clements**<sup>2</sup>, Mo Heydari<sup>3</sup>, Umesh Rosyara<sup>4</sup>, Mathias Lorieux<sup>5</sup>, Alexis Dereeper<sup>6</sup>, Angel Villahoz-Baleta<sup>7</sup>, Venice Margarette Juanillas<sup>8</sup>, Victor Jun M. Ulat<sup>9</sup>, Paulino Perez-Rodriguez<sup>10</sup>, Sivasubramani Selvanayagam<sup>11</sup>, Jose Crossa<sup>12</sup>, Ramil Mauleon<sup>13,14</sup>, Elizabeth Jones<sup>1</sup>, Yaw A. Nti-Addae<sup>15</sup>, Xuecai Zhang<sup>16</sup>, Clay Sneller<sup>17</sup>, Dmytro Chebotarov<sup>14</sup>, Juan D. Arbelaez<sup>14</sup>, Fernando Toledo<sup>4</sup>, Juan Burgueño<sup>18</sup> and Kelly Robbins<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)Johns Hopkins University, Eugene, OR, (3)Johns Hopkins University, Baltimore, MD, (4)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (5)IRD-CIAT, Cali, Colombia, (6)IRD, UMR IPME, Montpellier cedex 5, France, (7)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (8)International Rice Research Institute, Metro Manila, Philippines, (9)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Edo. de Mexico, Mexico, (10)Colegio de Postgraduado, Texcoco, Mexico, (11)ICRISAT, Hyderabad, India, (12)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, EM, Mexico, (13)Southern Cross Plant Science, Southern Cross University, Australia, (14)International Rice Research Institute, Los Baños, Philippines, (15)Genomic Open-Source Breeding Informatics Initiative, Ithaca, NY, (16)CIMMYT, Texcoco, Mexico, (17)OHIO STATE UNIVERSITY, WOOSTER, OH, (18)International Maize and Wheat Improvement Center (CIMMYT), El Batán, TEXCOCO, EM, Mexico

Excellence in Breeding (EiB) Platform <https://excellenceinbreeding.org/> aims to modernize breeding programs in the developing world by providing access to cutting-edge tools, services, best practices, training, and advice. As part of EiB, the mission of Genomic Open-source Breeding informatics initiative (GOBii) (<http://cbsugobii05.biohpc.cornell.edu/wordpress/>) project is to enable routine implementation of marker-assisted and genomic selection (GS) by building genomic data management and marker application tools (<http://asia.gobii.org:8081/gobii-portal/>). Collaborating with Galaxy developers and statisticians, we have put a suite of bioinformatic analysis tools, R-packages, and visualization tools in Galaxy (<http://galaxy-demo.excellenceinbreeding.org/>) to manage routine GS and GWAS analysis as reproducible crop-specific workflows.

In this workshop, attendees will learn GS and GWAS analysis tools implemented in Galaxy, including file conversion, imputation, principle component analysis (PCA), best linear unbiased estimate or prediction (BLUE or BLUP), and genomic estimated breeding value (GEBV) calculators. GS and GWAS analysis using maize, wheat, or rice datasets will be demonstrated too. Targeted audience of this hands-on session include any breeders or genomic scientists who are interested in running GS or GWAS analysis yourselves but can't do programming.

### **W509: Gene Expression Analysis**

#### **Insights into Plant Transcriptomes from the 1KP Project**

**Michael K. Deyholos**, University of British Columbia, Kelowna, BC, Canada

The One Thousand Plant (1KP) Transcriptomes Initiative, with participants from over 130 different institutions, recently published RNA-Seq assemblies of 1,147 species from a wide range of plant taxa representing diverse lineages of land plants, as well as green algae, red algae, and glaucophytes. These data have been used to make inferences about whole genome duplications and gene family expansion/contraction, in the context of diversification and innovations in plant evolution (*Nature* 574, 679–685). Here we briefly review these findings and other 1KP outcomes, as well as new analyses into the diversity of transcripts, with a particular focus on small transcripts that are predicted to encode conserved, novel proteins.

### **W510: Gene Expression Analysis**

#### **Machine Learning Based Comparative Analysis of Gene Regulatory Networks in Monocots**

**Donat Wulf** and Andrea Braeutigam, Bielefeld University, Bielefeld, Germany

Gene regulatory networks (GRN) control plant development in time and space and responses to environmental cues. The vast amount of publicly available RNAseq data can be harnessed with machine

learning algorithms such as random forest prediction to assign transcription factors to their target genes. We hypothesized that evolutionary comparisons in five monocot species will identify evolutionarily conserved transcription factors for core traits and changes in the GRN structure for derived traits.

The analysis is focused on photosynthesis as a core trait and C4 photosynthesis as a derived trait. The network of the species *Zea mays*, *Sorghum bicolor*, *Triticum aestivum*, *Hordeum vulgare* and *Oryza sativa* ssp. *Japonica* were inferred. We overcome the machine learning based high error rate in the prediction of targets by combining it with enrichment analyses. These enrichments determine transcription factors which are involved in the regulation of photosynthesis and C4 photosynthesis. By drawing comparisons between the five monocot species we are able to identify conserved regulatory patterns in these species. Comparing the C4 species *Zea mays* and *Sorghum bicolor* with the C3 species *Hordeum vulgare*, *Triticum aestivum* and *Oryza sativa* ssp. *Japonica* reveals that the C4 genes share the regulators with the conserved photosynthesis regulators in all species. Motifs were predicted for the putative C4 regulators based on network data.

### **W511: Gene Expression Analysis**

#### **Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis**

**Shivraj Braich**<sup>1</sup>, Rebecca C. Baillie<sup>1</sup>, Larry S. Jewell<sup>1</sup>, German C. Spangenberg<sup>2</sup> and Noel O.I. Cogan<sup>1</sup>, (1)Agriculture Victoria Research, Bundoora, VIC, Australia, (2)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, Australia

Cannabinoids are the main medicinal compounds of interest in the plant *Cannabis sativa*, that are primarily synthesised in the glandular trichomes; found on female floral buds. The content, composition and yield of secondary metabolites (cannabinoids and terpenoids) is influenced by the plant's genetics and environment. Some initial gene expression experiments have been performed from strains of this plant species that contrasted in cannabinoid production, however the present knowledge about detailed trichome transcriptomics in this species is limited. An extensive transcriptome atlas was generated by RNA sequencing using root, shoot, flower and trichome tissues from a female plant strain (Cannbio-2) and was enhanced with the addition of vegetative and reproductive tissues from a male cannabis plant. Differential gene expression analysis identified genes preferentially expressed in different tissues. Detailed trichomics was performed from extractions specifically from glandular trichomes as well as female floral tissues at varying developmental stages, to identify stage-specific differentially expressed genes. Candidate genes involved in terpene and cannabinoid synthesis were identified and the majority were found to have an abundant expression in trichomes. The comprehensive transcriptome is a significant resource in cannabis for further research of functional genomics to improve the yield of specialised metabolites with high pharmacological value.

### **W512: Gene Expression Analysis**

#### **Lowering the Cost of Photorespiration using Synthetic Biology**

**Amanda P Cavanagh**, University of Illinois at Urbana-Champaign, Urbana, IL, Paul South, Louisiana State University and Don Ort, University of Illinois, Urbana-Champaign, Urbana, IL

Meeting food demands for the growing global human population requires improving crop productivity, and large gains are possible through enhancing photosynthetic efficiency. In C3 plants grown under ambient [CO<sub>2</sub>] at 25°C, 15-20% of the fixed carbon dioxide is lost to photorespiration following Rubisco oxygenation reactions. Furthermore, rates of photorespiration increase with rising temperature, as higher temperatures favour increased Rubisco oxygenation, such that losses can exceed 25% above 30-35°C. Multiple attempts have been undertaken to overcome this yield penalty and increase biomass production in plants, with limited success to date. Introducing alternative metabolic pathways to native photorespiration pathway hold promise as a strategy to mitigate the impact of photorespiratory losses, but until recently these designs have been grown in controlled conditions, and have not been tested under stressful environments. We designed and tested a synthetic pathway that metabolizes 2-

PGlycolate inside plant chloroplasts. Field grown transgenic tobacco plants expressing this pathway show strongly enhanced biomass production, demonstrating that our manipulations could be used to improve crop yields. We then grew tobacco plants expressing this pathway under ambient and elevated temperatures (+5°C) in agricultural field conditions to determine if our pathways offered a thermal protection. We found transgenic plants sustained up to 20% less yield losses under heated conditions compared to unmodified plants, demonstrating that bypassing photorespiration can mitigate high temperature induced losses in plant productivity.

### **W513: Gene Expression Analysis**

#### **Integrative Analysis of Plant Single-Cell Transcriptomes**

**Benjamin J. Cole**, DOE-Joint Genome Institute, Berkeley, CA

Single-cell transcriptome profiling of heterogeneous tissues can provide high-resolution windows into developmental dynamics and environmental responses, but its application to plants has been limited. Recently, a number of studies have demonstrated the use of single-cell RNA-seq (scRNA-seq) on multiple platforms in profiling numerous root cell types and developmental progressions. We have begun to build an integrative analysis of five publicly available single-cell transcriptome data sets from *Arabidopsis*. This assembled resource consists of > 40,000 cells, distributed across roughly 30 clusters with characteristic gene expression profiles. Our analysis highlights the inter-lab reproducibility of these assays, as nearly all cell clusters identified were composed of cells from all data sets. We also identified, in combination with a rebuilt Index of Cell Identity specification matrix, several clusters of lateral root cap cells, and well-defined clusters of xylem and phloem-associated cells. This integrative resource has potential to streamline future applications of scRNA-seq technology on plant roots.

### **W514: Gene Expression Analysis**

#### **Analysis Considerations for Handling Single Cell Transcriptomic Data**

**Jason Kim**, 10x genomics, Pleasanton, CA

### **W515: Gene Introgression**

#### **MAGIC and Pre-Breeding Advances**

**Kathy Dibley**<sup>1</sup>, C. Lynne McIntyre<sup>2</sup>, Bill Bovill<sup>2</sup> and Klara Verbyla<sup>3</sup>, (1)CSIRO Agriculture and Food, Brisbane, QLD, Australia, (2)CSIRO Agriculture and Food, Australia, (3)CSIRO, Data61, Canberra, Australia

Multi-parent advanced generation intercross (MAGIC) populations provide opportunities to dissect complex genetic relationships between traits and the environment that both increase our understanding of the underlying genetic basis of traits and our identification of elite germplasm. At CSIRO we have developed several wheat MAGIC populations involving 4, 8 or 16 diverse parents. In conjunction with phenomic tools, these populations have provided significant insight into the genetic complexity of traits, increased the frequency and robustness of QTL detection, increased the heritability of selection methods for major traits, provided a resource for the dissection of traits across the wheat genome, and assisted in the identification of candidate genes for traits of interest. These advances are now being used to develop elite germplasm more efficiently and effectively.

### **W516: Gene Introgression**

#### **Detection Activities of LMO Introgressions on Maize Landraces at the National Genetic Resources Center in Mexico**

**Moises Cortes Cruz**, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Guadalajara, JA, Mexico

## **W517: Gene Introgression**

### **New Alien Translocation Lines from *Thinopyrum intermedium***

Xianrui Guo, **Handong Su** and Fangpu Han, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Wheat stripe rust is a destructive disease in many cool and temperate regions around the world. TAI-14, a pair of telocentric chromosomes from *Th. intermedium*, was identified with resistance to CYR32, CYR33 and CYR34. Conversely, T14 without alien chromosome was susceptible to all three races and vast numbers of fungal spores spread on wheat seedling leaves. It could be inferred that at least one genetic locus on the alien chromosome were responsible for stripe rust resistance. In order to transfer the alien resistant gene to common wheat (*Triticum aestivum*), translocation lines were created by radiating the pollen of TAI-14 with alien chromosomes. We totally obtained 153 wheat-*Th. intermedium* translocation lines, including 66 short alien segmental translocation lines, 70 long alien segmental translocation lines and 17 intercalary translocation lines. Three specific molecular markers were screened from SSRs and ESTs of wheat group 7, and eleven markers were developed from de novo transcripts. These markers combining with seven different sized translocation lines were used to construct the physical map of the alien chromosome. Most of these markers were clustered at the end of the chromosome. One long alien segmental translocation line WT78 and one intercalary translocation line WT15 were selected after two-generation backcross with wheat cultivar Jimai 22. Both of them show high resistance to stripe rust at seedling stage and adult plant stage, and the resistance was co-segregated with the translocated chromosome. Finally, the gene *YrT14* was located at the alien part with specific marker T14K50 by analyzing the patterns of translocated chromosome from ten susceptible translocation lines. Therefore, the marker can be used for molecular-assisted breeding, and the new lines could potentially be used as new resistant source for wheat breeding program.

## **W518: Gene Introgression**

### **Tracking Alien Introgression in Cultivated Potato: An Example using Cyst Nematode Resistance from the Wild Potato Species *Solanum spegazzinii***

Ulrike Gartner<sup>1</sup>, Miles Armstrong<sup>2</sup>, Sanjeev Kumar Sharma<sup>1</sup>, Gail Wilde<sup>1</sup>, Ingo Hein<sup>1</sup>, Vivian Blok<sup>1</sup> and **Glenn Bryan**<sup>1</sup>, (1)The James Hutton Institute, Dundee, United Kingdom, (2)University of Dundee, United Kingdom

Potato cyst nematodes (PCN) pose an ever-increasing threat to the global potato industry. One way to address this problem is to find and deploy new sources of naturally occurring resistance in wild potato germplasm by introgressing them into potato breeding material. The diploid potato wild species *Solanum spegazzinii* Bitter shows resistance to both cyst nematode species, *Globodera pallida*, Pa1 and Pa2/3 and *Globodera rostochiensis*, Ro1. To investigate the resistance of *S. spegazzinii* to *G. pallida*, a backcross of the wild species to the susceptible diploid group Phureja cultivar Mayan Gold has been studied. After phenotyping a subset of 200 progeny clones of the backcross population for resistance to *G. pallida* population Lindley Pa2/3, resistant and susceptible bulks comprising 20 clones each were identified. Normalized DNA pools from these bulks as well as the two parental clones were subjected to targeted gene enrichment (RenSeq and GenSeq) and non-targeted genotyping-by-sequencing (GBS) next-generation sequencing approaches. The analyses resulted in the identification of single-nucleotide polymorphisms (SNPs) linked to the resistance locus on chromosome VI. SNP markers were converted to KASP<sup>TM</sup> (Kompetitive Allele Specific PCR) assays and further used to screen for recombinants in a much larger backcross population. Using the GBS data, several more SNPs were identified and tested on the remaining recombinants to narrow down the resistance gene location further. Moreover, introgression of the mapped resistance gene directly into relevant tetraploid cultivars is ongoing.

## **W519: Gene Introgression**

### **Increasing Durum Wheat Genetic Diversity by Wild Relative and D-Genome Introgressions**

**Manel Othmeni**<sup>1</sup>, Surbhi Grewal<sup>1</sup>, Stella Edwards<sup>1</sup>, Cai-yun Yang<sup>1</sup>, Duncan Scholefield<sup>1</sup>, Stephen Ashling<sup>1</sup>, Amor Yahyaoui<sup>2</sup>, Perry Gustafson<sup>3</sup>, Pawan Singh<sup>2</sup>, Ian P. King<sup>1</sup> and Julie King<sup>1</sup>, (1)University of Nottingham, United Kingdom, (2)CIMMYT, Mexico, (3)University of Missouri, Columbia, MO

The wild relatives of wheat provide an important source of genetic variation for wheat improvement. Much of the work in the past aimed at transferring genetic variation from wild relatives into wheat has relied on the exploitation of the *ph1b* mutant, located on the long arm of chromosome 5B. This mutation allows homologous recombination to occur between chromosomes from related but different genomes, e.g. between the chromosomes of wheat and related chromosomes from a wild relative resulting in the generation of interspecific recombinant chromosomes. However, the *ph1b* mutant also enables recombination to occur between the homologous genomes of wheat, e.g. A/B, A/D, B/D, resulting in the generation of wheat intergenomic recombinant chromosomes. In this work we report on the presence of wheat intergenomic recombinants in the genomic background of hexaploid wheat/*Amblyopyrum muticum* introgression lines. The transfer of genomic rearrangements involving the D-genome through pentaploid crosses provides a strategy by which the D-genome of wheat can be introgressed into durum wheat. Hence, a pentaploid crossing strategy was used to transfer D-genome segments, introgressed with either the A- and/or the B-genome, into the tetraploid background of two durum wheat genotypes Karim and Om Rabi 5 in either the presence or absence of different *Am. muticum* ( $2n = 2x = 14$ , TT) introgressions. Introgressions were monitored in backcross generations to the durum wheat parents *via* multi-colour genomic *in situ* hybridization (GISH). Tetraploid lines carrying homozygous D-genome introgressions, as well as simultaneous homozygous D- and T-genome introgressions, were developed. Introgression lines were characterized *via* Kompetitive Allele-Specific PCR (KASP) markers and multi-color fluorescence *in situ* hybridization (FISH). Results showed that new wheat sub-genomic translocations were generated at each generation in progeny that carried any *Am. muticum* chromosome introgression irrespective of the linkage group that the segment was derived from. The highest frequencies of homologous recombination were observed between the A- and the D-genomes. Results indicated that the genotype Karim had a higher tolerance to genomic rearrangements and T-genome introgressions compared to Om Rabi 5. This indicates the importance of the selection of the parental genotype when attempting to transfer/develop introgressions into durum wheat from pentaploid crosses.

## **W520: Gene Introgression**

### **Hybrid Wheat - Building on Diversity**

**David Bonnett**, BASF Corporation, Sabin, MN

Hybrid Wheat is the main focus of BASF's global wheat breeding investments; in North America it is the sole focus.

Commercially successful hybrid wheat has previously been elusive, but sometimes tantalizingly close. BASF's current hybrid wheat investments build on the successes and aim to overcome the deficiencies of past efforts.

Diversity is both a foundation and a challenge for hybrid wheat. Genetic diversity provides the basis for heterosis at commercially attractive levels in most published studies across a wide spread of environments and germplasm over several decades. This foundation needs building to achieve higher combining ability within germplasm that must also be competitive for grain quality, biotic and abiotic stress responses.

In addition to genetic diversity, cytoplasmic diversity provides a means of generating and restoring male sterility to facilitate hybrid production and assure that resulting hybrids are fully fertile. Achieving full sterility and complete fertility restoration is challenging but achievable.

Diversity for both male and female components of cross-pollinating ability is abundant but can be difficult to select. Cross-pollinating ability is one of the key drivers of the price of hybrid seed and the generally low cross-pollinating ability of wheat has been one of the key challenges to commercial viability.

Certainly, there is complexity in building and combining all the components to support commercially successful hybrid wheat, but good progress is being made towards launch of the first commercial hybrids in the mid-2020s, supported by a range of tools and technologies. Advanced genomics, predictive breeding and phenomics, together with speed breeding and state of the art field phenotyping are rapidly bringing together all the key components. Continued development of these technologies and integration of novel diversity from native sources and an array of alternative technologies will underpin future advances in hybrid wheat performance and commercial success.

### **W521: Gene Mapping by Segregation**

#### **Gene Mapping in Era of High-Throughput Phenotyping**

**Zhiwu Zhang**, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

With advances of sensing, especially remote sensing, phenotyping is available at large scale. This not only enlarge selection pressure for breeding, but also enhance the opportunity for gene mapping. This presentation presents the recent progress of High-Throughput Phenotyping, the applications and prospective of gene mapping in the new era of phenotyping.

### **W522: Gene Mapping by Segregation**

#### **RNA Polymerase Mapping in Plants Identifies Enhancer Candidates Enriched in Causal Variants**

Roberto Lozano<sup>1</sup>, Greg Booth<sup>2</sup>, Bilan Yonis-Omar<sup>3</sup>, Edward Buckler<sup>4</sup>, John Lis<sup>2</sup>, **Jean-Luc Jannink**<sup>5</sup> and Dunia Pino del Carpio<sup>2</sup>, (1)Cornell University, Ithaca, NY, (2)Cornell University, (3)SupAgro, France, (4)USDA-ARS-Cornell University, Ithaca, NY, (5)USDA-ARS, Ithaca, NY

Control of gene expression is fundamental at every level of cell function. Promoter-proximal pausing and divergent transcription at promoters and enhancers, which are prominent features in animals, have been reported to be absent in plants. PRO-Seq analysis in cassava (*Manihot esculenta*) identified peaks of transcriptionally-engaged RNA polymerase II (Pol2) at both the 5' and 3' end of genes, consistent with paused or slowly-moving Pol2. Additionally, we identified divergent transcription at potential intergenic enhancers. A full genome search for bi-directional transcription using an algorithm for enhancer detection developed in mammals (dREG) identified many enhancer candidates. These sites show distinct patterns of methylation and nucleotide conservation based on genomic evolutionary rate profiling (GERP). SNPs within these enhancer candidates predict significantly more variation in fitness and root composition than SNPs in chromosomal segments randomly ascertained from the same intergenic distribution, suggesting a functional importance of these sites. Maize GRO-Seq data showed RNA polymerase occupancy at enhancers consistent with patterns in cassava. Furthermore, putative enhancers in maize identified by dREG significantly overlapped with sites previously identified on the basis of open chromatin, histone marks, and methylation, and were enriched for reported eQTL. Our results suggest that bidirectional transcription can identify enhancer regions in plants that play an important role in transcription regulation and that their identification has the potential to aid crop improvement.

### **W523: Gene Mapping by Segregation**

#### **Genomic Approach for Enhancing Abiotic Stress Resilience in Alfalfa (*Medicago sativa* L.)**

**Long-Xi Yu**<sup>1</sup>, Cesar Medina<sup>2</sup>, Sen Lin<sup>2</sup>, Michael Peel<sup>3</sup>, Meijing Liang<sup>4</sup> and Zhiwu Zhang<sup>5</sup>, (1)USDA-ARS Plant Germplasm Introduction and Testing Research, PROSSER, WA, (2)USDA-ARS, Prosser, WA, (3)USDA, Logan, UT, (4)Dept. of Crop and Soil Science, Washington State University, Pullman, WA, (5)Washington State University, Pullman, WA

Alfalfa is a worldwide forage legume and is “queen of forage” due to its high nutritional value. However, the production of alfalfa is challenged by adverse environmental factors such as drought and high salinity. Developing resistance alfalfa is an important breeding target for enhancing alfalfa productivity in arid and semi-arid regions. Since drought and salinity occur simultaneously in these regions. To understand the genetic base of abiotic stress resilience, two alfalfa populations were used for evaluating

drought and salt resistance in the field. Genome-wide association studies with genotyping by sequencing were used for mapping resistance loci. Twenty-eight markers were associated with yield under drought. Most of the markers were identified across different harvest periods under water deficit, although different levels of significance were found among the three harvests. The loci associated with biomass yield under water deficit located throughout all chromosomes in the alfalfa genome agreed with previous reports. Our results suggest that biomass yield under drought may involve a different mechanism compared to that of non-stress. BLAST searches of the flanking sequences of the associated loci against DNA databases revealed several stress-responsive genes linked to the drought resistance loci, including leucine-rich repeat receptor-like kinase, B3 DNA-binding domain protein, translation initiation factor IF2 and phospholipase-like protein. Marker-trait association identified a total of 42 markers significantly associated with five traits associated with salt tolerance, including fresh and dry weights, plant height, relative water content, stomatal conductance. They were located on all chromosomes except chromosome 2 based on the alignment of their flanking sequences to the reference genome (*Medicago truncatula*). Of those identified, 13 were associated with multiple traits. Several loci identified in the present study were also identified in previous reports. BLAST search revealed that 19 putative candidate genes were collocated with 24 significant markers. Among them, B3 DNA-binding protein, Thiaminepyrophosphokinase and IQ calmodulin-binding motif protein were identified among multiple traits in the present and previous studies. With further investigation, the markers closely linked to drought and salt resistance can be used for MAS to accelerate the development of new alfalfa cultivars with improved resistance to drought and high salinity.

#### **W524: Gene Mapping by Segregation**

##### **Genetic Architecture and Gene Mapping of Cyanogenic Glucoside in Cassava**

**Alex C Ogonna**<sup>1,2</sup>, Guillaume J. Bauchet<sup>2</sup>, Eder Oliveira<sup>3</sup> and Lukas A. Mueller<sup>1,2</sup>, (1)Cornell University, Ithaca, NY, (2)Boyce Thompson Institute, Ithaca, NY, (3)Embrapa, Brasília, Brazil

Cassava is a root crop originating from Latin America and a major staple crop in the Tropics, including Africa. In this study, we focused on South American cassava germplasm and investigated Hydrogen Cyanide (HCN), a major component of tuber quality. HCN is involved in cassava plant defense against herbivores but a toxic compound upon tissue disruption. We genotyped 3,354 landraces and modern breeding lines originating from 26 different Brazilian states and phenotypically, 1,389 individuals were characterized across multi-year-location trials. All plant material was subjected to high density genotyping using Genotyping-By-Sequencing (GBS) and 27,000 Single Nucleotide Polymorphic (SNP) markers were selected. We performed the first association mapping (GWAS) to characterize the genetic architecture and gene mapping of HCN. Field experiment revealed phenotypic heritability of 0.74 (with SNP based heritability, 0.41) for HCN. Association studies revealed two major loci contributing 7 and 30% of the marker based estimated variance and indicated the presence of genes encoding for an ATPase and MATE protein respectively. We developed and validated sets of diagnostic markers for breeding applications and investigated evidence for domestication in HCN. Our findings were further validated in an African population and provides future resources for genetic studies of cyanide in cassava.

#### **W525: Gene Mapping by Segregation**

##### **High Throughput Image Techniques in Breeding**

**Yang Hu**<sup>1</sup>, Chunpeng (James) Chen<sup>1</sup>, Zhou Tang<sup>1</sup>, Long-Xi Yu<sup>2</sup> and Zhiwu Zhang<sup>1</sup>, (1)Dept. of Crop and Soil Science, Washington State University, Pullman, WA, (2)USDA-ARS Plant Germplasm Introduction and Testing Research, PROSSER, WA

As the most cultivated legume forage worldwide, alfalfa has been improved successfully for resistance to diseases and winter hardiness, however, the genetic improvements for economically important traits, such as forage and seed yields, have been limited due to phenotyping challenges. We developed a pair of software packages, GRID and GridFree, to overcome the challenges in phenotyping biomass and seed characteristics, respectively. GRID was designed to process images for fields with plots in grids, which is typical for alfalfa experiments. Based on the field images collected by an Unmanned Aerial

Vehicle, the vegetarian area of alfalfa plots extracted by GRID had a much higher correlation than the area extracted manually. GridFree was designed to process images without grid settings, including images of alfalfa seeds. Counting alfalfa seeds and measuring length and width are critical for alfalfa breeding. Large seeds not only lead to better germination but also have the potential for high seed yield. High throughput phenotyping methods have been developed by analyzing images of spreaded kernels on a surface. The challenges remain in the segmentation of kernels from background variation or separating kernels that are next to each other tightly. GridFree was developed to overcome these challenges. GridFree uses an unsupervised machine learning approach K-Means to separate kernels from the background by using color index-based approaches and uses Gaussian normal distribution as a dynamic criterion for a divide-and-combine strategy to segment adjacent kernels. Both GRID and GridFree were implemented in Python with a friendly graphic user interface and are freely available at <http://zzlab.net>.

## **W526: Gene Mapping by Segregation**

### **Mapping Gene Presence-Absence in a Recent Complex Polyploid Crop Genome**

**Iulian Gabur**, Department of Plant Breeding, Justus Liebig University, Giessen, Germany

Gene presence absence variation (gene PAV) is known to occur in many gene families of important crop species. The extent and influence of gene PAV on quantitatively inherited agronomic traits is largely unknown. Brassica napus is a recent allotetraploid crop genome that has undergone many events of gene loss and speciation and has a strongly rearranged genome. We investigated the association of gene PAV with resistance of oilseed rape (canola) to a fungal pathogen *Verticillium longisporum*, as an example for a complex, quantitatively inherited resistance. Genomic variation was assessed in QTL intervals based on the Brassica Illumina 60k SNP array of bi-parental and multi-parental mapping populations and compared with genome-wide resequencing data of parental lines. In addition, an exon capture array was specially designed to investigate the extent of gene PAV for genes within confidence intervals of quantitative trait loci (QTL). Our results provide insights into the prominent role of gene PAV in quantitatively inherited disease resistance. These findings will improve future breeding efforts on Brassica napus and other closely related species.

## **W527: Genome annotation resources at the EBI**

### **Accessing Plant and Animal Genomic Data with Ensembl**

**Astrid Gall**, EMBL-EBI, Hinxton, United Kingdom

Ensembl ([www.ensembl.org](http://www.ensembl.org)) provides a comprehensive and integrated source of mainly vertebrate genomic data, while its sister project Ensembl Genomes ([www.ensemblgenomes.org](http://www.ensemblgenomes.org)) contains data for plants, fungi, protists, bacteria and (invertebrate) metazoa.

Together, these resources cover more than 230 vertebrates, including livestock such as cow, pig, sheep, goat, turkey and chicken, more than 60 plants, including crops such as bread wheat, barley, maize, soy, sorghum and rice, and several important pathogens, pests and vectors.

Gene annotations are available for all species. Comparative genomics data, namely alignments, gene trees and homologue predictions, are available for all plant and animal species at the taxa level and for a subset of species between different taxa. Variation data includes short and structural variants, phenotypes, population genetics data as well as pathogenicity predictions and adds value for several species.

All data are freely available. They can be accessed via the browser websites, BioMart tool, FTP server, MySQL, REST and Perl APIs. A variety of tools allow to analyse both data in Ensembl and your own data, e.g. BLAST and the Variant Effect Predictor (VEP) which are available for all species.

This session will give an introduction to the data in Ensembl and Ensembl Genomes, highlight key displays in the browser websites and demonstrate the use of two of our tools: BioMart to export custom data sets and the VEP to analyse your own variation data. It will focus on agriculturally important plants and animal species.

## **W528: Genome annotation resources at the EBI**

### **Manual Genome Annotation in Ensembl**

**Jane Loveland**, EMBL-EBI, Cambridge, United Kingdom

The Human and Vertebrate Analysis and Annotation (HAVANA) team at the European Bioinformatics Institute (EBI) are part of the Ensembl team and undertake manual annotation of vertebrate genomic sequence. As part of the GENCODE project, we are responsible for producing detailed reference annotation of all human and mouse protein-coding genes, pseudogenes, long non-coding RNAs and small RNAs. Previously we have also annotated whole genomes and chromosomes for zebrafish, pig and rat and specific regions of interest, such as the Major Histocompatibility Complex (MHC), for selected organisms. We are working to refine the mature genesets for the human and mouse genomes, for example adding more than 100 new protein-coding genes, ~500 pseudogenes and ~1800 lncRNA loci to the human GENCODE gene set.

Updating GENCODE genesets with the deluge of new long transcriptomic sequencing methods such as PacBio and ONT is a significant challenge for manual gene and transcript annotation. Our TAGENE workflow incorporates new long read transcriptomics datasets into GENCODE with a manually supervised automated method to create and modify transcripts for non-coding genes and for adding additional coding transcripts to existing protein-coding genes. As such, the numbers of both genes and transcripts is expected to increase significantly in future.

For more than a decade, GENCODE gene annotation was created via the merging of two independent databases; one containing manual gene annotation and the other automated annotation. We have recently changed the workflow such that all annotation is stored in a single Ensembl database. This has several advantages, including the removal of complex and time consuming merge process, the assignment of permanent stable IDs to genes and transcripts at their creation and perhaps most importantly, the facilitation of manual editing of genes and transcripts produced by the Ensembl genebuild pipeline. As a result of these changes we have created an infrastructure to allow us to add manual annotation to, and edit computational annotation for any Ensembl species, for example farmed animal species.

Annotation is a continuous process and so between the regular releases of the Ensembl/GENCODE update cycle we release new annotation every 24 hours via an update track hub that can be accessed from all genome browsers and is available here:

[http://ftp.ebi.ac.uk/pub/databases/genocode/update\\_trackhub/hub.txt](http://ftp.ebi.ac.uk/pub/databases/genocode/update_trackhub/hub.txt)

The annotation can be downloaded from [gencodegenes.org](http://gencodegenes.org) and is the default annotation available from the Ensembl and UCSC genome browsers.

## **W529: Genome annotation resources at the EBI**

### **The European Variation Archive: Genetic Variation Archiving and Accessioning for All Species**

**Baron Koylass**, EMBL-EBI, Hinxton, United Kingdom and **Jose Miguel Mut-Lopez**, (EMBL) - The European Bioinformatics Institute, Cambridgeshire, United Kingdom

## **W530: Genome annotation resources at the EBI**

## **Submission, Archival and Visualisation of Single-Cell Sequencing Data**

**Nancy George**, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom, Irene Papatheodorou, European Bioinformatics Institute, Hinxton, United Kingdom and Gene Expression Group

Single Cell Expression Atlas (SCEA) ([www.ebi.ac.uk/gxa/sc](http://www.ebi.ac.uk/gxa/sc)) is the latest component of the Expression Atlas. The Expression Atlas knowledgebases form part of a comprehensive suite of functional genomics resources developed at the EBI from submission to visualisation. SCEA ingests and systematically reanalyses and visualises publicly available single cell RNA-sequencing (scRNA-seq) datasets from functional genomics archives such as ArrayExpress; NCBI's GEO and ENA. As of its latest release (December 2019), SCEA contains 132 datasets, across 12 species and over 1.3 million cells.

For inclusion into the Single Cell Expression Atlas all datasets are curated to a high standard complying with minimum metadata requirements (described in detail here: <https://arxiv.org/abs/1910.14623>). These standards are incorporated into the single cell submission template in our web submission tool Annotare. Upon acceptance, datasets are given a stable, citable accession, reviewed by curators and uploaded to ArrayExpress – a functional genomics data archive. ArrayExpress ([www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)) then displays each dataset as a self-contained entity providing the experiment information; sample metadata and links to raw and processed data under a single accession.

From other resources, such as NCBI's GEO or EGA, when a suitable dataset is discovered, raw data is retrieved from the corresponding archive or data submitter. Sample metadata is curated manually and where possible annotated to ontology terms for easier data search and retrieval. All raw data is processed through standardised analysis pipelines depending on the scRNA-seq technology used. All analysis workflows are available through our GitHub repositories: <https://github.com/ebi-gene-expression-group/> whilst tools to run these are made available via Galaxy here: <https://humancellatlas.usegalaxy.eu>.

Once reanalysed, the scRNA-seq analysis results are made available to the wider scientific community through the visualisations in the SCEA interface. Through this platform users can search for genes across datasets and filter the results for particular cell types or tissues. SCEA can also be used to identify in what conditions and populations a gene can act as a marker gene, i.e. define a specific cell population. For each experiment cell populations are displayed via a t-SNE plot. Cells can also be coloured with the underlying metadata. Gene expression at the single cell level can be explored in the neighbouring plot. The top 5 marker genes per cluster are also displayed and all analysis data and accompanying metadata are available for download.

## **W531: Genome annotation resources at the EBI**

### **Expert Manual Curation of Plant Protein Sequences and Plant Metabolic Pathways in UniprotKB/Swiss-Prot**

**Damien Lieberherr**, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland and The UniProt Consortium

Our group specializes in the development and maintenance of expert curated knowledge resources for the life sciences. These include the UniProt knowledgebase (UniProtKB, <https://www.uniprot.org>) a comprehensive, high-quality and freely accessible resource of protein sequences and functional information that covers thousands of plant species, and the Rhea knowledgebase (<https://www.rhea-db.org>), a resource of computationally tractable biochemical reactions.

Here we describe recent work designed to improve the utility of these knowledge resources for integrated computational and experimental analyses of plant metabolic systems, with a particular focus on natural product biochemistry.

Plants produce various natural products with extremely diverse molecular structures and activities. These natural products may have interesting medicinal properties (as antibiotics, anti-cancer treatments, analgesics or immune-suppressive drugs), as well as applications in the agronomy (as insecticides, fungicides and more), food (as flavors or pigments for example) and energy sectors (as biofuels). Expert curation of natural product pathways in UniProt and Rhea provides a means to link these chemical structures to the underlying genomic sequence of the gene clusters that synthesize them, and presents new opportunities for genomic data mining for the discovery of new biosynthetic routes for industry. We will present examples drawn from a wide range of natural product pathways and plants including the anti-malaria drug artemisinin in *Artemisia annua*, and morphine and derivatives analgesics in *Papaver somniferum* in UniProt using Rhea.

## **W532: Genome annotation resources at the EBI**

### **Open-Access Tools for Genome Annotation**

**Sandra Orchard**, EMBL-EBI, Hinxton, United Kingdom

Adding functional information to entries describing genes and proteins for which there is no experimental work being published is an ever-growing problem. UniProt is a long-standing collection of databases that enable scientists to navigate the vast amount of sequence and functional information available for proteins. The reviewed section of the UniProt knowledgebase (UniProtKB/Swiss-Prot) contains over 560 000 sequences that have been created by our expert biocuration team. For these entries experimental information has been extracted from the literature and organized and summarized, greatly easing scientists' access to protein information. The unreviewed section of the database (UniProtKB/TrEMBL) currently provides a further 190 million sequences that have been largely derived from high throughput sequencing of DNA. These entries are annotated with high degree of accuracy by our automatic annotation systems such as UniRule, consisting of annotation rules which are created by the biocurators as part of the process of curation of the experimental literature for UniProtKB/Swiss-Prot and the Statistical Automatic Annotation System (SAAS), a completely automatic decision-tree-based rule-generating algorithm. We will describe how we are sharing these annotation systems with the research community and we have developed a freely-available rule engine, UniFire, which enables external researchers to annotate the protein sequences of newly sequenced genomes.

Additionally, we are now faced with an influx of largely redundant proteomes as increasing numbers of strains/breeds of plants and animals are sequenced. We look for input from the community as to how we can best present this data to you, whilst still making it possible for programs such as BLAST to run efficiently.

## **W533: Genome annotation resources at the EBI**

### **The Global Biodata Coalition: Sustaining Life Science Data Resources**

**Charles E Cook**, Global Biodata Coalition, Hinxton, United Kingdom

Research in the life sciences is increasingly data-driven, and researchers across the world are dependent on the data integration and analysis enabled by open-access biodata resources. These resources form a major global infrastructure that is crucial to current and future biomedical and life sciences research, but the infrastructure has developed piecemeal over recent decades without coordination among funders or data resource managers, and there is no comprehensive understanding of the infrastructure at a global level. Many key data resources rely on short-term funding from a small set of funders and are at risk of losing their funding.

The Global Biodata Coalition (GBC) has been formed with the support of public and charitable research funders to address the challenges in maintaining the biodata infrastructure. The GBC is a forum for funders of data resources to coordinate and share approaches for the efficient management and growth of the infrastructure. The GBC will work with funders to stabilize and ensure support for the global biodata infrastructure and identify a set of Global Core Data Resources that are most essential for

biomedical and life sciences research. The goal of this process is to ensure long-term support for these key components of the global infrastructure that are crucial for the health of the entire infrastructure.

### **W534: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment**

#### **eGreenhouse In-a-Box and Other Gadgets from the Openly Published Environmental Sensing Lab**

**Chet Udell**, Oregon State University, Corvallis, OR

We live in an unprecedented era where the variety of food available to us can be sourced any time of the year regardless of season or climate conditions, largely thanks to greenhouses. Additionally, indoor farms have been able to equal the production capacity of 2 acres of traditional farmland in just 320 square feet. However, there is an ever increasing disparity in our country's aging greenhouse infrastructure versus the capabilities afforded by new technologies to make greenhouse management more automated and efficient. A common solution employed by mega-growers is to renovate or construct new infrastructure with the latest technologies built in. However, 84% of Oregon's farms are sole proprietorships with 64% of Oregon's agricultural land managed by operators age 55 and older. Building new, state-of-the-art facilities is not an option for many of these mid-to-small scale family businesses. What if there were a kit that, like installing a drip irrigation system, one could easily distribute a number of internet-controlled devices to interact with sprinklers, grow lights, heat lamps, and fans with no programming or electronics experience? And what if those devices could be automated by a transporting sensors like soil moisture, CO<sub>2</sub>, NDVI, temperature, and RH throughout the greenhouse? The speaker will present a new open-source out-of-the-box eGreenhouse kit, and other technologies to meet these challenges.

### **W535: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment**

#### **Multifunctional Sensing of the Plant Root Zone: Water Status, Gas Exchange, Nutrient Level and Beyond**

**Scott Jones**, Utah State University, Logan, UT

The rapidly advancing realm of bioinformatics is growing the value of research in plant biotechnology as developing bioinformatic tools integrate the resulting large data sets generated by expanding '-omic' technologies. However, improved management and monitoring of the plant's environmental conditions and stresses are needed in order to fully understand and improve candidate plants, such as global food crops, under what are often marginally managed or poorly understood environmental conditions of targeted stress (e.g., drought, flooding, salinity, etc.). In response to the looming global food crisis and the urgent need to sustainably secure and more efficiently manage resources such as water and nutrients to produce food, considering competing resource demands, we are developing a transformative sensor technology. Thermo-TDR has tremendous potential to improve our awareness of soil and porous media water, air, and nutrient content in addition to advancing understanding of processes therein. The proposed novel multifunctional Thermo-TDR measures and outputs information on the soil water content, nutrient status (via electrical conductivity), temperature, and thermal properties, which are all currently available in separate sensors. The novel aspect of a commercialized Thermo-TDR sensor is providing unprecedented in-field soil bulk density and porosity determination capability, leading to advances in agricultural production, management and sustainability as well as for other disciplines (e.g., engineering, food science, etc.). Soils serve many important functions, whether in crop growth, horticulture and agronomy, golf course and landscape management, etc. Practitioners, researchers, farmers all have common needs of measuring or determining soil water content, temperature and salinity but would also benefit from determination of bulk density, air-filled- and total-porosity. Porous medium bulk density and porosity are presently determined in a laboratory from, for example, field-based soil samples, a process that takes several days and therefore limits timely water and nutrient management decision making. Furthermore, determination of soil air-filled porosity and pore water salt (solute) concentration are not generally determined, but would be valuable real-time assessment tools for agricultural production activities.

### **W536: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment**

#### **Integrating High-Resolution Sensors in Convergent Systems to Manage Risks to Crop Productivity**

**D. Keith Roper**, Utah State University, Logan, UT

Climate and ecological variability, extreme weather events, and abiotic and biotic stresses reduce crop yields and limit farmer profitability. Irrigation, fertilization and pest/pathogen chemicals are important to remedy crop stress but their inefficient use can deplete soil, energy and water resources *and* deteriorate water *via* nutrient runoff and air *via* greenhouse gas emissions, thus threatening sustainable production. Stress-resilient cultivars and decision support systems for sustainable intensification in precision agriculture to manage irrigation, fertilizer, pests and pathogens could mitigate such factors. However, the lack of workable, high-resolution sensors has resulted in gaps like sparsity and heterogeneity of data to phenotype and model plants, and little-known interactions between genetic, environmental, management and socioeconomic (GEMS) systems, and barriers like cost, information deficit, social norms and trait correlation. These preclude systematic risk/benefit-informed responses to stress.

This presentation will discuss a convergent systems approach to structure and secure data from higher-resolution sensors, integrate models of plants, soils and hydrology with empirical models of climate and socioeconomic tradeoffs, and derive boundaries and GEMS linkages using data science and artificial intelligence to prescribe **(i)** selection of stress-resilient cultivars; **(ii)** sustainable management of crops with practices that adapt to environmental, genetic and socioeconomic variability; and **(iii)** training a tech/infosavvy agri-ecoengineering workforce. Crowd-sourcing has been used to prioritize gaps, barriers, existing capabilities and opportunities for integrated data collection, analysis, systems modeling, and forecasting in risk-based decision support systems. It includes results from surveys, workshops, roundtables, symposiums and consortia development. Mentored-student contests in ethics and design have attracted top talent and identified bold new innovations based on **(i)** I-Corps Lean LaunchPad principles; **(ii)** proven innovation strategies; and **(iii)** community input. Engagement with underrepresented and economically disadvantaged participants and minority serving institutions has been supported by tiered mentoring, communities of practice and internships where innovation, design, teamwork, communication, ethics and leadership are modeled.

### **W537: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment**

#### **Sensors and Robots—the Future of Plant Phenotyping?**

**Grégoire Hummel**, Phenospex, Heerlen, Netherlands, Stefan Schwartz, Phenospex, Netherlands and Claude Juriens, ecoRobotix SA, Switzerland

There are more and more autonomous robots available on the market, for different application in agriculture. Those robots seem to be ideal tools to carry sensors to phenotype plots and plants in the field. The robots are light, flexible, cost efficient and fully autonomous, hence an ideal tool to carry sensors for the purpose of plant phenotyping. In this talk we present insights and data of an autonomous robot, which was equipped with a multispectral 3D laser scanner to automate routine application of plant breeders in the field.

### **W538: Genome Variation and Somatic Cell Breeding**

#### **Reconstruction of Cell Karyotypes and Chromosome Evolutionary Trajectories of Plants**

**Xiyin Wang**, North China University of Science and Technology, Tangshan, China

### **W539: Genome Variation and Somatic Cell Breeding**

#### **Deciphering the High Quality Genome Sequence of Coriander**

**Xiaoming Song**, North China University of Science and Technology, Tangshan, China and Xiyin Wang, North China University of Science and Technology, Watkinsville, GA, China

Coriander (*Coriandrum sativum* L.  $2n = 2x = 22$ ), a plant from the Apiaceae family, also called cilantro or Chinese parsley, is a globally important crop used as vegetable, spice, fragrance, and traditional medicine. Here, we report a high-quality assembly and analysis of its genome sequence, anchored to 11 chromosomes, with total length of 2,118.68 Mb and N50 scaffold length 160.99 Mb. We found that two whole-genome duplication events, respectively dated to ~45-52 and ~54-61 million years ago, were shared by the Apiaceae family after their split from lettuce. Unbalanced gene loss and expression observed between duplicated copies produced by these two events. Gene retention, expression, metabolomics and comparative genomic analyses of Terpene synthase (TPS) gene family, involved in terpenoid biosynthesis pathway contributing to coriander's special flavor, revealed that tandem duplication contributed to coriander TPS gene family expansion, especially compared to their carrot counterparts. Notably, a TPS gene highly expressed in all 4 tissues and 3 development stages studied, is likely a major-effect gene encoding linalool synthase and myrcene synthase. The present genome sequencing, transcriptome, metabolome and comparative genomic efforts provide valuable insights into the genome evolution and spice trait biology of Apiaceae and others related plants, and facilitated further research into important gene functions and crop improvement.

#### **W540: Genome Variation and Somatic Cell Breeding**

##### **The Reductive Genome of *Zygnema circumcarinatum* Provides New Insights into Plant Terrestrialization**

**Yanbin Yin**, University of Nebraska-Lincoln, Lincoln, NE

The charophyceangreen algae (CGA) species, *Zygnema circumcarinatum*, represents the closest algal relatives to land plants. Sequencing its genome thus will contribute to the understanding of the origin and early evolution of land plants. The genomes of four *Z. circumcarinatum* strains: hypothetically diploid UTEX 1559 [unfunctional mating +], and haploid UTEX 1560 [mating type -], SAG 698-1a [mating type +] and SAG 698-1b [mating type -] have been sequenced by our lab. Comparing the *Zygnema* genomes with genomes of other green algae and land plants can elucidate the genomic basis of adaptation to the terrestrial environment. Using the 3rd generation long read technology, an Oxford Nanopore sequencing of UTEX 1559, SAG 698-1a and 1b were finished. Meanwhile, the 2nd generation Illumina sequencing reads have also been generated for the four *Zygnema* strains. RNA sequencing has also been done for UTEX 1559 and SAG 698-1b with desiccation and cold treatments. A hybrid assembly of the draft genome was finished for SAG 698-1b, which had a genome size of 65Mb (contig N50=154,887bp), close to the estimated genome size of ~64Mb by a DAPI experiment staining the nuclear DNA, a flow cytometry experiment and a k-mer frequency analysis. This is the smallest genome compared to six recently sequenced charophyote genomes. Preliminary genome annotation indicated that this genome contained 10.95% repeat elements, 0.56% retro elements, 0.56% LTR elements and 0.26% DNA transposons, respectively. An ab initio gene prediction using the program MAKER with the help of RNA-seq data predicted 17,460 protein coding genes. Other analyses, such as evolution features in terrestrialization, cell wall evolution, phytohormone regulation, are still on going.

#### **W541: Genome Variation and Somatic Cell Breeding**

##### **The *Arachis hypogaea* Genome Provides Insight into the Polyploid Origin and Subgenomes Differential Evolution**

**Weijian Zhuang**, Fujian Agriculture and Forestry University, Fuzhou, China

*Arachis hypogaea* (peanut or groundnut) is a globally important oil and food legume crop. The most exciting events this year were the completion of genomes sequencing which provides a landmark for peanut research. Employing multiple methods for genomic sequence comparisons and Ks value estimation, the origin of peanut polyploidy was dated about 400,000 years ago clearly contradictory to the previous estimates. Through genomic comparison, differential evolution of A and B subgenome was found after polyploidy based on the Shitouqi genome. B subgenome derived from *A. ipaensis* with more anchored contigs (55.49% of assemblies) and shared more colinear genes. It also showed more gene expansion than A subgenome. More A subgenome genes (58.7% of total 629) were converted by their B

subgenome counterparts. A total of 30,596 non-redundant genes were identified including 24,208 homeologs pairs which demonstrated widespread differential expressions between two subgenomes, and B subgenome contained much more dominant expression homeologs than A subgenome. Congruently, we found A subgenome underwent greater structural rearrangements and large scale LTR transposable elements expanded after tetraploidization, reconstructing the chromosomes. These not only explain the prevalence of dysfunctional expression or loss of A subgenome homeologs, but propose that the sequenced *A. duranensis* was not the A subgenome donor. The results provide insights into the peanut domestication and genetics enhancement.

#### **W542: Genome Variation and Somatic Cell Breeding**

##### **Expression Patterns and Epigenetic Regulation under Subgenomic Interaction in *Brassica napus***

**Chaobo Tong**, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, China

#### **W543: Genome Variation and Somatic Cell Breeding**

##### **Patterns of Adaptive Genetic Variation across *Coffea canephora***

**Valérie Poncet**, IRD UMR DIADE, Montpellier cedex 5, France

Understanding how organisms respond to their environment by altering physiological processes will increase our capacity to make predictions about adaptation to global climate change. Adaptive clines have been increasingly studied in plant species within temperate zones to understand adaptation of organism in natural populations. However, they are still poorly understood in tropical environments. *Coffea canephora*, cultivated as Robusta, is an interesting tropical tree model to investigate adaptation in the tropics, as it is largely distributed within the range of the lowland tropical rain forests of Africa. In particular, modifications occurring in genes related to abiotic stress tolerance make these genes candidate for enhanced resilience to future climate change. We combined the use of both captured regions sequenced for a set of candidate genes related to drought tolerance and whole genome SNP markers. Leveraging on a robust statistical approach combining multiple neutrality statistics, we provided a comprehensive map of selection signals in the genome of the *C. canephora* both at the species level and within its major genetic groups.

The genotype-environment association suggests regional adaptation to spatially varying environments of the recent past, with a special focus on the Eastern edge of the distribution, in Uganda. More specifically, we found signals of selection tightly linked to several genes involved in response to biotic and abiotic stress and in caffeine biosynthesis. Our detection of selection signals support the hypothesis of present ecological gradient contributing to the structure of the genetic diversity. Moreover, assessing the genomic vulnerability of the present populations will help to predict their response to future environmental changes.

Denoeud, F. *et al.* (2014) The coffee genome provides insight into convergent evolution of caffeine biosynthesis. *Science* **345**, 1181-1184.

Dereeper, A. *et al.* (2015) The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Research* **43**, D1028-D1035.

Marraccini, P. *et al.* (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *Journal of Experimental Botany* **63**, 4191-4212.

Merot-L'anthoene, V. *et al.* (2019) Development and evaluation of a genome-wide Coffee 8.5 K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. *Plant Biotechnology Journal* **17**, 1418-1430.

Gomez, C. *et al.* (2016) Shift in precipitation regime promotes interspecific hybridization of introduced *Coffea* species. *Ecology and Evolution* **6**, 3240-3255.

#### **W544: Genomic Advances in Fruit and Vegetable Breeding**

##### **Multi-Omics Identification of Flavor and Aroma Genes in Strawberry**

**Chris Barbey**<sup>1</sup>, Kevin M. Folta<sup>1</sup>, Vance M. Whitaker<sup>2</sup> and Seonghee Lee<sup>2</sup>, (1)University of Florida, Gainesville, FL, (2)University of Florida, Wimauma, FL

Flavor is an important consumer trait in the commercial strawberry. However, breeding improvement in this area has been slow due to the immense genetic and chemical complexity of flavor in octoploid strawberry. This work describes multi-omics strategies to identify causal biosynthesis genes at the octoploid subgenome level. Fruit volatile metabolomes were derived from over 800 individuals using statistical alignment techniques on non-targeted GC/MS data. A subset of 212 individuals were selected for array-based genotyping across eight pedigree-connected biparental populations in multiple seasons. Fifty-five fruit transcriptomes were assembled based on the subgenome-scale octoploid genome to identify candidate genes via trait/transcript correlation and expression-QTL co-segregation. Novel fruit volatile QTL were discovered for methyl anthranilate, methyl 2-hexenoate, methyl 2-methylbutyrate, ethyl butanoate, ethyl hexanoate, mesifurane, and various mono- and sesquiterpenes. These terpenes including linalool, 3-carene,  $\beta$ -phellandrene,  $\alpha$ -limonene, linalool oxide, nerolidol,  $\alpha$ -caryophellene,  $\alpha$ -farnesene, and  $\beta$ -farnesene. An abundantly fruit-expressed methyl transferase is located 0.01 Mb (two genes) from the most-correlated QTL marker shared by three separate methyl ester compounds, including the grape-like methyl anthranilate. In a separate QTL specific to methyl anthranilate, an abundantly fruit-expressed anthranilate synthase gene is located 0.16 Mb (24 genes) from the most-correlated marker. For mesifurane, an epistatic interaction was detected between the known causal gene (O-METHYL TRANSFERASE 1) and a novel QTL likely corresponding to a furaneol glucosyltransferase. Strawberry mono- and sesquiterpenes each co-locate to an identical genomic hotspot containing various terpenoid synthesis pathway components, including the known biosynthesis gene NEROLIDOL SYNTHASE 1 (*FanNES1*). Differences in linalool and other monoterpene levels are partially explained by co-segregation with a *FanNES1* eQTL. Additional evidence show likely quantitative effects from other terpenoid-pathway genes in this narrow hotspot.

#### **W545: Genomic Advances in Fruit and Vegetable Breeding**

##### **Genome Assembly of Homozygous Octoploid Strawberry Inbred Line Provides Insights into Crop Improvement**

**Jinhee Kim**<sup>1</sup>, Do Sun Kim<sup>1</sup>, Sun Yi Lee<sup>2</sup>, Abinaya Manivannan<sup>3</sup>, Jong Pil Hong<sup>4</sup>, Byoung-Cheorl Kang<sup>5</sup>, Jungeol Yum<sup>6</sup>, Eun-Su Lee<sup>4</sup>, Hye-Eun Lee<sup>7</sup> and Ye Rin Lee<sup>8</sup>, (1)National Institute of Horticultural and Herbal Science, Wanju-gun, Jeollabuk-do, South Korea, (2)National Institute of Horticultural & Herbal Science, Wan-ju gun, South Korea, (3)NIHHS, RDA, Ministry of Agriculture, Korea, Jeonju, South Korea, (4)National Institute of Horticultural and Herbal Science, Wanju-gun, South Korea, (5)Seoul National University, Seoul, Korea, Republic of (South), (6)Seoul National University, Seoul, South Korea, (7)National Institute of Horticultural & Herbal Science, Wanju\_Gun, South Korea, (8)National institute of horticultural and herbal science, Wanju\_Gun, South Korea

Allopolyploid plants have a higher level of genome complexity, which leads to additional challenges to improve the quality of reference genome. In the current study, the de novo sequencing, assembly, and annotation of the octoploid inbred line (*Fragaria ×ananassa*) were performed using two highly homozygote inbred lines ( $S_{10}$ ) derived from cv. Benihoppe and Chandler. A total of 804 Mb genome size was sequenced ( $\sim 80\times$  coverage), and

subsequent assembly of contigs and scaffolding process were carried out using Hi-C in combination with the HiRise scaffolder. The gap-filling and error-correction were performed by PBJelly. A linkage map was constructed using the F<sub>2</sub> mapping population derived from the cross of two inbred lines. A total of 81,233 raw SNPs were obtained from the GBS analysis. Further, 138,099 markers were obtained additionally from Axiom array. From the total of 219,332 SNPs (81,233 from GBS analysis and 138,099 from Affymetrix Axiom 90k SNP array), 6,077 polymorphic SNPs were selected based on the Mendelian segregation ratio and used to construct a linkage map, consisting 47 linkage groups. We verified the collinearity between our high-density genetic map and the reference genome assembly data. The ab initio gene prediction approaches was used to annotate predicted genes (151,891 genes). Furthermore, several gene functions associated with fruit quality were determined by computational analysis, indicating the high quality of reference genome. Taken together, our results show that the new octoploid reference genome from inbred lines can be a valuable resource for studying genetics and genomics of octoploid strawberry, and further improving commercial varieties.

#### **W546: Genomic Advances in Fruit and Vegetable Breeding**

##### **Using Synteny and Candidate Genes to Identify Loci Controlling Fruit Sweetness in Blackberry**

Jason D. Zurn<sup>1</sup>, Mandie Driskill<sup>1</sup>, Sook Jung<sup>2</sup>, Dorrie Main<sup>2</sup>, Melinda H. Yin<sup>3</sup>, Melissa C. Clark<sup>1</sup>, Lailiang Cheng<sup>4</sup>, John R. Clark<sup>3</sup>, Margaret L. Worthington<sup>3</sup>, Chad E. Finn<sup>5</sup> and **Nahla Bassil**<sup>1</sup>, (1)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (2)Department of Horticulture, Washington State University, Pullman, WA, (3)Department of Horticulture, University of Arkansas, Fayetteville, AR, (4)School of Integrative Plant Science, Horticulture Section, Ithaca, NY, (5)USDA-ARS Horticultural Crops Research Unit, Corvallis, OR

A synteny-based approach was used to identify candidate genes responsible for sugar production in blackberry (*Rubus* L.). Three sugar quantitative trait loci (QTL) were identified from the GDR QTL database that are conserved among apple (*Malus domestica*), peach (*Prunus persica*), and alpine strawberry (*Fragaria vesca*). The physical regions for these QTLs were identified in the *F. vesca* v1.1 assembly and 26 genes with functions associated with sugar production were extracted. Additionally, 789 sugar-associated genes were extracted from the *M. domestica* v3.0.a1 assembly. The *Fragaria* and *Malus* genes were used to conduct a BLAST search in the GDR *Rubus* reference transcriptome. Of 279 *Rubus* candidate transcripts identified, predicted exons were used to design 9,355 Hyb-Seq baits. The baits had a 2X tiling density and covered 99.6% of the targeted regions. These baits were used in conjunction with PacBio sequencing to genotype 40 cultivars with high and low sugar content from the University of Arkansas and USDA blackberry breeding programs. A total of 430,167 high quality circular consensus sequences (CCS) were generated. Alignment to the 'Hillquist' blackberry and *Rubus occidentalis* genomes, followed by variant identification resulted in 929,430 and 1,324,854 markers, respectively. Welch's t-test and a Benjamini-Hochberg correction identified 467 and 312 significant loci from the 'Hillquist' and the *R. occidentalis* genotype tables, respectively. Population structure modeling identified a total of 173 loci that were significantly ( $\alpha = 0.05$ ) associated with sugar production regardless of population structure. We are in the process of validating these loci using KASP genotyping.

#### **W547: Genomic Advances in Fruit and Vegetable Breeding**

## **The Sweet Corn Genome Assembly and Comparative Analysis with Other Field Corn Genomes**

**Ying Hu**<sup>1</sup>, Vincent Colantonio<sup>1</sup>, Barbara Muller<sup>1</sup>, Kristen Leach<sup>1</sup>, Matheus Baseggio<sup>2</sup>, Kelly Swarts<sup>3</sup>, Michael A. Gore<sup>2</sup>, William F. Tracy<sup>4</sup> and Marcio F. R. Resende<sup>1</sup>, (1)University of Florida, Gainesville, FL, (2)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (3)Gregor Mendel Institute of Molecular Plant Biology GmbH, Vienna, Austria, (4)University of Wisconsin-Madison, Madison, WI

Sweet corn is a specialty crop and one of the most popular vegetables in the United States. Despite advances in genome assembly and annotation in multiple lines of field corn, a reference genome of sweet corn is currently unavailable. In this project, we sequenced a sweet corn inbred line, la453 with the mutated shrunken 2 allele (la453-sh2). This mutation increases sugar content and is present in the vast majority of commercial hybrids developed for the fresh market. We sequenced and de novo assembled a la453-sh2 reference genome through a combination of single-molecule real-time (SMRT) long-read sequencing, BioNano optical mapping and Dovetail Hi-C mapping technologies. The final assembly has a genome length of 2.286 Gb and contains 10 pseudomolecules and 8440 contigs with a N50 of 222.2Mb. BUSCO and long-terminal repeat (LTR) Assembly Index (LAI) analysis indicate that our assembly has higher continuity and completeness than B73 genome. Genes of the la453 genome were annotated using a modified Maker-P pipeline and 38,384 high-confidence protein-coding genes were predicted. Comparative genomics analysis was performed to understand the differences in the structure variation between sweet corn and field corn. Our study provides a high-quality reference genome of sweet corn for further follow-on studies as well as a set of target regions potentially relevant for the traits that are specific to sweet corn and are under selection when compared with field corn.

## **W548: Genomic Advances in Fruit and Vegetable Breeding**

### **Genomic Characterization of Novel Fruit Ripening Pathways**

**Seanna Hewitt**, Washington State University, Pullman, WA, Christopher Hendrickson, National University, San Diego, CA, Evan Stowe, Molecular Plant Sciences, Washington State University and Amit Dhingra, Molecular Plant Sciences, Washington State University, Pullman, WA

Climacteric fruits are characterized by a dramatic increase in autocatalytic ethylene production, which is accompanied by a spike in respiration, at the onset of ripening. The change in the mode of ethylene production from autoinhibitory to auto-stimulatory is known as the system 1 (S1) to system 2 (S2) transition. European pear (*Pyrus communis* L.) cultivars require a genetically pre-determined duration of cold-temperature exposure to induce autocatalytic system 2 ethylene biosynthesis and subsequent fruit ripening. What happens during the cold treatment is becoming clearer at the molecular level. Differential expression, functional annotation, and gene ontology enrichment analyses have provided interesting evidence for the involvement of cold-induced, vernalization-related genes and repressors of endodormancy release, and an unexpected involvement of AOX transcription at pre-climacteric stage. These genes have not previously been described to play a role in fruit during the ripening transition. Besides the need for cold, application of 1-methylcyclopropene in European pear irreversibly obstructs the onset of system 2 ethylene production resulting in perpetually unripe fruit. 1-MCP is an ethylene receptor antagonist which blocks ethylene perception and downstream ripening responses. In pear, application of exogenous ethylene, carbon dioxide and treatment to high temperatures is not able to reverse the blockage in ripening. Activation of AOX via exposure of 1-MCP treated 'D'Anjou' pear fruit to glyoxylic acid has been shown to trigger an accelerated ripening response. Ripening is consistently evident in decrease of fruit firmness and onset of S1-S2 ethylene transition. Transcriptomic and functional enrichment analyses have helped in identifying genes and ontologies implicated in glyoxylic acid mediated ripening, including alternative oxidase, TCA cycle, fatty acid metabolism, amino acid metabolism, organic acid metabolism, and ethylene responsive pathways. These data point to the glyoxylate cycle as a metabolic hub linking multiple pathways to stimulate ripening through an alternate mechanism. The results have provided information regarding how blockage caused by 1-MCP may be circumvented at the metabolic level, thus opening avenues for consistent ripening in pear and possibly other fruit. Understanding metabolic intervention points at which ripening responses can be manipulated provide key, species- and cultivar-specific gene targets which can be altered via gene editing or

transgenic approaches for proactive modulation of ripening to enable development of strategies or new cultivars for reducing overall postharvest wastage.

#### **W549: Genomic Advances in Fruit and Vegetable Breeding**

##### **Genetics Characterization of New and Existing Sources of Tomato Yellow Leaf Curl Virus Resistance in Tomato**

**Upinder Gill**<sup>1</sup>, Rafaela Ruschel<sup>2</sup>, Anuj Sharma<sup>2</sup> and Samuel F. Hutton<sup>3</sup>, (1)North Dakota State University, (2)University of Florida, (3)University of Florida, GCREC, Wimauma, FL

Tomato yellow leaf curl virus (TYLCV) is an important viral pathogen in the Southeastern United States and other tropical and subtropical regions in the world. Genetic resistance against TYLCV has been a major breeding priority at the University of Florida since the 1990s, and several *Ty* resistance genes have been discovered and deployed. Our current research involves identification of new sources of TYLCV resistance in the wild relatives of tomato and characterization of existing *Ty* genes (single or in combination) for TYLCV resistance under field conditions. We recently identified TYLCV resistance in three *Solanum pimpinellifolium* accessions (*viz* LA2093, LA2102, and LA2173) under natural disease pressure, and we confirmed this resistance under controlled TYLCV inoculations using viruliferous whitefly inoculations. Quantitative trait loci (QTL) analysis of a RIL population developed previously from the cross between NC EBR-1 and LA2093 identified a single major QTL (LOD score >13) on chromosome 6, which was responsible for TYLCV resistance in LA2093. The physical location of this QTL encompassed that of the *Ty-1/Ty-3* alleles from *S. chilense*, suggesting either further allelic diversity at the *Ty-1/Ty-3* locus or a new *Ty* gene located within this region. The *Ty-1/Ty-3* locus was significantly associated with resistance in the F<sub>2</sub> populations developed from LA2093 and LA2173 but not from LA2102, indicating that resistance in LA2102 is conferred by a separate locus (loci). An additional locus (loci) besides *Ty-3* may also be present in LA2093 and LA2173, since partial resistance was observed in some *Ty-3* negative plants. Additionally, we have developed near-isogenic lines (NILs) by incorporating single and double *Ty* genes into the backgrounds of several fresh market tomato inbreds. These NILs were controlled inoculated with TYLCV and evaluated for three seasons (Fall 2018, Spring 2019, and Fall 2019) under field conditions for TYLCV disease severity and virus titer. The latest data from our trials will be presented.

#### **W550: Genomic features and chromosome functionality**

##### **The DNA “Jigsaw Puzzle” Structure Model As the Molecular Basis of Biology: Variation and Interaction of Genome Constituting Elements**

**Hongbin Zhang**<sup>1</sup>, Meiping Zhang<sup>1</sup>, Yun-Hua Liu<sup>1</sup>, Xiaoli Qi<sup>2</sup> and Xiaojun Su<sup>3</sup>, (1)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, (2)College of Life Science, Jiamusi University, Jiamusi, China, (3)Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, China

The genomes of all higher organisms sequenced today consist of limited types of fundamental functional elements (FFE), including genes and non-gene elements, such as DNA transposable elements and retro-transposable elements. Therefore, it is essential to study the genome of a higher organism in an integrated and systems manner, known as Systems Genomics, in order to comprehensively understand the variation, evolution and biology of an organism. We previously discovered that the genome (DNA) of a higher organism is structured as “Jigsaw Puzzle” of these FFEs (Wu et al. 2006, *Genomics* 88:394-406; Liu et al. 2014, *Genome* 57:9-19) and found that (1) the number and ratio (Zhang et al. 2010, *Nucleic Acids Research* 38:6513-6525), (2) sequence and order (Zhang et al. 2011, *BMC Genomics* 12:447), (3) interaction, and (4) modification of the FFEs vary dramatically among genotypes within a species and significantly contribute to the genetic variation, evolution and biology of an organism. The number variation of FFEs is now known as the disposable elements of a species' pan-genome. Here we report the variation in the number of FFEs within a bi-parental population, interaction among FFE families, and their potential roles in plant genetic variation, evolution and biology. We found that the number of genes in a gene family also substantially varied within a biparental population and the FFEs

consisting of a genome were correlated in association with the genetic variation, evolution and biology of a species. These findings provide insights into the molecular basis of genome variation, evolution and biology of a species.

**W551: Genomic features and chromosome functionality**

**Preferential Retention of Cold Tolerance Genes Increases Robustness of Cold Resistance Regulatory Network in Paleo-Polyploid Plants**

**Xiaoming Song**, North China University of Science and Technology, Tangshan, China

TBA

**W552: Genomic features and chromosome functionality**

**Interactions between Meiotic Recombination and Sequence Diversity in Arabidopsis**

**Ian R Henderson**, University of Cambridge, Cambridge, United Kingdom

During meiosis DNA double strand breaks undergo interhomolog repair to yield crossovers, which has a diversifying effect on genetic variation. Sequence polymorphism between the homologous chromosomes is known to feedback onto recombination in *cis* and *trans*. To investigate how genetic polymorphism affects crossover distributions genome-wide, we sequenced multiple Arabidopsis recombinant populations. I will present our work detailing both positive and negative effects of sequence polymorphism on meiotic recombination.

**W553: Genomic features and chromosome functionality**

**Karyotype Evolution and Tetraploid Formation of Peanuts**

**Xiyin Wang**, North China University of Science and Technology, Watkinsville, GA; North China University of Science and Technology, Watkinsville, GA, China

Recently, different lines of peanut genomes have been sequenced, which laid a solid foundation for peanut biological and agricultural research. With the peanut genome sequences and those of the other legumes, we managed to reconstruct the karyotype evolution of legumes. By using collinear genes, we dated the formation of tetraploid peanut to have occurred millions of years, excluding the possibility of human work in its origination. However, different approaches reached divergent estimation. Here, we have to alert method selection in genomic data analysis and explanation.

**W554: Genomic features and chromosome functionality**

**Fast and Accurate Genome Assembly and Polish Tools for Third Generation Sequencing (TGS) Reads**

**Jiang Hu**, GrandOmics Biosciences Co.,Ltd., Wuhan, China

The third-generation long-read sequencing, such as the PacBio single-molecule real-time (SMRT) sequencing and Oxford Nanopore sequencing (ONT), can overcome the challenges that are inherent to short-read and can resolve complex and repetitive genomic regions in genomes. However, the high error rate of both sequencing technologies requires software tools that can efficiently correct those errors in the long reads. These tools are essential to produce a high quality sequence assembly by correcting the sequence errors at the initial assembly step as well as at the consensus generation step from assembled contigs, which usually are cost-inhibitive, time-consuming, and computer-resource intensive. Therefore, we developed two software, NextDenovo, a string graph-based *de novo* assembler for long reads, and NextPolish, a fast and efficient genome polishing tool, to generate high quality genome assemblies with great efficiency. Compared with other existing tools with similar functions, both NextDenovo and NextPolish perform better by producing more contiguity assemblies with higher accuracy, and consuming

significant less compute resources. Binary releases can be found at <https://github.com/Nextomics/NextDenovo> and <https://github.com/Nextomics/NextPolish>.

### **W555: Genomics-Assisted Breeding**

#### **Smart Targeting of Germplasm Bank Exploration**

**Sarah Hearne**, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, Jorge Franco, Facultad de Agronomía, Paysandu, Uruguay, Huihui Li, Institute of Crop Sciences, CIMMYT-China office, CAAS, Beijing, China and José Crossa, International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico

Germplasm banks worldwide are reserves of native genetic variation. Despite this the use of germplasm bank materials by breeders is woefully limited. This is not due to a lack of desire for new variation for crop improvement rather significant challenges related to accession use including multifaceted linkage drag, photoperiod incompatibility, poor adaptation and lack of useful documentation of germplasm bank resources. The relative affordability of sequence based genotyping has ushered in a new era in germplasm bank exploration. Initiatives in the three major commodities exemplifies this interest. In maize, CIMMYT together with Mexican and international partners has fully characterized the CIMMYT international germplasm bank maize collections using DArTseq and in addition characterized a sub set of the collection using GbS. Past and ongoing exploration of this data reveals new paradigms and opportunities for variant discovery and allele mining opportunities extending the value of germplasm bank collections not just as resources for novel variation but as test-beds for identification of high value standing variation in elite breeding materials. This understanding is being integrated with efforts on pre-breeding to better inform exotic germplasm selection and refine and redefine breeding approaches.

### **W556: Genomics-Assisted Breeding**

#### **Combining Mutant Analysis and Genome Wide Association for Root Genetics Dissection in Barley**

**Silvio Salvi**, Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy

Knowledge about the genetic control of root development and architecture is lagging behind others plant traits and functions. This is negatively impacting modern plant breeding addressing drought tolerance and nutrient use efficiency. In this study, we investigated the genetic control of barley root architecture by exploring both induced and native genetic variation. We screened the barley TILLMore mutagenized population to identify root architecture mutants at seedling stages (2 weeks) using a semi-hydroponic system. We identified approx. 40 mutant lines, which grouped in three categories: root growth rate/length (short and long, 77%), root morphology (coiling or geotropic, 15%) and root hairs (hairless or shorthairs, 8%). Several mutants were tested for Mendelian inheritance and confirmed. SNP-based bulk-segregant analysis combined with exome and/or whole-genome shotgun sequencing enabled us to identify root candidate genes. Using the same root phenotyping protocol, a collection of >400 barley landraces and cultivars was phenotyped and GWA was carried out taking advantage of exome-seq based SNP analysis. Interestingly, the mutant loci and GWA-based QTL showed little if any overlaps, suggesting the presence of a largely undiscovered genetic system controlling root architecture in barley.

### **W557: Genomics-Assisted Breeding**

#### **Breeding Pedigrees Demonstrate Asymmetrical Subgenome Recombination and Selection in Polyploid *Brassica napus***

**Yuanyuan Zhang**, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China

Pyramiding excellent QTLs/genes plays a pivotal role of breeding strategies in crop domestication and improvement, especially pedigree breeding of founder parents. To exploit key genome regions during breeding, we deploy the pedigree-based analysis on an elite rapeseed, Zhongyou 821, by deep

resequencing of 11 cultivars (50×) in two pedigrees. Tracing the inherited and changed genome blocks with pedigrees and information on SNPs revealed the chromosomal recombination during breeding, which showed great difference between An- and Cn-subgenome that the changed regions ratio of An-subgenome in Zhongyou 821 is significantly higher than that of Cn-subgenome in two pedigrees. Meanwhile, An-subgenome exhibited a higher frequency of the small fragments and lower frequency of the large fragments, whereas Cn-subgenome showed the opposite distribution of that in An-subgenome. Furthermore, the genome recombination rate of the changed blocks in An subgenome is significantly higher than that in Cn subgenome. We also found several genomic regions decreasing genetic diversity which might be caused by a recent human selection in rapeseed breeding. With the verification on the genes regulating seed glucosinolate and erucic acid contents, and the aid of their pedigree information, we clarified the dynamics of asymmetrical genome recombination during rapeseed breeding process. Our study provides genetic dissection of polyploid crop breeding and sheds new light on how to perform genome selective breeding.

### **W558: Genomics-Assisted Breeding**

#### **Predicting a Path for Increased Genetic Gain in Wheat using Artificial Intelligence**

**Kai Voss-Fels**, The University of Queensland, Brisbane, QLD, Australia, Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia, Bertus Jacobs, LongReach Plant Breeders, Australia and Ben J. Hayes, The University of Queensland, Brisbane, Australia

The rate of genetic gain in wheat must be doubled over the next 2-3 decades to secure global food supply. Production trends worldwide suggest that current breeding strategies need to be improved in order to achieve this goal. One of the most critical steps in a wheat breeding program is to select parents for targeted crossing but strategies for decision support are not readily available. Conventionally, parents are selected based on their performance per se or their breeding value which limits the inference that can be made about the probability for maximising the number of favourable chromosome segments in the offspring of a given cross. This is particularly challenging when a combination of quantitative traits (yield) and mono- or oligogenic traits (e.g. disease resistance, quality) are considered simultaneously. This problem is well suited to evolutionary computation approaches, where algorithms inspired by biological evolution, such as reproduction, mutation, recombination, and selection are employed to find solutions to complex problems.

In this ongoing study, we use evolutionary computing in a large commercial wheat data set comprising more than 34,000 genotyped breeding lines which have been tested over multiple years and locations across Australia. The algorithms predict optimal crosses that most efficiently stack complementary alleles for quality, disease resistance and yield. Generated offspring from those crosses is advanced through “speed breeding” which allows a rapid turnaround of several generations and breeding cycles per year. Finally, these approaches will be compared for efficiency with alternative breeding strategies, such as standard genomic selection and phenotypic selection.

### **W559: Genomics-Assisted Breeding**

#### **Taking Cool-Season Grain Legume Breeding to the Next Level: The Key Role of the Pea Genome Sequence**

**Nadim Tayeh**, INRAE, UMR1347 Agroécologie, Dijon, France

The transition from a standalone phenotypic selection to a marker-assisted selection has been seen as a great step forward to improve the breeding process and reach the expected goals. More recently, the genomic revolution has also had its great impact on breeding. -omics are now part of the required toolkit

for a successful, cost and time-efficient breeding. The genome sequence of pea (*Pisum sativum*) has been made available in 2019 through a collaborative international effort. This is a great tool for the pea community in general and the Fabaceae community in particular. Current challenges facing pea and other Fabaceae production are numerous. A large number of traits has also to be tackled in order to adapt the products to the various intended markets while guaranteeing their quality and quantity. This talk will highlight ongoing efforts to take pea breeding to the next step with genomics-assisted breeding. Examples will be given to illustrate how the pea resource is also important for lentil and faba bean programs.

### **W560: Genomics-Assisted Breeding**

#### **Genomic Tools for Early Generation Selection Delivers Improved Peanut Cultivars for Cultivation.**

**Pasupuleti Janila**, ICRISAT, Hyderabad, India

Peanut breeding program at ICRISAT deployed process innovations in breeding pipeline to enhance selection efficiency through early generation selection, and operational efficiency. ICRISAT's global partnership with NARS, advanced research institutes, private sector, and platforms (HTGP) resulted in development of genomic tools, and their deployment in peanut breeding pipeline of ICRISAT and NARS programs. First high oleic peanut varieties were identified for commercialization in India during 2019; Girnar 4 (ICGV 15083) and Girnar 5 (ICGV 15090) have an oleic acid content of 80+2% and recorded superior pod yield and agronomic performance under national testing conducted under All India Coordinated Research Project on Groundnut (AICRP-G). High oleic peanut varieties are preferred by the food industry for extended shelf-life benefits and studies have established consumer health benefits. Selection efficiency is enhanced through the early generation selection using genotyping for FAD mutant alleles, and robust-phenotyping using Near Infrared Reflectance Spectroscopy (NIRS). The 'high oleic' peanut lines in Spanish and Virginia Bunch types suitable for cultivation in Asia and Africa under rainfed conditions are bred at ICRISAT and over 100 lines were shared by ICRISAT with collaborators in nine countries including Australia. The fast-track development in 8-years as against 12-years was possible through reduced generation interval and multi-location testing. High Throughput Genotyping Platform (HTGP) enabled the use of SNP based markers for high oleic trait by ICRISAT and NARS. The use of F2 seed-chip instead of leaf disc of F2 plants further enhanced operational efficiency in deploying SNP genotyping. The 10-SNP panel has markers for two QTLs associated with two important foliar fungal diseases, rust and late leaf spot (LLS) resistance and thus the 10-SNP panel is extensively used in the peanut breeding program at ICRISAT to improve resistance to rust and LLS, and FAD2B mutant allele associated with the high oleic trait.

### **W561: Genomic Selection and Genome-Wide Association Studies**

#### **Overcoming Six Major Challenges to Implement Genomic Prediction**

**Andrew Cromie**, ICBF, Bandon, Co. Cork, Ireland

There are six major challenges that must be addressed in the implementation of a genomic improvement programme. First, there must be a sound data infrastructure for collection, storage and reporting of data and information. Second, a good body of phenotypic and genotypic data is required to generate a suitable training population so that selection candidates can be ranked. Third, the relevant stakeholders must gain confidence in prototype genomic predictions through some form of validation. Fourth, there must be an understanding and communication of the analytical pipeline from molecular measures of the genome through to predictions of breeding values. Fifth, a breeding program with a good value proposition must be identified. Finally, there needs to be a transition from the traditional breeding scheme to the one that exploits genomics. The Irish Cattle Breeding Federation punches well above its weight because it has addressed these six challenges in manners which are the envy of dairy and beef cattle industries throughout the world.

### **W562: Genomic Selection and Genome-Wide Association Studies**

## **Integrating Gene Expression Mapping, Epigenetics and GWAS to Understand the Genetic Control of Agronomic Traits in Hexaploid Wheat**

Fei He<sup>1</sup>, Wei Wang<sup>2</sup>, William B. Rutter<sup>3</sup>, Alina Akhunova<sup>4</sup>, Monica F. de Soto<sup>5</sup>, Les J. Szabo<sup>6</sup>, Katherine Jordan<sup>2</sup>, Matthew N. Rouse<sup>7</sup> and **Eduard Akhunov**<sup>1</sup>, (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Kansas State University, Department of Plant Pathology, Manhattan, KS, (3)USDA ARS, (4)Integrated Genomics Facility, Kansas State University, Manhattan, KS, (5)Integrated Genomics Facility, Kansas State University, (6)USDA-ARS, University of Minnesota, St. Paul, MN, (7)USDA-ARS Cereal Disease Lab, St Paul, MN

In recent years, wheat genetics was revolutionized by the development of new genomic resources including the reference genomes of wild and cultivated wheat, re-sequenced populations of genetically and geographically diverse lines, TILLING populations, and gene expression atlases. Combined with phenomic data these resources have substantially accelerated the identification of genomic regions harboring causal genes and variants that underlie trait variation of agronomic significance. In spite of these advances, our understanding of the genetic control of complex traits remains rather limited. To advance the construction of genome-to-phenome map in wheat, we have performed RNA-seq profiling of a genetically and geographically diverse population of wheat accessions and investigated the genetic basis of gene expression variation. Integrated analyses of eQTL, chromatin accessibility and epigenetic data was used to identify cis- and trans-regulatory SNPs affecting transcript levels in the wheat genome. The data was used to build a gene regulatory network and demonstrate the importance of genetic redundancy created by two rounds of whole genome duplication and repetitive DNA expansion in regulating gene expression levels in wheat tissues. Strong enrichment of gene expression controlling variants in wheat GWAS was detected confirming the relevance of identified eQTL in defining the studied traits. The joint effects of genomic variants affecting both gene expression levels and phenotypic variation were analyzed resulting in identification of pathways and genes underlying agronomic trait variation in wheat. Our study demonstrates how transcriptome and chromatin accessibility dimensions added to the genome-to-phenome map can help to establish connection between genomic variation, gene expression and organism-level phenotypes, improve our ability to predict phenotypic outcomes of a particular genotype, and, in the future, facilitate selection of targets for engineering a biological system with the desired properties.

## **W563: Genomic Selection and Genome-Wide Association Studies**

### **A Multi-Omics Approach to Improving the Metabolic Health of Dairy Cows**

**Timothy D. W. Luke**<sup>1</sup>, Thuy T. T. Nguyen<sup>2</sup>, Simone J. Rochfort<sup>3</sup>, Hans D. Daetwyler<sup>3</sup> and Jennie E. Pryce<sup>3</sup>, (1)Agriculture Victoria Research and La Trobe University, Bundoora, VIC, Australia, (2)Agriculture Victoria Research, Australia, (3)Agriculture Victoria Research and La Trobe University, Australia

Improving animal health and resilience is an increasingly important breeding objective for the dairy industry. Most disease events affecting dairy cows occur in the first 30 days after calving, and many of these events are associated with metabolic disorders. One of the most sensitive and objective ways of monitoring the metabolic health of cows is serum metabolic profiling, which employs well-established epidemiological associations between the concentrations of several metabolites in serum, and the occurrence of periparturient diseases. We used data collected from a genotyped female reference population (approximately 1400 animals), to estimate the genetic parameters and the accuracies of genomic prediction for a suite of serum biomarkers of health in early lactation dairy cows. Heritability estimates varied significantly (0.07 to 0.49) as did the magnitude of genetic correlations between traits (0.01 to 0.96). Interestingly, most genetic correlations were favourable suggesting that selecting for normal concentrations of one biomarker may result in improvements in the concentrations of other biomarkers. Accuracies of genomic prediction were low (0.05 to 0.27) but were consistent with trait heritabilities and a small reference population. In order to increase genomic prediction accuracies, we are working to increase the size of the reference population using novel high-throughput phenotyping technologies such as mid-infrared spectroscopy of milk. We are also using molecular phenomic techniques such nuclear magnetic resonance spectroscopy to better understand the complex biology

underlying these metabolic biomarkers, and to identify new intermediate phenotypes. Our aim is to combine information from these different omics datasets to identify and prioritize important functional variants, and thereby increase genomic prediction accuracies.

#### **W564: Genomic Selection and Genome-Wide Association Studies**

##### **Genomic Prediction for Plant Breeding: Beyond GEBVs and Cross Validation**

**Aaron Lorenz**, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Genomic prediction was introduced as a concept nearly 20 years ago and gained attention in the plant breeding world over 10 years ago. Since this time, countless research papers have been published on the potential and optimization of the genomic prediction approach for predicting genetic values of new progenies. As the field transitions into the next decade, it should pivot towards research for expanding the scope of selection and genomic mating, as well as practical application of genomic selection in actual cultivar development programs. Research and challenges related to these lines of inquiry will be presented from the perspective of a university plant breeder.

#### **W565: Genomic Selection and Genome-Wide Association Studies**

##### **The Fascinating Genetic Architecture of Honeybees**

**Gertje Petersen**, AbacusBio, Dunedin, New Zealand, Peter Dearden, Department of Biochemistry, University of Otago, New Zealand; Genomics Aotearoa, Dunedin, New Zealand and Peter F. Fennessy, AbacusBio Ltd, New Zealand

With the increase in focus on our diet composition, crops with high nutritional and monetary value such as nuts, fruits and vegetables are increasingly in demand. In addition to high water requirements resulting from this increased demand, more pressure is put on insect pollinators to ensure the fruit set and subsequent production. Since production is often localised in areas with good growing conditions, pollinators need to be managed proactively. After decades of sinking numbers of commercial beekeepers due to the constant availability of cheap industrial sugars replacing honey, commercial beekeeping is once again on a rise as a result. As the only insect pollinator that can be managed at scale, the European or Western honeybee, *Apis mellifera*, finds itself at the forefront of commercial provision of pollination services.

Breeding strategies for honeybees have evolved away from experimental crosses (such as the ones done by Gregor Mendel) towards scientifically backed selection based on genetic evaluation of candidate queens. Simultaneously, honeybee research has been flourishing, often focused on a globally distributed ectoparasitic mite, *Varroa destructor*, and the mechanisms of resistance or tolerance observed in particular populations of bees. However, although both quantitative and molecular honeybee genetics are on the rise, they are facing some restrictions related to lack of fundamental knowledge on the details of bee biology and the peculiar architecture of honeybee populations. Polyandry, long-term sperm storage by the queen and mating on the wing bring with them some obstacles to a structured breeding program through the absence of a traditional pedigree and difficulties surrounding the mating management, while haplodiploidy, lack of defined breeds and exportation/importation bottlenecks during the distribution of European honeybees across multiple continents create peculiarities in their genetic architecture.

In addition to the realities of honeybee genetics, social aspects of both honeybees (eusocial organisation in colonies of thousands of individuals) and beekeepers (a shift from a hobby to industrialised livestock farming) can influence the feasibility of the application of genomics methodologies. At the same time, genomics offers attractive solutions to some of the conundrums in bee breeding, for example closing the gap of missing pedigree information.

Walking the line between pure honeybee research, e.g. answering fundamental questions on bee biology, investigating disease resistance mechanisms and other important traits and ensuring that

outcomes are relevant to the modern beekeeping environment as well as conservation efforts is one of the most challenging tasks for today's honeybee geneticists. Finding a way to address this tension does however offer an exciting reward: contribution to the understanding and improvement of a species with wide-reaching economic, environmental and social impact through securing global food supplies.

#### **W566: Genomics of Crop Ecosystem Services**

##### **Towards a Viable Perennial Grain Sorghum through Integrated Genetic and Economic Analysis**

**Tara Maireid Conway**, Plant Genome Mapping Laboratory, UGA, Athens, GA, Wenqian Kong, University of Georgia, Athens, GA, T. Stan Cox, The Land Institute, Salina, KS and Andrew H. Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

The replacement of annual cereal crops with perennial varieties has great potential for increased ecosystem resilience and improved food security. Successful establishment of such cropping systems requires development of both system “hardware” (perennial grain varieties) and “software” (agronomic and ecological systems facilitated by the perennial germplasm). Integration of the “hardware” and “software” development holds great potential for increasing the efficiency of perennial grain system establishment. Sorghum (*S. bicolor*) the world's fifth most important cereal crop, is an ideal target for perenniality due to its close relation to two wild perennial grasses, *S. halepense* and *S. propinquum*, and its potential to grow across highly diverse landscapes. The key to perenniality lies in the development of rhizomes, subterranean stems that sprout to form the next season's crop. Quantitative trait locus (QTL) analysis of a novel F2:3 population derived from an *S. bicolor* x *S. halepense* cross illuminated genomic regions pertinent to perenniality and can contribute to marker-assisted selection in breeding. Simultaneous investigations into Georgia sorghum farm budgets highlighted differential yield targets contingent on perennial cropping scenario as well as other breeding targets to further improve viability for Georgia farmers.

#### **W567: Genomics of Crop Ecosystem Services**

##### **Applied Genomics to Improve IWG Breeding Efficiency**

**Jared Crain**, Kansas State University, Manhattan, KS, Jesse Poland, Department of Plant Pathology, Kansas State University, Manhattan, KS and Lee DeHaan, The Land Institute, Salina, KS

Twenty-first century genomic tools provide a complement to traditional phenotypic selection, especially for developing new crops such as perennial grains that can provide ecosystem services in addition to food and fiber. Ecosystem services facilitate soil and water quality improvement, increased nutrient cycling and retention, and enhanced soil carbon sequestration, providing long-term environmental benefits. Since the 1980's, intermediate wheatgrass (IWG, *Thinopyrum intermedium*) is a species that has been undergoing domestication to establish a perennial grain crop. While phenotypic selection has substantially improved grain yield and seed size, it is estimated to take a minimum of an additional 20 years of equal phenotypic selection gains to reach similar yield of annual wheat. The advent of affordable molecular markers has provided the opportunity to use genomic selection (GS) within IWG breeding. This has reduced the selection cycle from two years to one year, theoretically doubling selection gains. Beginning in 2017, The Land Institute (Salina, KS) has been using GS, recently completing the 3<sup>rd</sup> cycle of GS in 2019. Current results indicate that over 10% selection gains per cycle can be made in desired traits including free-threshing, reduced shattering, seed size and yield. Along with driving genetic gains, GS model predictions and observed traits have been highly correlated (up to  $r = 0.73$ ) for many traits. Application of GS has the potential to rapidly develop IWG, leading to a perennial grain that can provide economic and environmental benefits along with increased ecosystem services.

#### **W568: Genomics of Crop Ecosystem Services**

##### **Silphium Population Genomics: Progress and Applications to Breeding Programs**

**Andy Raduski**<sup>1</sup>, Adam Herman<sup>2</sup>, Nolan Kane<sup>3</sup>, David Van Tassel<sup>4</sup> and Yaniv Brandvain<sup>2</sup>, (1)University of Minnesota, St. Paul, MN, (2)University of Minnesota, Saint Paul, MN, (3)University of Colorado, Boulder, CO, (4)The Land Institute, Salina, KS

*Silphium integrifolium*, a long-lived, stress tolerant species, is being domesticated as a potential perennial oilseed crop. To date, domestication progress has been made without a deep understanding of the levels and distribution of genetic variation within and across wild populations. Here I will describe research into developing '-omic' resources for *Silphium* and their integration into existing breeding programs. Specifically, population transcriptomic data from wild plants grown in a common garden have guided future collections of diverse germplasm to include into breeding programs. Interestingly, the populations that are the source of the breeding program are genetically depauperate compared to the rest of the species range, suggesting that large amounts of potentially useful variation exist outside of the existing breeding program. Recent advances towards a genotyping platform based upon the *Silphium* exome will also be discussed.

## **W569: Genomics of Crop Ecosystem Services**

### **High Throughput Phenotyping and Population Genomics of *Lupinus polyphyllus***

**Matthew Rubin**<sup>1</sup>, Claudia Ciotir<sup>2</sup>, Niyati Bhakta<sup>3</sup>, Emma Frawley<sup>4</sup>, Zachary Harris<sup>5</sup>, Brandon Schlautman<sup>6</sup> and Allison Miller<sup>3</sup>, (1)Donald Danforth Plant Science Center, Olivette, MO, (2)University of Haifa, Israel, (3)Donald Danforth Plant Science Center, (4)Washington University, (5)Saint Louis University, St. Louis, MO, (6)The Land Institute, Salina, KS

The majority of domesticated plant species are herbaceous annuals and woody perennials. However, there are many herbaceous perennial species that hold potential for future agricultural systems. In addition to multiyear harvest, herbaceous perennials (e.g. members of the legume family) provide many ecosystem services including lessening of erosion as a result of their large and continually present root systems. We performed a population genetic study and phenotypic survey of *Lupinus polyphyllus* which included populations from the native range in western North America, feral populations from Europe and New Zealand, and cultivated varieties. Using a genotyping-by-sequencing (GBS) approach, we identified ~10,000 single nucleotide polymorphisms (SNPs). We found levels of genetic diversity varied within and among populations in the native range, and that feral populations are genetically similar to cultivated varieties. We quantified variation in seed area and found that cultivated seeds were larger than seeds from natural populations. Seed area was likely not the target of selection during the development of the forage and ornamental varieties of *L. polyphyllus* but rather a correlated response to selection on another trait. The population genetic results and observed phenotypic change in seed area suggest that there is sufficient standing genetic variation within natural populations to initiate a pre-breeding program focused on the development of *L. polyphyllus* as a perennial agricultural crop.

## **W570: Genomics of Genebanks**

### **Genomic Approaches for Exploring and Utilizing Germplasm Diversity in Lentil**

**Hai Ying Yuan**, Bert Vandenberg and Kirstin Bett, University of Saskatchewan, Saskatoon, SK, Canada

Domestication and adaptation history of cultivated lentil (*Lens culinaris*) produced bottleneck effects resulting in a narrow genetic base in breeding programs. The diversity of various traits exhibited by wild lentils as well as exotic material makes them potential genetic resources for sustained improvement of cultivated lentil. Introgression of favorable genes from exotic germplasm and crop wild relatives is necessary for adapting crop varieties to changing disease pressures, environmental stresses and emerging market demands. Therefore, there is a need for tools to better understand germplasm being used in introgression breeding. The availability of a sequenced genome coupled with next-generation sequencing (NGS) technologies provides viable options to uncover useful genetic variants within lentil germplasm. An overview of the genomic tools that were developed and used within our group for classification of the germplasm, trait genetic analysis and the identification of potential variants for marker-assisted selection will be presented. Canada is one of the world's largest producers and exporters of lentil. These genomic tools are enabling us to identify and incorporate important allelic

variation from many lentil germplasm collections to ensure we continue to develop varieties for a growing industry.

## **W571: Genomics of Genebanks**

### **Diversity and Structure of the USDA *Fragaria* x *Ananassa* Collection**

Jason D. Zurn<sup>1</sup>, Michael A. Hardigan<sup>2</sup>, Steven J. Knapp<sup>2</sup>, Kim Hummer<sup>1</sup> and Nahla Bassil<sup>1</sup>, (1)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (2)Department of Plant Sciences, University of California, Davis, CA

The cultivated strawberry (*Fragaria* x *ananassa*) arose through a chance hybridization of two wild American species in Versailles, France, in the 1750s. Since then, breeders have improved flavor, production traits, harvest season, and plant adaptation of this important crop. Diverse germplasm is crucial for strawberry breeders to meet next generation needs in a climate-changing world. The USDA-ARS National Clonal Germplasm Repository in Corvallis, Oregon, maintains the U.S. strawberry collection. In the 1980s, a core subset was selected based on morphological traits and origin, to represent genetic diversity. Recent development of high-throughput genotyping platforms could provide a better understanding of the diversity and structure of a germplasm collection, while offering insights for efficient plant management. Our objective was to evaluate the diversity of this collection and define core collections based on molecular techniques. Genotyping was conducted on 534 *F. xananassa* samples using either the iStraw35 or the 50K octoploid strawberry axiom array. These arrays share 5,809 markers that are distributed across the 'Camarosa' genome assembly. Data for the shared markers were curated for call quality, missing data, and minor allele frequency resulting in 4,033 markers. K-means clustering analysis revealed eight sub-populations associated with different geographic breeding centers. Two core collections were determined: one represented a uniform distribution of the gene space, and the other its maximum genetic diversity. Pedigrees were confirmed for triads and sports in the collection. Genotypic data for this collection will be publically available and could allow identification of genomic regions controlling valuable traits when phenotypic data are obtained.

## **W572: Genomics of Genebanks**

### **Characterizing and Comparing International Soybean Germplasm Collections**

Anne V. Brown, USDA-ARS-CICGRU, Ames, IA

There are over 230,000 soybean accessions in germplasm repositories worldwide, making the identification of useful germplasm difficult. As high throughput genotyping costs have dropped, the dense genotyping of large germplasm collections has become feasible. Nevertheless, significant challenges remain due to the sheer volume of such genotype data. Therefore, tools to investigate this data are needed. A new on-line tool called GCViT (Genotype Comparison Visualization Tool) is designed to further investigate and visualize data contained within VCF (Variant Call Format) files. GCViT allows users to pick a set of accessions within a dataset and visually see where there are similarities or differences on the chromosomes between the given accessions. This enables a user to identify regions of interest, track introgressions and visualize pedigree relationships. This talk will provide examples and demonstrate how this tool is beneficially to plant breeders worldwide.

## **W573: Genomics of Genebanks**

### **Genebank Genomics - a Barley Case Study**

Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Plant Genetic Resources (Plant GenRes) may hold the key for adapting crops to a changing climate. Their actual use in crop improvement, however, is limited and is in stark contrast to their potential and promise. Major international initiatives like DivSeek have advocated for systematically generating genotypic and phenotypic data for genetic resources stored in genebanks. At IPK we have accomplished a pilot study on the genotyping of all >20,000 barley accessions of our ex situ gene bank. This study has opened new avenues in the way we can make information available to users of genetic resources, the

way we (want to) manage genetic resources and it has revealed the composition of global domesticated barley diversity – laying the foundation for exploring the pan-genome of the species for the profit of research and application. By the barley example, IPK has initiated in the context of other national or international collaborative projects the systematic genotyping of further collections such as of wheat, *Solanaceae* crops, *Phaseolus* beans thus paving the way of turning the gene bank from a seed archive into a bio-digital resources center.

#### **W574: Genomics of Genebanks**

##### **Diversity Lost and Diversity Regained: The Promise and Perils of Tapping Crop Wild Relatives for Improvement**

**Loren H. Rieseberg**, University of British Columbia, Department of Botany and Biodiversity Research Centre, Vancouver, BC, Canada

Genetic diversity is indispensable for crop improvement. However, most crops have undergone a domestication bottleneck and harbor less genetic diversity than their wild relatives. Thus, there is considerable interest in reaching across the crop-wild boundary to re-acquire useful genetic variation that was lost during domestication and improvement. Here I explore both positive and negative impacts of wild introgressions and discuss how the latter can be minimized, using cultivated sunflower (*Helianthus annuus* L) and its wild relatives as a case study. Building on a data set of several reference sequences, numerous genetic maps, and > 2,400 re-sequenced genotypes, I show that around 10% of the cultivated sunflower genome is derived from introgressions with wild species. These introgressions have introduced new disease resistance alleles into cultivated sunflowers, as well as components of the hybrid production system. On the other hand, numerous chromosomal rearrangements differentiate cultivated sunflower and wild populations. For example, a frequently employed wild donor species, *H. petiolaris*, differs from the cultivated sunflower by 6-8 translocations and 50-60 inversions, many of which are associated with important traits such as seed size and tolerance to low fertility soils. These rearrangements hamper introgression from much of the genome and, if successfully introgressed into cultivars, can introduce genetic load and reduce recombination rates. Gene presence/absence variation, which affects 27% of the genes in the cultivated sunflower pan genome, is also often associated with introgressions from wild species. While gene presence/absence variation appears to underlie key agronomic traits in cultivated sunflower, including disease resistance, fertility restoration, and flowering time, it also likely represents an important cause of linkage drag. Such negative consequences of wild introgressions can be reduced by focusing pre-breeding efforts on a crop's least divergent wild relatives and by moving to a hybrid production system, which permits complementation of deleterious introgressions.

#### **W575: Genomics of Non-Classical Model Animals**

##### **The Importance of *cis*-Regulatory Divergence for Eye Degeneration in Subterranean Mammals**

**Michael Hiller**, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

I will present our work on combining comparative and functional genomics to investigate the fate of *cis*-regulatory elements in subterranean mammals that exhibit highly-degenerated eyes. Using "Forward Genomics" to perform a genome-wide screen for increased divergence in selected evolutionary lineages, we identified thousands of conserved non-coding genomic regions that exhibit preferential sequence divergence in independent subterranean lineages. These diverged genomic regions are significantly associated with genes having key roles in eye development and function. Furthermore, these diverged regions significantly overlap regulatory elements active during mouse eye development, as identified by ATAC-seq. We show that sequence divergence in these genomic elements resulted in an extensive loss of relevant transcription factor binding sites. We used Crispr-Cas9 to replace the sequence of a mouse eye regulatory element by the orthologous blind mole rat sequence and found that this genetic change is sufficient to change target gene expression in the lens. Together, our genome-wide analysis suggests that not only losses of eye-related genes but also decay of eye-related regulatory elements contribute to eye degeneration in subterranean mammals.

## **W576: Genomics of Non-Classical Model Animals**

### **Detection and Interpretation of Archaic Ancestry in Extant Humans**

Nathan K. Schaefer, University of California - Santa Cruz, Santa Cruz, CA, Beth Shapiro, Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA and **Richard E. Green**, University of California, Santa Cruz, Santa Cruz, CA

Many humans carry genes from Neanderthals, an important legacy of past admixture. Several methods have been described for detecting this archaic hominin ancestry within human genomes using patterns of linkage disequilibrium or direct comparison to Neanderthal genomes. Each of these methods is limited in sensitivity and scalability. We describe a new ancestral recombination graph inference algorithm that is scalable to large genome-wide data sets and demonstrate its accuracy on real and simulated data. We then generate the first genome-wide ancestral recombination graph of both human and archaic hominin genomes. From this, we generate a map within human genomes of archaic ancestry and of genomic regions devoid of genes shared with archaic hominins by either admixture or incomplete lineage sorting. We find that only 1.5-7% of the modern human genome is uniquely human. We also find evidence of at least two bursts of adaptive changes specific to modern humans within the last 600,000 years, consisting of both coding and regulatory changes, many of which may relate to brain development and function.

## **W577: Genomics of Non-Classical Model Animals**

### **What Comparative Studies of Parrot Genomes Can Teach Us about Longevity, Large Brains and Cognition**

**Claudio V. Mello**, Department of Behavioral Neuroscience, Oregon Health and Science University, Portland, OR

Parrots have highly distinct and intriguing characteristics, including a large brain expansion, highly developed cognitive and vocal communication abilities, and a significantly longer lifespan than other birds of comparable size. To address the genetic basis of these unique traits, we have examined the genomes of several parrot species in comparison with birds and some outgroups of comparative relevance. This effort included generating a high-coverage, annotated assembly of the genome of the blue-fronted Amazon (*Amazona aestiva*), a highly vocal and long-lived species emblematic of the Brazilian avifauna, and carrying out comparisons with 30 other avian species, including 4 other parrots. Taking advantage of methods previously developed for analysis of other avian genomes, we identified several genomic features unique to parrots, including novel genes and modifications to coding and regulatory sequences of known genes. We also identified genomic features under selection in parrots and other long-lived birds, including genes previously associated with lifespan determination (from the GenAge database) as well as several hundred new candidate genes. These genes support a range of cellular functions, including telomerase activity, DNA damage repair, control of cell proliferation, cancer, immunity, and anti-oxidative mechanisms. We also identified brain-expressed, parrot-specific paralogs with known functions in neural development or vocal-learning brain circuits. Parrot-specific changes in otherwise ultra conserved non-coding elements (UCNEs) were also detected that were associated with genes linked to cognitive abilities and that have undergone similar selection in the human lineage, suggesting convergent evolution. These findings contribute insights into the genetics and evolution of longevity and cognition, and suggest novel targets to be explored further by mechanistic studies in genetic model organisms.

## **W578: Genomics of Non-Classical Model Animals**

### **Everyone Is Your Friend! the Molecular Architecture of Hypersocial Canines**

**Bridgett vonHoldt**, Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ

Although considerable progress has been made in understanding the genetic basis of morphologic traits in dogs and wolves, the genetic basis of their behavioral divergence is poorly understood. We analyzed a 5-Mb genomic region on chromosome 6 previously found to be under positive selection in domestic dog

breeds and contains putative candidate genes. A deletion of this region in humans is linked to Williams-Beuren syndrome (WBS), a multisystem congenital disorder characterized by hypersocial behavior. Further, this behavior is also a core element of domestication that distinguishes dogs from wolves. Here, we associated quantitative data on behavioral phenotypes symptomatic of WBS in humans with structural changes in WBS-linked genes in dogs. We found that this region also harbors a large number of highly polymorphic structural variants in canines, some of which are private to an individual dog or breed. Notably, our study revealed a statistically significant association between transposable element (TE) insertions in *GTF2I* and *GTF2IRD1*, basal transcription factors that regulate vertebrate development, with measures of human-directed social behavior typical of WBS. With a larger sample size and leveraging behavioral phenotypes from breed stereotypes, we found a significant association between the copy number of these TEs and behavior. Hence, it is conceivable that selection acting on hypersociability-associated TEs may have helped shape the evolution of the canid family. Further, our findings suggest that there are commonalities in the genetic architecture of WBS and canine tameness and that directional selection may have targeted a unique set of linked behavioral genes of large phenotypic effect, allowing for rapid behavioral divergence of dogs and wolves, facilitating coexistence with humans. In light of our findings, we propose a unifying hypothesis to explain one aspect of canid domestication, where individuals with hypersocial tendencies were favored under selective breeding, accentuating a behavior likely influenced by structural variants in the canine WBS locus. Our findings provide insight into one genetic mechanism by which the hypersocial response of domestic dogs toward humans compared with human-reared wolves can be acted on and shaped by selection during species domestication. This mechanism is expected to predispose dogs for hypersocial responses toward any bonded companion. This is consistent with the finding that domestic dogs appear to maintain, or even increase, the duration of social engagements with humans and conspecifics as they approach adulthood, with the opposite trend found in wolves. In summary, our findings suggest that the same region affected by structural variants in human WBS is associated with the exuberant sociability of domestic dogs.

## **W579: Genomics of Phytoremediators, Metal Accumulators and Relatives**

### **Identification of Functionally Redundant Genes in Heavy Metal Tolerance and Accumulation**

**Qingqing Xie**, Division of Biological Sciences Cell and Developmental Biology Section & Ctr for Mol. Genetics, San Diego, CA and Julian I. Schroeder, University of California, San Diego, La Jolla, CA

Heavy metal contamination in soils caused by human activities is an urgent problem. Heavy metals impair plant growth and survival. Therefore, plants have developed a complex and dynamic system to detoxify and sequester heavy metals to minimize detrimental effects. While previous studies mainly concentrated on the identification of the function of a single gene, identification and characterization of functionally overlapping genes that mediate rapid detoxification in plants has not been previously feasible. Artificial microRNAs (amiRNAs) are small non-coding RNA molecules that down-regulate gene expression in plants by targeting mRNAs for cleavage or translational repression. amiRNAs provide a powerful strategy to alleviate the limitation of functional redundancy in traditional screening methods in plants. Therefore, our laboratory has computationally designed genome-wide subclade-specific amiRNA libraries with a high-performance computing cluster, for genome-wide knock-down of homologous gene family members (F. Hauser, et al. Plant Cell). We have now generated over 14,000 Arabidopsis T2 lines expressing these amiRNAs for use by the community (F. Hauser, P. Ceciliato, et al. J. Exp. Bot. 2019). AmiRNA lines have been screened in response to cadmium or arsenic treatment to identify new genes that affect this toxic heavy metal and metalloid tolerance. We report here new genes that are key to cadmium and arsenic resistance and accumulation. We have isolated amiRNA lines in chloroplast transporters, phosphate uptake transporters, ERF transcription factors, protein kinase superfamily members and many other interesting homologous gene sets that affect arsenic and/or cadmium sensitivity.

In summary, our results provide novel insights and uncover novel genes and mechanisms involved in heavy metal tolerance and responses in plants.

## **W580: Genomics of Phytoremediators, Metal Accumulators and Relatives**

## The Genomics of Endophytic Biofilms in the Phytoremediation of Arsenic

**Robert Tournay**, Sharon L. Doty, Andrea Firrincieli, Shruti Parikh and Dominic Sivitilli, University of Washington, Seattle, WA

A legacy of agricultural and industrial activities has resulted in contaminated soils and groundwaters on a global scale. Because of their persistence in the environment and toxicity at low levels, the heavy metals and metalloids, such as cadmium, lead, and arsenic are of particular concern. Phytoremediation provides a low-cost, eco-friendly solution for the remediation of environmental pollutants. Important goals for improving phytoremediation outcomes lie in reducing phytotoxicity and increasing accumulation of the pollutants in harvestable tissues. Endophytes, beneficial microorganisms internally colonizing plants, may improve phytoremediation outcomes directly by detoxifying the pollutants, and/or indirectly by improving plant health; e.g. promoting root growth, reducing stress responses, and increasing nutrients acquisition.

The formation of biofilms, assemblages of microorganisms embedded in a three-dimensional matrix of extracellular polymeric substances (EPS), is a common mechanism by which microbes protect themselves from environmental stresses; e.g. turbulence, desiccation, UV, and predation. In clinical bacterial strains, biofilm formation has been associated with increased microbial resistance to antibiotics and toxic metals, and while the role of EPS in plant pathogens is well studied, little is known about its role with endophytes and their host plants. Genome sequence data have revealed that EPS synthesis genes are common in endophytes and we are investigating whether the heavy metal/metalloid detoxification mechanisms of endophytic biofilms might benefit phytoremediation efforts by reducing phytotoxicity.

We have sequenced the genomes of four arsenic-tolerant endophyte strains that form biofilms under arsenic stress. To identify candidate genes involved in arsenic-induced biofilm formation, a genome analysis is underway on three endophyte strains, *Pseudomonas putida* PD9R, *Rahnella aquatilis* PD12R, and *Pseudomonas koreensis* WW6. The fourth strain, *Enterobacter* sp. PDN3, was subjected to random barcoded Tn-Seq experiments in both arsenate and arsenite. The Tn-Seq technology will provide a global, top-down look at the genes required for arsenic tolerance.

## W581: Genomics of Phytoremediators, Metal Accumulators and Relatives

### Understanding High Potassium Stress Responses of the Extremophyte, *Schrenkiella parvula* using Systems Biology Approaches

**Pramod Pantha**<sup>1</sup>, Dong-Ha Oh<sup>1</sup>, David Longstreth<sup>2</sup> and Maheshi Dassanayake<sup>1</sup>, (1)Louisiana State University, Baton Rouge, LA, (2)Louisiana State University

*Schrenkiella parvula*, a close relative of *Arabidopsis thaliana* and Brassica crops, grows in the shores of Lake Tuz, Turkey, one of the largest hypersaline lakes in the world. Soils in the Lake Tuz region is high in multiple salts including Na<sup>+</sup> and K<sup>+</sup>. *S. parvula* can complete its life cycle in the presence of multi-ion salt concentrations lethal to most plants including *A. thaliana*. In this study, we investigated the underlying genomic basis for high K<sup>+</sup> tolerance using physiological, ionic, transcriptomic, and metabolomic approaches between stress-adapted *S. parvula* and the stress-sensitive *A. thaliana* treated with 150 mM KCl for four successive time points (0, 3, 24, and 72 hours after treatment, HAT).

In response to high K<sup>+</sup>, both the primary and lateral root architecture changes significantly compared to control conditions, and the phenotypic change in *A. thaliana* is more pronounced than the effect observed for *S. parvula*. We further showed KCl is more toxic than NaCl at similar osmotic strength.

The ionome profiles of *A. thaliana* showed increased accumulation of K<sup>+</sup> in both root and shoot over time whereas it was only observed in 72HAT for *S. parvula*. Under K<sup>+</sup> stress *A. thaliana* was unable to maintain the macro and micro-nutrient homeostasis while *S. parvula* remained unaffected.

We conducted co-expressed gene network analysis to identify genetic mechanism that may lead to the differential ionic responses in these two species. The *S. parvula* transcriptome responds to high K<sup>+</sup> stress by adjusting only a limited number of genes in selected pathways whereas *A. thaliana* shows a large number of genes significantly changed. Clusters of co-expressed ortholog pairs showed the “stress-preparedness” in *S. parvula* and those clusters were enriched in response to salt stress and transport function. The largest down-regulated clusters in *A. thaliana* roots were enriched with root hair formation and cell wall organization, supporting the physiological response of root growth interruption in response to high K<sup>+</sup> stress. Contrastingly, response to salt stress, transport, and ABA regulated pathways were enriched in the up-regulated gene networks in *A. thaliana* shoots. In line with the reduced photosynthesis observed in *A. thaliana* under stress, shoot transcriptomes also showed significant reduction in gene expression of pathways related to photosynthesis and cell cycle.

At the metabolome level, similar to the transcriptomic response, *S. parvula* showed minimal yet distinct responses. For example, accumulation of protective osmolytes and antioxidants including proline, myo-inositol, and their precursors in *S. parvula* root and shoot were highly increased under stress.

In conclusion, we observed reduced photosynthesis, alteration in root architecture, and cell wall re-organization as major responses to high K<sup>+</sup> stress in *A. thaliana* while *S. parvula* showed minimal interruption to growth and development, facilitated by its stress-preparedness strategies exemplified at all levels. The ionic, transcriptomic, and metabolomic profiles supported the physiological responses to high K<sup>+</sup> stress.

## **W582: Genomics of Phytoremediators, Metal Accumulators and Relatives**

### **Extremophyte Arabidopsis Relatives Thrive in Soils with Toxic Concentrations of Sodium**

**Kieu-Nga T Tran**, Dong-Ha Oh and Maheshi Dassanayake, Louisiana State University, Baton Rouge, LA

Salinity stress is one of the major threats that affects crop yield worldwide. Yet, we have not been successful in developing salt tolerant cultivars for many of our major crops. One bottleneck to overcome this challenge has been our lack of understanding of how naturally adapted plants respond and thrive under salt stress using orthologs of genes known to regulate salt stress responses in crops. To address this need, the emerging extremophyte models, *Schrenkiella parvula* and *Eutrema salsugineum* have been identified as unique systems to study genetic mechanisms underlying stress adaptation. These plants are closely related to the premier model, *Arabidopsis thaliana*. This allows more accurate transfer of knowledge on gene functions to identify key regulatory networks using comparative genomics and systems biology approaches. In this study, we used a comparative transcriptomic, ionic, and a metabolomic experimental design to gain insight on how these extremophytes respond to salt stress differently from the stress sensitive model, *A. thaliana*.

We found that *S. parvula* and *E. salsugineum* accumulated less Na<sup>+</sup> and maintained higher K<sup>+</sup> in the shoots compared to *A. thaliana* under salt stress. However, Na<sup>+</sup> accumulation in *E. salsugineum* was comparable to that of *A. thaliana* in the roots. Among the two extremophytes, only *S. parvula* maintained a low Na<sup>+</sup> concentration in the roots upon salt exposure. Salt stress led to a reduction in the abundance of macro and micro-nutrients in *A. thaliana* roots, while both halophytes could maintain their overall nutrient status similar to control levels at comparable salinities.

Concordant with the ionomes, the shoot metabolomes of *S. parvula* and *E. salsugineum* showed minimal changes compared to *A. thaliana*. Among the notable changes to the root metabolome, *S. parvula* increased in abundance of primary sugars, amino acids, and other intermediates in the TCA cycle and ATP biosynthesis while the abundance of these metabolites decreased in *E. salsugineum* under salt stress. However, this reduction did not lead to a significant net reduction of these metabolite pools as *E. salsugineum* had higher basal levels of these metabolites in roots compared to *S. parvula* and *A. thaliana*. Both extremophytes indicated distinct yet different strategies for stress preparedness at the metabolome level during salt stress.

The transcriptomic responses to salt stress further supported the metabolic adjustments of each species. Up-regulated genes in all three species were enriched in stress-related pathways including ROS scavenging and osmolyte biosynthesis. In addition, up-regulated genes in *S. parvula* were enriched in various catabolic processes that suggested the use of alternative energy sources to support growth during salt stress. Genes associated with photosynthesis were enriched in down-regulated transcripts in all three species. However, the transcriptomic effect on photosynthesis was significantly less in the extremophytes. The extremophytes showed stress preparedness, both at the transcriptome and metabolome levels to allow sufficient nutrient uptake to promote growth and development under salt stress levels that impaired growth in *Arabidopsis*.

### **W583: Genomics of Plant Development**

#### **Structural Genome Variations Associate with Juvenile Development and Flowering Time Modulation in Elite Winter Canola (*Brassica napus*)**

**Paul Vollrath**, Justus Liebig University, Giessen, Germany

A multiparental mapping population created by German canola breeding companies, comprising 354 DH lines from crosses among seven commercial winter canola (*Brassica napus*) varieties has been studied in three-year field trials in eight locations. The population was assessed for flowering time and juvenile development before winter. All 354 accessions were genotyped using the *Brassica* 60K Illumina Infinium SNP array and genome-wide association studies (GWAS) were performed. 60K SNP array data was also filtered for markers segregating with Single Nucleotide absence Polymorphism (SNaP) patterns. Inclusion of SNaP markers in genetic mapping revealed a number of QTL affected by presence-absence variation (PAV). To dissect the role of long- and short-range PAV events in more detail, the genomes of the seven parental lines of the multiparental mapping population were sequenced with a 20x genome coverage using the MinION device from Oxford Nanopore Technologies. The sequencing data allowed us to detect deletions and insertions in elite canola genomes and evaluate their role in juvenile development as well as flowering time modulation.

### **W584: Genomics of Plant Development**

#### **Whole Genome Duplication by Demand: Biotroph-Induced Localized Host Endoreduplication**

**Mary Wildermuth**, Dept of Plant and Microbial Biology, Berkeley, CA

### **W585: Genomics of Plant Development**

#### **Soybean Tilling By Target Exome Sequencing: Applications in Disease Resistance and Seed Composition**

**Mallory A. Cullen**<sup>1</sup>, Naoufal Lakhssassi<sup>2</sup>, Zhou Zhou<sup>3</sup>, Abdelhalim El Baz<sup>1</sup> and Khalid Meksem<sup>3</sup>, (1)Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL, (2)Southern Illinois University, Carbondale, IL, (3)Southern Illinois University at Carbondale, Carbondale, IL

Soybean (*Glycine Max* (L.) Merr) is an important agricultural crop that is grown worldwide mainly for the production of oil and protein meal. Soybeans are a host for a variety of pests and diseases that limit soybean production. The soybean cyst nematode (SCN), the most damaging pest of soybeans, is capable of causing over 1 billion US dollars in yield losses annually. The best disease management tool for diseases such as the soybean cyst nematode is through planting resistant lines.

Targeting induced local lesions in the genome (TILLING) is a method for high throughput screening of a mutagenized population. TILLING can be utilized as a reverse genetics approach to identify the functions of genes in different pathways. Our study focuses on using TILLING by Target Exome Sequencing for the identification of induced mutations within genes related to disease resistance pathways and seed composition traits. A mutant library of 4000 mutagenized soybeans from a resistant cultivar Forrest was created, DNA from every mutant family was extracted and arrayed in 2 dimensional pools, probes designed targeting the exomes of several genes within the soybean genome was performed. After

Illumina sequencing, the data revealed an abundance of single nucleotide polymorphisms (SNPs) between the mutants and wild type Forrest. The purpose of this work is to identify SNPs within these genes to better understand their roles in disease resistance and seed composition networks.

#### **W586: Genomics of Plant Development**

##### **Rice Circadian Clock Regulates Tillering and Panicle Development through Sugar Response and Strigolactone Signaling**

**Z. Jeffrey Chen**, The University of Texas at Austin, Austin, TX

#### **W587: Genomics of Plant Development**

##### **Transcriptomic Response of *Sorghum bicolor* to a Brassinosteroid Inhibitor: Gene Targets for Developing a “Smart Canopy”**

**Martha Ibore**<sup>1</sup>, Maria Betsabe Mantilla Perez<sup>2</sup>, Michelle A. Graham<sup>3</sup>, Peng Liu<sup>4</sup> and Maria G. Salas Fernandez<sup>2</sup>, (1)Department of Agronomy, Ames, IA, (2)Iowa State University, Ames, IA, (3)USDA, Ames, IA, (4)Department of Statistics, Iowa State University, Ames, IA

Brassinosteroids (BRs) are a group of steroid phytohormones that regulate several morphological, biological, and physiological processes during plant growth and development. One of the agronomically important morphological traits controlled by BRs is leaf angle, the inclination between the main stem and the leaf blade. Leaf angle is determined by the number and size of cells in the collar. In cereals, leaf angle affects plant density, light interception efficiency, photosynthetic rate, and thus, yield per unit of land. The proposed leaf angle ideotype, called a “smart canopy”, is one in which upper leaves are more erect and angles gradually increase towards the middle and lower canopy layers. However, preliminary evidence in sorghum demonstrated that leaf arrangement throughout the canopy is the opposite of a “smart canopy”.

In phenotyping experiments, the application of propiconazole (pcz), an inhibitor of BR biosynthesis, significantly reduced leaf angle. To discover genes and gene networks involved in the response of sorghum to pcz, RNA sequencing was conducted using two contrasting genotypes (PI 656015-large leaf angle, and BTx623-small leaf angle). The comparison between mock and pcz-treated samples revealed a greater number of differentially expressed genes for PI 656015 compared to BTx623. Genes that responded to pcz treatment were involved in phytohormone biosynthesis, signaling or response (mainly auxin, brassinosteroids, jasmonic acid, and abscisic acid), lipid metabolism, and cell wall modification. Future studies can utilize selected candidate genes as targets for leaf angle modification in sorghum and other cereals, to achieve the “smart canopy” ideotype and increase crop productivity.

#### **W588: Genomics of Tissue Regeneration in Plants and Animals**

##### **Myocardial Regeneration: Uncommon Sense for Common Problems**

**Mark Sussman**, SDSU Heart Institute,, San Diego, CA

Myocardial regenerative research remains an area of intensive study despite over a decade of frustratingly slow progress and modest clinical efficacy. A fundamental limitation in myocardial regeneration is inherently poor reparative capacity of adult mammalian heart which declines over lifespan. Augmentation of repair requires unnatural solutions to overcome normal adult myocardial biology using Regeneration Associated Cellular Effectors (RACE) to deliver functionally competent therapeutic interventions. The logic and rationale of four distinct RACE conceptual strategies will be presented including CardioEnhancers (genetic engineering), CardioChimeras (cell chimerism), CardioClusters (multi-cell three dimensional clustering), and CardioEvolvers (increased ploidy). Each RACE approach addresses a distinct biological limitation that impairs current cell-based treatments for myocardial damage, and different RACE approaches can be combined to promote synergism of biological potentiation. These next-generation approaches represent the future of myocardial

regenerative research, ultimately translating into novel clinical treatments achieving desperately needed treatment of heart failure.

## **W589: Genomics of Tissue Regeneration in Plants and Animals**

### **Documenting the Changes in DNA Methylation following Tissue Culture in Maize**

**Nathan M. Springer**<sup>1</sup>, Peter Crisp<sup>2</sup>, Sarah N. Anderson<sup>2</sup>, Scott C Stelpflug<sup>3</sup> and Shawn Kaeppler<sup>4</sup>, (1)University of Minnesota, St. Paul, MN, (2)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (3)University of Wisconsin-Madison, Madison, WI, (4)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI

DNA methylation plays roles in genome stability and regulation of expression. In general, DNA methylation exhibits faithful inheritance following cell division. However, there is substantial interest in understanding how DNA methylation may vary as a result of environmental changes. Tissue culture is one of the few conditions that has been demonstrated to effect DNA methylation patterns. We have monitored DNA methylation in a set of maize plants that have been subjected to tissue culture to document changes in DNA methylation. While the DNA methylation levels and general patterns of DNA methylation are similar following tissue culture there are locus-specific changes. A subset of these changes are common to multiple different independent tissue culture events suggesting consistent changes in DNA methylation can be triggered by tissue culture. In general, many of the loci that exhibit altered DNA methylation exhibit Mendelian inheritance of the DNA methylation levels suggesting that the changes can be stably inherited following culture. We have also assessed the changes in expression of transposable elements during tissue culture to understand the potential consequences of the changes in DNA methylation.

## **W590: Genomics of Tissue Regeneration in Plants and Animals**

### **Regeneration of Sensory Neurons in Planarians**

**Ricardo M. Zayas**, San Diego State University, San Diego, CA

Regeneration of missing body parts has long fascinated biologists, yet the mechanisms underlying regenerative processes remain poorly understood. My laboratory uses the freshwater planarian, a classic model of regeneration studies, to investigate molecular mechanisms underlying tissue regeneration. Planarians are able to form lost body parts and regenerate entire worms from very small body pieces. These organisms are endowed with a population of adult pluripotent stem cells that support their capacity for regeneration, which provides an excellent opportunity to identify genes involved in stem cell maintenance, proliferation, and differentiation. Work from several laboratories has demonstrated that tissue-specific transcription factors are expressed in subsets of planarian stem cells and their progeny, and that many of these factors play essential roles in neuronal cell fate specification and differentiation. My laboratory discovered that the SoxB1 gene, *soxB1-2*, is required for regeneration and maintenance of epidermal and sensory neuron populations in the planarian *Schmidtea mediterranea*. Our initial work uncovered a candidate set of *soxB1-2*-regulated target genes required for mechanosensory regeneration, but the mechanism underlying *soxB1-2*<sup>+</sup> stem cell specification and terminal differentiation remains largely unknown. In my talk, I will discuss our on-going work aimed at understanding how *soxB1-2* regulates sensory neuron regeneration and function in planarians.

## **W591: Genomics of Tissue Regeneration in Plants and Animals**

### **Signaling for Regeneration**

**Moshe Reuveni**, ARO, Volcani Center, Bet Dagan, Israel

The regeneration of shoots from leaf tissue is investigated and shown to occur from multiple cells and not a single cell. The regeneration process requires cellular contact and second messengers transfer. Embryogenesis is the process whereby a cell divides into cellular cooperatives that are shaped into differentiated tissues and organs. Organogenesis via plant tissue culture is thought to be a similar

process that starts from a de-differentiating single cell. Here we show that in tobacco plants that undergo shoot regeneration from leaf tissue organogenesis occurs via cellular collectives. These cellular collectives are dependent on communication between the cells prior to the collective decision to go into shoot formation.

## **W592: Graft Genetics and Genomics**

### **Scion-to-Rootstock Transfer of Endogenous Small RNAs in Sweet Cherry**

**Guo-Qing Song**, Michigan State University, East Lansing, MI and **Dongyan Zhao**, Michigan State University

Cherry grafting is a well-established agricultural practice in cherry production for clonal propagation, disease/pest-resistance, altered plant vigor and architecture, increased tolerance to abiotic stresses, precocity, and higher yield. Mobile molecules (e.g., water, hormones, nutrients, DNA, RNAs, and proteins) play an essential role in rootstock-scion interactions. Small RNAs (sRNAs) are short (~19 to 30 nucleotides) RNA molecules that are a group of mobile signals in plants. Rootstock-to-scion transfer of transgene-derived small interfering RNAs enables virus resistance in nontransgenic sweet cherry scion. To determine whether there are long-distance scion-to-rootstock transfer, we compared sRNAs profiles in bud tissues of a non-grafted cherry rootstock 'Gisela 6', a sweet cherry 'Emperor Francis' scion as well as its 'Gisela 6' rootstock. We identified 1-29 scion-to-rootstock transfer of sRNAs from 574 cherry transcripts, of which 296 was annotated using the reference gene set of *Arabidopsis thaliana*. Of the transported 19- to 30-nucleotide (nt) sRNAs, there are three major groups, including 38.0% of 24-nt, 18.9% of 21-nt, and 10.2% of 22-nt sRNAs. Furthermore, the transported sRNAs to rootstock (sink) showed a positive correlation with those in the scion (source). Overall, the profiles of the transported sRNAs and their annotations revealed in this study facilitate a better understanding of the role of the long-distance transported sRNAs in sweet cherry rootstock-scion interactions.

## **W593: Graft Genetics and Genomics**

### **Rootstock by Scion Interaction in Polyploid Shrub Willow (*Salix*)**

**Dustin G. Wilkerson**<sup>1</sup>, **Laura E. Dougherty**<sup>2</sup>, **Chris O. Hernandez**<sup>3</sup> and **Lawrence B. Smart**<sup>1</sup>,  
(1)Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY,  
(2)Horticulture Section, School of Integrative Plant Science, Cornell AgriTech, Cornell University, Geneva, NY, (3)Plant Breeding and Genetics Section, School of Integrative Plant Sciences, Cornell University, Ithaca, NY

Advancements in biomass production of shrub willow (*Salix* spp.) breeding programs have focused on the capture of heterosis through clonal propagation in interspecific hybrids. Especially in triploids formed from the crossing of diploid and tetraploid species, there is improvement in yield, vigor and biomass composition that recent research suggests is due to an increased number of differentially expressed genes in the hybrid. Using diploid *S. viminalis*, tetraploid *S. miyabeana*, and one of their triploid progeny selected due to its high mid-parent heterosis, this study sought to determine if the increased yield and vigor of the triploid could be conveyed across the graft junction and influence its grafted counterpart. The triad was grafted in all combinations including self and ungrafted checks using dormant cuttings and planted in a greenhouse as an RCBD with six replications for eight weeks. A myriad of phenotypes were collected including daily bud break at the onset of the experiment, weekly height measurements, and during the final week of the experiment, leaf area, SPAD, stomatal conductance, stem number and diameter, cumulative leaf weight, and total above and below ground biomass. Additionally, using four of the six replications, leaf samples were sent for ionomic analysis and shoot tip and root tissue were submitted for 3'RNAseq.

## **W594: Graft Genetics and Genomics**

### **Rootstocks As an Exciting Research Frontier in Blueberry Science**

**Wei Qiang Yang**, Oregon State University, Aurora, OR

Rootstocks have been used successfully in fruit production for decades. Its applications in canopy management, drought tolerance, disease resistance, yield management, and harvest timing are widely studied in many fruit and vegetable cropping systems. There is currently no commercial blueberry rootstock available. Our blueberry rootstock research in the last ten years has brought us closer to the first commercial release of blueberry rootstocks that will grow well in relatively high pH soils. Most importantly, the availability of rootstocks will change blueberry plant from a bushy growing habit to a tree form. Our studies demonstrated rootstock and cultivar interactions in terms of compatibility, yield, soil pH adaptation, nutritional requirements, and root system function. The basic nutrient needs and growing habits in a nursery setting were also studied with using clonal blueberry rootstocks. As an emerging research area, much remains to be learned from exploring a blueberry tree based production system.

## **W595: Graft Genetics and Genomics**

### **Understanding Scion/Rootstock Interactions in Grapevine**

**Nathalie Ollat**<sup>1</sup>, Duyen Prodhomme<sup>1</sup>, Clément Chambaud<sup>1</sup>, Cyril Hévin<sup>1</sup>, Josep Valls Fonayet<sup>2</sup>, Ghislaine Hilbert<sup>1</sup>, Claudine Trossat-Magnin<sup>1</sup>, Tristan Richard<sup>2</sup>, Giovanni Bortolami<sup>3</sup>, Gregory A. Gambetta<sup>4</sup>, Lysiane Brocard<sup>5</sup> and Sarah Jane Cookson<sup>1</sup>, (1)EGFV, Bordeaux Sciences Agro - INRA - University of Bordeaux, ISVV, Villenave d'Ornon, France, (2)Faculté des Sciences Pharmaceutiques, Unité de Recherche Œnologie EA 4577, USC 1366 INRA, Equipe Molécules d'Intérêt Biologique (Gesvab) - ISVV, University of Bordeaux, Villenave d'Ornon, France, (3)INRA, UMR 1065 Santé et Agroécologie du Vignoble, Bordeaux Sciences Agro, Villenave d'Ornon, France, (4)EGFV, Bordeaux Sciences Agro - INRA - University of Bordeaux, ISVV, Villenave d'Ornon, France, (5)Bordeaux Imaging Centre, Plant Imaging Platform, UMS 3420, INRA-CNRS-INSERM, University of Bordeaux, Villenave d'Ornon, France

Grafting is used in viticulture to allow grapevine cultivation in soils infected with Phylloxera, an American soil-dwelling insect pest introduced to Europe at the end of the 19th century. Graft union formation is a complex biochemical and structural process that begins with a wound response, then callus formation and the establishment of a functional xylem and phloem connecting the two grafted partners. Despite the importance of the scion/rootstock interface in horticulture, we know little of the processes involved in forming a successful graft union.

We have studied the developments at the scion/rootstock interface of grapevine using a variety of techniques. Morphological developments have been studied using microscopy techniques and high resolution computed tomography. Xylem connectivity has been assessed using a high pressure flow meter. Microarrays have been used to identify the genes differentially expressed between the wood and graft interface tissues of homo-grafts (the same genotype grafted together) and at the graft interface between different scion/rootstock combinations (hetero-grafts). Primary and secondary metabolite profiling has been done as well as the quantification of hormone concentrations.

An overview of the interdisciplinary approaches currently being used to piece together the puzzle of graft union formation in an important woody, perennial crop will be presented.

## **W596: Graft Genetics and Genomics**

### **The Long-Distance Signaling and Tree Maturity - microRNA Perspective**

**Muhammad Umair Ahsan**<sup>1</sup>, Alice Hayward<sup>2</sup>, Neena Mitter<sup>2</sup> and Margaret Frank<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)The University of Queensland, St Lucia, Australia

Grafting is the common propagation method for avocado and primarily benefits orchard production by reducing the time to tree productivity. It also allows use of scions and rootstocks specifically selected for improved productivity and commercial acceptance. Rootstocks in avocado may be propagated from mature tree cuttings ('mature'), or from seed ('juvenile'). While the use of mature scion material hastens

early bearing/maturity and economic return, the molecular factors involved in the role of the scion and/or rootstock in early bearing/reduced juvenility of the grafted tree are still unknown. To explore this, we utilized juvenility and flowering associated miRNAs; miR156 and miR172 and their putative target genes to screen pre-graft and post-graft material in different combinations from avocado. The abundance of mature miR156, miR172 and the miR156 target gene *SPL4*, showed a strong correlation to the maturity of the scion and rootstock material in avocado. Graft transmissibility of miR156 and miR172 has been explored in annual plants. Here, we show that the scion may be responsible for grafted tree maturity involving these factors, while the rootstock maturity does not significantly influence miRNA abundance in the scion. We also demonstrate that the presence of leaves on cutting rootstocks supports graft success and contributes towards intergraft signalling involving the carbohydrate-marker *TPS1*. In conclusion, we suggest that the scion largely controls the molecular 'maturity' of grafted avocado trees, however, leaves on the rootstock not only promote graft success, but can influence miRNA and mRNA abundance in the scion. This constitutes the first study on scion and rootstock contribution towards grafted tree maturity using the miR156-*SPL4*-miR172 regulatory module as a marker for juvenility and reproductive competence.

## **W597: Graft Genetics and Genomics**

### **Effects of in-Row Spacing on Grafted Watermelon Productivity and Fruit Quality**

**Zheng Wang**, University of California Cooperative Extension, Modesto, CA

Vegetable grafting has its merits, but the increased cost and the additional needs of field management remain the major limit to widespread adoption. Practices which maintain the advantages of grafting while making it more cost-effective must be evaluated under commercial field conditions. To answer the question, the team at University of California Cooperative Extension used watermelon as the study commodity to evaluate the impacts of different rootstock-scion combinations and in-row transplant spacing on yield and fruit quality. The field trial was initiated in the spring 2019 in Turlock, California on a commercial watermelon field. The study was a split-split plot design with 3 in-row spacing (93 cm <3 feet>, 123 cm <4 feet>, and 183 cm <6 feet>), two scions (7187 and Fascination), and 4 rootstocks (RS841, Flexifort, UG29A, and XSQ9901) as main, sub-, and sub-sub-factors, respectively. Non-grafted scion plants were also included as the control. The entire set of treatments were replicated four times. Each row (plot) is 9.3 m long containing 10, 7, and 5 plants for 93 cm, 123 cm, and 183 cm spacing, respectively. Wildcard Plus was used as the pollenizer, and each pollenizer plant was placed after every three triploid plants. The fertility and field management followed grower's standard practices. Fruit were harvested five times and sampled for quality and yield measurements. Data will be presented on the workshop.

## **W598: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**

### **Introduction to Gramene Workshop**

**Doreen Ware**, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Gramene (<http://www.gramene.org>) is an integrated reference and curated data resource for comparative functional genomics in plants. It supports a discovery environment with well-annotated genomic data sets and tools to enable plant researchers, breeders and educators to undertake analysis and visualization of reference plant genomes and pathways and make data-driven hypotheses for experimental analysis. Currently, our genome portal hosts 67 reference plant genomes ranging from unicellular autotrophs to higher plants, annotations for 2,146,774 genes, 96,607 gene families, and genetic diversity datasets from 12 plant species in the form of ~224 million SNPs. The Plant Reactome pathway portal (<https://plantreactome.gramene.org>) hosts 306 curated reference rice pathways and their gene homology-based projection to 97 species. The Expression Atlas provides transcriptomic data from 819 experiments in 24 plant species. Our integrated search platform facilitates querying across these various data portals. Tools are available to users to visualize genome alignments, synteny, gene features and neighborhoods, genetic variation and their consequences on gene function and structure, and gene expression profiles. These data sets support users in exploring nuances of speciation, ploidy, adaptation, effects of domestication and natural selection on the genome structure, gene function, plant structure,

phenotype, pathways and plant metabolism. Gramene is committed to open access and FAIR data principles. We support data access in various standardized formats as well as via web-based programming interfaces (APIs). Extensive use of ontologies, database cross-references, common data formats, metadata, community engagement and open-source software promotes interoperability within the ecosystem of informatics data and services. Gramene is supported by NSF grant IOS-1127112, with partial support from USDA-ARS (1907-21000-030-00D).

### **W599: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities** **Single Cell RNA-Sequencing in Plants: EMBL-EBI's Resources for Data Submission, Archival and Visualization**

**Nancy George**, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Single-cell sequencing (scRNA-seq) is one of the latest development in the field of genomics. With the ability to detect gene expression at the level of single cells, researchers can use this technology to identify rare cell populations, investigate complex cell-cell interactions such as those between parasites and hosts or map localised cell response to environmental stimuli such as drought. Although such technology is exceptionally useful, the number of scRNA-seq plant datasets remains relatively limited due to the difficulty of isolating single plant cells. Thus, those single-cell datasets that have been generated are of high value to the plant scientific community and their reuse and reproducibility of particular interest.

The Gene Expression group at EMBL-EBI have developed a number of tools that facilitate the submission, archival and visualisation of functional genomics data with particular focus on plant and single cell sequencing datasets. In collaboration with Gramene, the functional genomics web submission tool, Annotare, was updated with plant material templates designed to ensure accurate capture of plant sample metadata. Information required for reproducibility of single-cell sequencing datasets were used as a basis to design single-cell submission templates and extensive guidelines on reporting scRNA-seq studies are published as a preprint here: <https://arxiv.org/abs/1910.14623>. Upon successful submission datasets including raw data are made available in the ArrayExpress archive under a stable, citable accession.

Publicly available scRNA-seq datasets from ArrayExpress and other public archives are systematically reanalysed and visualised in our added-value resource Single Cell Expression Atlas (SCEA) ([www.ebi.ac.uk/gxa/sc](http://www.ebi.ac.uk/gxa/sc)). As of December 2019, SCEA contains 5 landmark *Arabidopsis thaliana* single-cell datasets as part of 132 datasets, across 12 species and over 1.3 million cells. Through this platform users can search for genes across these datasets and filter the results for particular cell types or tissues. SCEA can also be used to identify in what conditions and populations a gene can act as a marker gene, i.e. define a specific cell population. For each experiment cell populations are displayed via a t-SNE plot with clusters defined using the *Scanpy* clustering algorithm. Cells can also be coloured over with the underlying metadata or post analysis sequencing cell identity (inferred cell type) where available. Gene expression at the single-cell level can be explored in the neighbouring plot. The top 5 marker genes per cluster are displayed in a heatmap and all analysis data and accompanying metadata are available for download.

### **W600: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities** **NGS-Enabled Fast Gene Discovery from Sorghum Pedigreed Mutant Library**

**Zhanguo Xin**, USDA ARS, Lubbock, TX

CRISPR-cas9-derived genome editing tools are rapidly evolving technologies that will revolutionize medicine and plant breeding. However, known promising targets are required for effective application of this exciting technology. Chemical, such as ethyl methane sulfonate, induced mutagenesis is a powerful tool to generate new variations including those that have potential to improve breeding. We have developed a pedigreed mutant library in sorghum inbred line BTx623, which is used to generate

reference genome sequence. The mutant library harbors a wide range of phenotypes, many of which may have potential to improve sorghum. To rapidly identify the causal mutations for these altered phenotypes, we have developed a bioinformatic pipeline based on bulk segregant analysis (BSA) of pooled homozygous mutants selected from segregating F<sub>2</sub> populations. In the last few years, we have discovered over 30 causal mutations underlying important agronomy traits. However, the presence of large number of unrelated background mutations may hamper their direct uses in breeding. It may take years to remove the unlinked background mutations by genetic backcrosses and even longer for the linked background mutations. With the precision and low off-target mutations, CRISPR-cas9 genome editing tools can be used to re-create the causal mutations in elite lines to accelerate breeding. A combination of fast causal gene discovery from the mutant library with the precision genomic editing will truly revolutionize plant breeding.

### **W601: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**

#### **Effect of Sequence Depth and Length in Long-Read Assembly of the Maize Inbred NC358**

**Shujun Ou**, Department of Ecology, Evolution, and Organismal Biology (EEOB), Iowa State University, Ames, IA

Recent improvements in the quality and yield of long-read data and scaffolding technology have made it possible to rapidly generate reference-quality assemblies for complex genomes. Still, generating these assemblies is costly, and an assessment of critical sequence depth and read length to obtain high-quality assemblies is important for allocating limited resources. To this end, we have generated eight independent assemblies for the complex genome of the maize inbred line NC358 using PacBio datasets ranging from 20-75x genomic depth and N50 read lengths of 11-21 kb. Assemblies with 30x or less depth and N50 read length of 11 kb were highly fragmented, with even the low-copy genic fraction of the genome showing degradation at 20x depth. Distinct sequence-quality thresholds were observed for complete assembly of genes, transposable elements, and highly repetitive genomic features such as telomeres, heterochromatic knobs and centromeres. This study provides a useful resource allocation reference to the community as long-read technologies continue to mature.

### **W602: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**

#### **Investigating the Maize Genome with Undergraduate Researchers**

**Rebecca L. Seipelt-Thiemann**, Middle Tennessee State University, Murfreesboro, TN

Undergraduate research is known to provide students with valuable experience and impact their view of themselves as scientists, as well increase persistence and success in science. However, time, space, and access to these experiences is limited. Course undergraduate research projects are an excellent way to involve more students in this practice of science in a more equitable way. With this in mind, a course research project involving undergraduate genetics students was developed in which students worked to investigate genome annotation resources and work on annotating select genes of the most recent Maize genome (V5). Students investigated sets of genes involved in specific functional categories using Plant Reactome and primary literature, identified transcripts and gene models from reference sequence resources using Gramene, used RNA evidence in the maize Apollo annotator to construct gene models accommodating all evidence, re-constructed isoforms in a region of interest for a single gene, identified potential domain structures in the region of interest using Gramene, designed primers for RT-PCR to detect isoform/gene model differences using Primer3Plus, performed RNA extract on plant tissue, produced cDNA, performed RT-PCR and agarose gel electrophoresis, compared expected and observed RT-PCR products to make inferences about isoform diversity and functional differences among three plant tissues as well as provide evidence to guide further manual curation of the gene models. Students presented this work in a scientific poster format at the end of the course. The workflow, challenges, and successes of this project will be discussed.

### **W603: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**

#### **Accessing the Maizecode Data from Gramene and Sciapps**

**Liya Wang**, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

MaizeCODE is a project aimed at identifying and analyzing functional elements in the maize genome. In its initial phase, MaizeCODE assayed up to five tissues from four maize strains (B73, NC350, W22, TIL11) by RNA-Seq, Chip-Seq, RAMPAGE, and small RNA sequencing. To facilitate reproducible science and provide both human and machine access to the MaizeCODE data, we developed SciApps, a cloud-based portal, for analysis and distribution of both raw data and analysis results. Based on the SciApps workflow platform, we generated new components to support the complete cycle of the MaizeCODE data management. These include publicly accessible scientific workflows for reproducible and shareable analysis of various functional data, a RESTful API for batch processing and distribution of data and metadata, a searchable data page that lists each MaizeCODE experiment as a reproducible workflow, and integrated JBrowse genome browser tracks linked with workflows and metadata. MaizeCODE data are also integrated into the Gramene platform, so that users can load the data into Gramene's genome browser, examine the associated metadata, and relaunch the reproducible workflows.

#### **W604: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities Gene Regulatory Network Construction Based on Machine Learning Enables Phylogenetic Comparisons in the Monocots**

**Andrea Braeutigam**, Bielefeld University, Bielefeld, Germany

Plant development in space and time and plant's responses to environmental cues are at least in part governed by gene regulatory networks (GRNs). At least for some species, sufficient amounts of publicly available RNA-seq data in public databases enables GRN prediction. We explored machine learning as a possible route towards establishing transcription factor (TF) target gene relationships. We hypothesized that evolutionary comparisons between different monocotyledonous species will reveal conserved GRNs.

Machine learning, more precisely random forest decision tree based regression, successfully constructs GRNs for *Zea mays*, *Triticum aestivum*, *Hordeum vulgare*, and *Oryza sativa* ssp. *Japonica* while data paucity prevents construction in *Sorghum bicolor*. We detect high false positive rates in the networks and overcome this limitation by introducing enrichment analyses. Phylogenetic analyses of TFs and targets reveal functionally conserved and diverged branches of TF family trees among the monocots. GRNs also enable TF binding site analyses of transcription factors solely based on publicly available ATAC-seq and DNase-seq datasets in conjunction with the regulatory networks.

#### **W605: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities Plant Reactome 2020: A Plant Pathway Knowledgebase**

**Justin Preece**<sup>1</sup>, Sushma Naithani<sup>2</sup>, Parul Gupta<sup>2</sup>, Peter D'Eustachio<sup>3</sup>, Justin L. Elser<sup>2</sup> and Pankaj Jaiswal<sup>1</sup>, (1)Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, OR, (2)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, (3)NYU School of Medicine, New York, NY

Plant Reactome (<https://plantreactome.gramene.org>) is an open-source, comparative plant pathway knowledgebase of the Gramene project. It uses *Oryza sativa* (rice) as a reference species for manual curation of pathways and extends pathway knowledge to another 97 plant species ranging from unicellular autotrophs to higher plants via gene-orthology projections. It hosts 306 reference pathways, including metabolic and transport pathways, transcriptional networks, hormone signaling pathways, and plant developmental processes. Plant Reactome supports (i) quick search for gene/protein, metabolites, pathways and browsing of pathways; (ii) upload and analysis of gene-expression and gene-gene interaction data in the context of plant pathways; and comparisons of reference rice pathways with pathways from any of 97 projected species hosted by Plant Reactome. Our curators engage researchers on gene and pathway curation by offering workshops and online tutorials and mentor undergraduate students on research projects focused on gene/pathway data and biocuration. The Plant Reactome

supports, implements and collaborates with the wider community to make data and tools related to genes, genomes, and pathways Findable, Accessible, Interoperable and Re-usable (FAIR).

## **W606: Grape Genome Initiative**

### **The Wild Side of Grape Genomics**

**Dario Cantu**<sup>1</sup>, Melanie Massonnet<sup>1</sup>, Noe Cochetel<sup>1</sup>, Andrea Minio<sup>1</sup>, Yongfeng Zhou<sup>2</sup>, Aline Marie Muyle<sup>2</sup>, Abraham Moralez-Cruz<sup>2</sup>, Jerry Lin<sup>1</sup>, Jadran Garcia<sup>1</sup>, Amanda Vondras<sup>1</sup>, Rosa Figueroa-Balderas<sup>1</sup> and Brandon S. Gaut<sup>2</sup>, (1)University of California Davis, Davis, CA, (2)University of California Irvine, Irvine, CA

The cultivation, sustainability and security of grapes (*Vitis vinifera*) rely on wild *Vitis* species as sources of resistance to biotic and abiotic stresses. Despite the importance of non-vinifera *Vitis* species, very few genomic resources are available. We have been generating reference genomes for wild *Vitis* species that either have been used, or have demonstrated promise, for breeding. These include multiple accessions of *V. vinifera* ssp. *sylvestris*, as well as North American species such as *V. arizonica*, *V. girdiana*, *V. berlandieri*, *V. acerifolia*, *V. riparia*, *V. aestivalis*, *V. monticola*, *V. mustangensis*, and Central Asian species, such as *V. piazeskii* and *V. romanetii*. All genomes were sequenced using single molecule real-time sequencing (SMRT; Pacific Biosciences) and optical maps (Bionano), and are being assembled into completely phased pseudochromosomes. SMRT sequencing was also used to sequence full-length cDNA (Iso-Seq) and, together with high-depth short-read libraries, reconstruct the transcriptomes of all species. These annotated reference genomes will be shared with dedicated genome browsers as a community resource and have been used as a foundation for own studies. For example, living collections of thousands of North American genotypes have been genotyped and phenotyped to study wild grape evolution in the American Southwest as well as to identify novel sources of genetic resistance to Pierce's Disease resistance and tolerance to salinity. These resources have been valuable also to study the genetic basis of other important domestication and agronomic traits.

## **W607: Grape Genome Initiative**

### **Role of microRNAs in Mediating the Effect of Leafroll Virus on Fruit Ripening in Table Grapes**

**Michael J Considine**, University of Western Australia, Perth, WA, Australia

Infection by grapevine leafroll associated virus can significantly retard ripening in table grapes. This study exploited a set of clones of cv Crimson Seedless, which were established by infection with either GLRaV-3 or a mixture of GLRaV-3, GLRaV-9 and GVA. These show significant differences in quality and ripening. We report sequencing of small RNA of six replicate plants of each clone. Enrichment analysis showed cellular developmental and cellular differentiation processes were represented commonly across the infected and uninfected clones. We show differential effects of virus infection on classes and families of microRNAs; both qualitative and quantitative differences were seen. Detailed analysis of the putative grapevine and viral targets will be reported.

## **W608: Grape Genome Initiative**

### **Population Genomics of Structural Variants in Grapevine Domestication**

**Yongfeng Zhou** and Brandon S. Gaut, University of California Irvine, Irvine, CA

Structural variants (SVs) are a largely unexplored feature of plant genomes. Understanding these dynamics is critical for understanding both the contributions of SVs to phenotypes and the likelihood of identifying them as causal genetic variants in genome-wide associations. We identify SVs and study their evolutionary genomics in clonally propagated grapevine cultivars and their outcrossing wild progenitors. To catalogue SVs, we assembled the highly heterozygous Chardonnay genome, for which one in seven genes is hemizygous based on SVs. Using an integrative comparison between Chardonnay and Cabernet Sauvignon genomes by whole-genome, long-read and short-read alignment, we extended SV detection to population samples. We found that strong purifying selection acts against SVs but

particularly against inversion and translocation events. SVs nonetheless accrue as recessive heterozygotes in clonally propagated lineages. They also define outlier regions of genomic divergence between wild and cultivated grapevines, suggesting roles in domestication. Outlier regions include the sex-determination region and the berry colour locus, where independent large, complex inversions have driven convergent phenotypic evolution.

## **W609: Grape Genome Initiative**

### **Molecular Characterization of the Sex Loci in Wild and Domesticated Grapes**

**Cheng Zou**<sup>1</sup>, Fred E. Gouker<sup>2</sup>, Avinash Karn<sup>2</sup>, Anne Fennell<sup>3</sup>, Dario Cantu<sup>4</sup>, Xia Xu<sup>5</sup>, Matthew D. Clark<sup>6</sup>, Bruce Reisch<sup>7</sup>, Qi Sun<sup>1</sup>, Lance Cadle-Davidson<sup>8</sup> and Jason Londo<sup>8</sup>, (1)Cornell University, Ithaca, NY, (2)Cornell University, Geneva, NY, (3)South Dakota State University, Brookings, SD, (4)University of California Davis, Davis, CA, (5)ARS-Grape Genetic Research Unit, Geneva, NY, (6)University of Minnesota, St. Paul, MN, (7)Horticulture Section, School of Integrative Plant Science, Cornell AgriTech, Cornell University, Geneva, NY, (8)USDA-ARS Grape Genetics Research Unit, Geneva, NY

Hermaphroditic (perfect) flowers were a key trait in grapevine domestication, enabling a drastic increase in yields due to the efficiency of self-pollination in the domesticated grapevine (*Vitis vinifera* L. ssp. *vinifera*). In contrast, all extant wild *Vitis* species are dioecious (each plant having only male or female flowers), with known genetic dominance of male (M) > hermaphrodite (H) > female (f). While the genetic region underlying flower sex has been delineated to several candidate genes in the European grapevine, this locus has not been explored in North American and Asian wild *Vitis*, species for which markers from European accessions fail to predict flower sex. In this study, we de novo assembled *Vitis cinerea* 'B9', a male accession of a North American wild grapevine. We identified the male(M) and female(f) chromosome haplotypes by the bulk sample resequencing of 13 male and 13 female accessions, observing the expected heterozygosity (Mf) in the male individuals and homozygosity (ff) in the female individuals. A comparative genomic analysis from shotgun resequencing data of wild and domesticated accessions suggested the H chromosome region is evolved from M. A statistical model based on these data successfully predicted flower sex in 96.1 % of a diversity panel spanning 200 wild and hermaphrodite accessions from all three gene pools. We also hypothesize a molecular model for sex determination in *Vitis*, in which the female sterile is attributed to a homozygous deleterious mutation, while the male sterile is attributed to a male-specific over-expression. This mechanism is supported in all *Euvitis* germplasm (from North America, Europe, Asia and domesticated grapevine), and could explain the observed genetic dominance of M> H> f.

## **W610: Grape Genome Initiative**

### **Elucidating the Clonal Diversity of *Vitis vinifera* cv. Cabernet Sauvignon**

Claudio Urra<sup>1</sup>, Catalina Pavez-Reyes<sup>2</sup>, Andrea Minio<sup>3</sup>, Matthieu Miossec<sup>4</sup>, Cristobal Balladares<sup>1</sup>, Dayan Sanhueza<sup>5</sup>, Alvaro Castro<sup>2</sup>, Dario Cantu<sup>3</sup> and **Claudio Meneses**<sup>1</sup>, (1)Universidad Andrés Bello, Centro de Biotecnología Vegetal, Santiago, Santiago, Chile, (2)UC Davis-Chile, Life Sciences Innovation Center, Chile, (3)University of California Davis, Davis, CA, (4)Universidad Andres Bello, Center for Bioinformatics and Integrative Biology, Santiago de Chile, Chile, (5)Universidad Andrés Bello, Centro de Biotecnología Vegetal, Santiago, Chile

Cabernet Sauvignon (CS) is the most iconic variety for the Chilean viticulture industry. Numerous selections or clones of CS are available, which have significant differences in agronomic and enological traits. These differences have been originated from the accumulation of somatic mutations during thousands of asexual propagation cycles. However, the genetic variation that underlies these differences remains mostly unknown. Our goal was to characterize the clonal diversity in *Vitis vinifera* cv. Cabernet Sauvignon. Thus, we re-sequenced eight CS clones using the Illumina HiSeq2500 sequencer. We obtained a mean coverage of 30X with a 98% of mapped reads in average and detected by GATK an average of 3,581,252 variants (83.4% SNPs and 16.6% InDels) among CS clones compared to the reference genome. The variant frequency and distribution were homogeneous detecting 5.3 SNP/Kbp

and 1.1 InDel/Kbp for all analyzed clones. Considering the global nucleotide diversity, three clones showed significant differences, while five seem to be the same genome. We found an average of 1,395 unique variants from CS clones with a mean of 1.66% located in coding regions. We observed that most of the total variants were located on intergenic repetitive regions while the unique variants were located mainly on non-repetitive elements. We validated the most promising clone-specific variants by amplicon sequencing. Besides, using a similar approach, we detected clone-specific markers for 'Sauvignon blanc', 'Chardonnay', 'Merlot' and 'Pinot noir'. The results will be used to develop a high-throughput genotyping platform for clonal identification. This work was supported by FONDECYT 1160584 and CORFO 13CEI2-21852.

## **W611: Grape Genome Initiative**

### **Influence of Seed Development of the Transcriptional Program of Grape Berries during the Ripening Phase**

Satyanarayana Gouthu<sup>1</sup>, Landry Rosssdeutsch<sup>1</sup>, Joseph Schmidt<sup>2</sup>, Yu Wang<sup>3</sup>, Albert Batushansky<sup>4</sup>, Sara Zenoni<sup>5</sup>, Mario Pezzotti<sup>5</sup>, Aaron Fait<sup>6</sup> and **Laurent G. Deluc**<sup>7</sup>, (1)Oregon State University, (2)North American Plants, (3)OSU, (4)OMRF, (5)University of Verona, Verona, Italy, (6)Ben-Gurion University of the Negev, Blaustein Insts for Desert Research, Sde Boqer campus, Israel, (7)Dept. of Horticulture, Oregon State University, Corvallis, OR

The ripening initiation is a transitional developmental phase characterized by a complete reprogramming that leads to dramatic changes on the central and secondary metabolisms of the fruit, and that contributes to the final fruit composition in terms of flavor and aromas. In grape berry research, much scientific efforts have emphasized on understanding how environmental changes (macro, meso, and microclimate), and plant statuses are integrated to the fruit development. The influence of developmental cues such as seed and embryo development has yet to be characterized on the ripening program. One peculiar feature of a grape berry cluster, conserved across cultivated and wild species, is the uneven pattern of the developmental progression of individual fruits within a cluster. In our lab, we found a poor correlation between anthesis and the ripening status of individual berries at the ripening initiation; rather the seed content relative to the berry mass (**Seed Index**) showed a stronger correlation in seeded berries. Seeds are seen as a main reservoir for plant hormones dedicated to the embryo growth and differentiation embryo development. Whether the dynamics of hormone accumulation in the seed follow a similar trend during the ripening onset in the other tissues of the pericarp is unclear. We hypothesized that seeds contribute to the timing of ripening of individual berries through a gradual decline a seed-derived transport of auxin, a ripening delayer, near the ripening onset. We used a streamlined phenological procedure to segregate two populations of berries with distinct Seed Index at different developmental stages (pre- and post-ripening, and maturity stages). Combining hormone and metabolite profiles with genome-wide RNA sequencing analyses, we were able to identify clear transcript, hormone and metabolite signatures between the two classes initiated at a discrete pre-ripening stage, with the maximum differences observed around the ripening onset, before experiencing a developmental convergence at transcript and metabolite levels towards maturity stage. Such ability of individual fruit to first diverge and then converge can only be explained by a modification of their rate of ripening. The relationships between transcript, hormone and metabolite signatures overtime in both classes and difference in rate of ripening are discussed. Differences in transcript abundance associated Plant hormone signalings (ABA, Auxin, and brassinosteroids) and their transports seem to contribute to the developmental divergence and convergence. Research perspectives of these new findings to understand the timing of ripening initiation and the further developmental adjustment observed during the ripening are discussed.

## **W612: Grasslands (Lolium Genome Initiative)**

### **Mapping of Self-Compatibility for Hybrid Breeding in Perennial Ryegrass (*Lolium perenne* L.)**

**Claudio Cropano**<sup>1</sup>, Iain Place<sup>2</sup>, Chloé Manzanares<sup>1</sup>, Boas Pucker<sup>3</sup>, Javier Do Canto<sup>4</sup>, Steven Yates<sup>5</sup>, Dario Copetti<sup>1</sup>, Daniela Holtgräwe<sup>3</sup>, Nic Boerboom<sup>6</sup>, Lukas Wolter<sup>6</sup>, Michael Koch<sup>7</sup>, Bernd Weisshaar<sup>3</sup>,

Torben Asp<sup>8</sup>, Thomas Lübberstedt<sup>9</sup>, Daniel Thorogood<sup>2</sup> and Bruno Studer<sup>10</sup>, (1)Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zürich, Zurich, Switzerland, (2)Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, United Kingdom, (3)Department of Biology, Bielefeld University & CeBiTec, Germany, (4)Instituto Nacional de Investigación Agropecuaria, Uruguay, (5)Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zürich, Switzerland, (6)DSV Zaden B.V., Netherlands, (7)Deutsche Saatveredelung AG, Germany, (8)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, (9)Iowa State University, Ames, IA, (10)Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

Perennial ryegrass (*Lolium perenne* L.) is one of the most important forage crops worldwide. Its outcrossing nature is caused by self-incompatibility (SI), a genetic mechanism preventing self-pollination in most forage grasses. SI represents a constraint for F1 hybrid breeding, as it impedes the development of homozygous parental inbred lines by repeated selfings. The need to develop higher-yielding grass varieties is driving scientists and forage breeders to find ways to overcome SI. A possible solution is to exploit non-functional alleles at genes determining the initial self/non-self recognition or mutations on the downstream cascade genes, leading to self-compatibility (SC). In fact, SC has been found in different SI grass species, such as perennial ryegrass, but its genetic and molecular basis remain unclear. Here we report on the mapping of SC in two distinct bi-parental populations, each segregating for a unique SC locus. For each population, a different mapping strategy is used to identify the genomic region harbouring the SC locus. The first relies on a marker-trait association using genome-wide markers generated by genotyping-by-sequencing (GBS). Molecular markers associated to SC are then used in a classical fine-mapping experiment to identify candidate genes. The second strategy uses whole-genome sequencing of pools of individuals showing contrasting SC phenotypes. This strategy, often referred as bulked segregant analysis (BSA), allows to identify a region with candidate genes based on differential allele frequencies of the pools. Together, these approaches will contribute to unravelling the genetics of SC in perennial ryegrass. In addition, they lay the ground for an efficient development of inbred lines and their exploitation in F1 hybrid breeding schemes.

## **W613: Grasslands (Lolium Genome Initiative)**

### **Advanced Phenomics Integrated into Rapid Genomic Sub-Selection in Perennial Ryegrass**

Luke Pembleton<sup>1</sup>, Chaya Smith<sup>2,3</sup>, Michelle M. Malmberg<sup>1</sup>, Preeti Thakur<sup>1</sup>, Carly Elliot<sup>2</sup>, Michelle Drayton<sup>1</sup>, Maiko Shinozuka<sup>1</sup>, Pieter Badenhorst<sup>2</sup>, Kevin F. Smith<sup>4</sup>, German Spangenberg<sup>1,3</sup> and **Noel O.I. Cogan**<sup>1,3</sup>, (1)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, Australia, (2)Agriculture Victoria, Hamilton, Australia, (3)School of Applied Systems Biology, La Trobe University, Bundoora, VIC, Australia, (4)University of Melbourne, Hamilton, Australia

Although perennial ryegrass (*Lolium perenne* L.) is the most important temperate forage species and is recognized as having a relatively high level of nutritive value, its quality is often inadequate to meet the requirements of high-producing dairy cows. During seasonal production times, supplements are often necessary to address feed and nutrition deficits. Nutritive traits are known to be under both genetic and environmental control and the costs associated with repeated screening across seasons and environments have limited the number of active breeding programs that target these traits for improvement, despite their economic importance. Novel advanced phenomic tools have been developed that enables rapid non-destructive in-field screening for herbage quality, delivering high correlations between in-field and laboratory data sets. High throughput screening has been conducted on industry relevant germplasm generating large robust datasets across seasons, generating the ability to rapidly breed for productive plants with high nutritive value. Approaches for rapid genomic sub-selection of

varieties have been developed and exemplified enabling up to 5x the rate of genetic gain to be realised in comparison to conventional breeding. The incorporation of herbage quality genomic predictions into the rapid breeding approach has been achieved by developing a selection index for optimal outputs in collaboration with the commercial breeding partner. The delivery of advanced forage cultivars to the farming sector with a vastly broader array of traits for a wider variety of scenarios, can now be achieved, enabling the farming sector to be more productive, agile and responsive.

#### **W614: Grasslands (Lolium Genome Initiative)**

##### **Genomic Selection for Better Clover-Rhizobium Symbiosis**

Sean Weith<sup>1</sup>, Andrew G. Griffiths<sup>2</sup>, Ross Ballard<sup>3</sup>, Craig Anderson<sup>2</sup>, Sai Krishna Arojj<sup>2</sup>, M. Z. Zulfi Jahufer<sup>2</sup>, Grace O. Ehoche<sup>2</sup>, Dongwen Luo<sup>4</sup>, E. Eirian Jones<sup>1</sup> and Rainer Hofmann<sup>5</sup>, (1)Lincoln University, New Zealand, (2)AgResearch Grasslands Research Centre, Palmerston North, New Zealand, (3)PIRSA-SARDI, Australia, (4)AgResearch Grasslands Research Centre, New Zealand, (5)Lincoln University, Christchurch, New Zealand

White clover (*Trifolium repens* L.) is an integral component of mixed pastures in temperate agriculture, providing quality feed and a sustainable source of fixed nitrogen (N) through its symbiosis with soil-dwelling *Rhizobium* bacteria. While there has been much focus on identifying and applying *Rhizobium* strain inoculants with improved N-fixation, there has been less attention on identifying and exploiting plant genetic factors to develop cultivars that routinely form effective *Rhizobium* symbioses. The complexity of the clover-*Rhizobium* interaction makes for a challenging but valuable breeding target which represents an important strategy for reducing N-fertiliser use.

To determine the feasibility of deploying genomic selection for clover-*Rhizobium* symbiotic response, we grew 32 seedlings each of 118 half-sibling (HS) families in vermiculate/McKnight's solution where N was provided either by a commercial *Rhizobium* strain (TA1) or supplied mineral N (positive control). Traits shoot and root dry matter (DM) were recorded after 35 days of growth. Symbiotic potential (SP) for each HS family was calculated as DM produced in the *Rhizobium* treatment as a proportion of the positive control DM. The traits showed significant additive variation and narrow-sense heritabilities ranged from 0.24-0.33. A Smith-Hazel index to facilitate multi-trait selection for the largest plants with the greatest SP was derived. Cross validation of the genomic selection model KGD-GBLUP, based on 110,00 SNPs, generated predictive abilities ranging from 0.23-0.36.

These data provide insight into applying genomic selection for complex plant-microbial traits and will be used to make proof-of-concept among-and-within HS family selections for improved clover-*Rhizobium* symbiosis.

#### **W615: Grasslands (Lolium Genome Initiative)**

##### **Genomic Selection for Forage Yield in Tetraploid Perennial Ryegrass**

Stephen Byrne<sup>1</sup>, Dan Milbourne<sup>1</sup>, Sai Krishna Arojj<sup>1</sup> and Patrick Conaghan<sup>2</sup>, (1)Teagasc, Crop Science Department, Carlow, Ireland, (2)Teagasc, Grassland Science Research Department, Carlow, Ireland

Studies have shown that perennial ryegrass breeders have succeeded in achieving significant breeding gain for forage yield. Indirect selection using genome-wide markers is one tool that offers breeders an opportunity to further accelerate genetic gain for forage yield. In perennial ryegrass this can be achieved by enabling multiple cycles of genomic selection (GS) to be completed in the same time it takes to perform a single cycle of conventional selection. We established a small population of half-sib families by intercrossing plants of an elite commercial tetraploid cultivar. Maternal plants were genotyped using a genotyping-by-sequencing strategy and half-sib progenies were phenotyped for forage yield in replicated sward plots over 2 years under two cutting managements. Predictive models were developed using maternal genotype data and predictive ability for forage yield varied across different cutting periods and managements. Even low to moderate predictive abilities should encourage use of GS in routine selection

given our ability to complete at least five cycles of GS in the time it takes to complete a single cycle of conventional selection.

## **W616: Grasslands (Lolium Genome Initiative)**

### **Step-up to Gain: Prediction of Genomic Breeding Values and Variance in Perennial Ryegrass (*Lolium perenne* L.) Crosses**

**Leif Skot**<sup>1</sup>, Sarah Palmer<sup>2</sup>, Alan Lovatt<sup>1</sup>, Nastasiya Grinberg<sup>3</sup> and Rhys Kelly<sup>2</sup>, (1)IBERS, Aberystwyth University, Aberystwyth, United Kingdom, (2)IBERS, Aberystwyth University, United Kingdom, (3)University of Cambridge, United Kingdom

Perennial ryegrass is the predominant forage crop in temperate grassland agriculture. Reasons for this include its high digestibility, response to nitrogen fertilisation and persistency under grazing. However, acceleration of the rate of improvement of genetic gain is desirable. Genomic selection (GS) has the potential to achieve this, but prediction accuracies need to be sufficiently high to compete with phenotypic selection. Factors preventing adoption of GS include costs of genotyping, high sample numbers and the timeliness against the window of opportunity for GS to influence selections prior to sexual recombination in the breeding cycle. We believe our novel 'step-up' approach will deliver genetic gain beyond what is currently possible using conventional genotypic recurrent selection alone, reduce the numbers of progeny for GS to a practicable and realistic number. The step-up to gain approach is as follows: in parallel with the half-sib progeny test (HSPT), pairwise genomically estimated breeding values (GEBVs) will be predicted and targeted polycrosses (called Step-ups) created with few elite mother plants (MPs) predicted to give the most attainable genetic gain. Progeny of these step-ups will then be re-incorporated into the subsequent breeding nursery. In the current system the selected 400 plants from the nursery are allocated in sets of 100 into four pollen proof chambers. This gives us  $(100 \times 99/2) \times 4 = 19800$  possible pairwise combinations. Based on the genotyping of all 400 parents there are however,  $400 \times 399/2 = 79800$  theoretically possible pairwise combinations, and some of these missed combinations (because these were physically isolated in separate chambers) are predicted to be superior to the progeny actually obtained. Step-ups target these specific combinations with the aim of ensuring they become foundation MPs of the subsequent recurrent selection generation. We use the PopVar programme to provide predicted means, variances and superior progeny value from all theoretical pairwise combinations, based on genotypic data. The polycross group information and shortlisting of optimal pairwise combinations of traits is used to identify the most optimal 2-16 way crosses, which we proceed with in our Step-up generation. This approach will allow us to include promising combinations not physically possible from the polycross arrangements. Progeny from step-ups are evaluated in terms of superiority of phenotype measured from HSPT plot trials, compared to baseline progeny.

## **W617: Honeybee Genomics**

### **European Honeybee Subspecies Genomes at the Individual and Population Levels**

**Alain Vignal**<sup>1</sup>, Benjamin Basso<sup>2,3</sup>, Kaspar Bienefeld<sup>4</sup>, Kamila Canale-Tabet<sup>5</sup>, Cecilia Costa<sup>6</sup>, Cécile Donnadieu<sup>7</sup>, Sonia Eynard<sup>8</sup>, Aleš Gregorc<sup>9</sup>, Christophe Klopp<sup>10</sup>, Melanie Parejo<sup>11,12</sup>, M. Alice Pinto<sup>13</sup>, Yves Le Conte<sup>2</sup>, Bertrand Servin<sup>14</sup> and David Wragg<sup>15</sup>, (1)GenPhySE, INRAE, ENVT, ENSAT, 31326, Castanet Tolosan, France, (2)INRAE, France, (3)ITSAP, France, (4)Bee Research Institute, Germany, (5)INRA, France, (6)CREA, Italy, (7)INRAE GeT Genomics Facility, Castanet-Tolosan, France, (8)GenPhySe, INRAE, ENVT, ENSAT, 31326, Castanet Tolosan, France, (9)Agricultural Institute of Slovenia, Slovenia, (10)MIAT, INRA, Université de Toulouse, Castanet-Tolosan, France, (11)UPV/EHU, Spain, (12)Agroscope, Switzerland, (13)CIMO, Portugal, (14)GenPhySE, INRAE, ENVT, ENSAT, Toulouse, France, (15)Roslin Institute, United Kingdom

Throughout western Europe, the endemic honeybee subspecies *Apis mellifera mellifera*, the black bee, has gradually been replaced in many beekeeper operations by other subspecies and hybrids of *A. m. ligustica*, *A. m. carnica*, *A. m. caucasica*, (C-type bees) and *A. m. mellifera* (M-type bees), which were

found to be more efficient producers of honey or royal jelly, and also to be less aggressive. In order to understand the impact of these practices on the genomes of bee populations we followed two complementary approaches. The first was to produce a high-continuity reference genome for M-type bees by PacBio sequencing, for comparison with the current Amel\_HAV3.1 reference - a C-type bee. A single haploid drone from a closed population was selected for sequencing and the overall genome assembly obtained is 227 Mb for the 16 chromosomes. The longest contig obtained is 11.6 Mb, the N50 contig size is 5.1 Mb and chromosome 16 was obtained as a single 7.2 Mb contig. Contigs were ordered along chromosomes by minimizing genetic recombination. Comparison with Amel\_HAV3.1 allowed the high confidence detection of inversions between the two reference genomes. Long-read sequencing also allowed the discovery of new highly repeated elements. The second approach is the sequencing of several hundred haploid drones from separate colonies, which includes black bee conservatories, French queen breeders (honey and royal jelly production) as well as out-groups from various European regions from where genetic stock is imported for breeding purposes. Varying degrees of admixture between the M- and C-type bees are observed, often reflecting different breeding practices. Regions of high differentiation between the two genetic types are also detected.

## **W618: Honeybee Genomics**

### **16andBee - Using Genomics to Unravel Unknown Family Relationships in Honeybees**

**Gertje Petersen**, AbacusBio, Dunedin, New Zealand, Peter F. Fennesy, AbacusBio Ltd, New Zealand and Peter Dearden, Department of Biochemistry, University of Otago, New Zealand; Genomics Aotearoa, Dunedin, New Zealand

Inbreeding and lack of genetic diversity can present a problem in livestock breeding since it reduces adaptability of the population and decreases the amount of genetic and phenotypic variation that can be selected from. Inbreeding effects in honeybees can be observed both in terms of reduced viability of inbred queens and workers as well as the sex determination system based on the gene *csd* (complimentary sex determiner). During embryo development of the honeybee, the number of *csd* alleles present in each individual determines the sex of the embryo via a gene expression cascade spearheaded by *csd*. Through this process, unfertilised eggs will develop into hemizygous haploid male drones and fertilised eggs with two different *csd* alleles into diploid female workers and queens. However, in populations with low numbers of *csd* alleles, there is a chance of “matched” mating of individuals with identical *csd* alleles. Their offspring, being homozygous at the *csd* locus, have no way of activating the gene cascade that would allow them to develop into females and instead develop into diploid males. While these are viable, they are usually detected as anomalies by worker bees and killed as larvae. Lack of genetic diversity in a honeybee population thus has a direct and devastating effects on worker brood viability.

Due to the global displacement of European Honeybees and resulting importation bottlenecks, populations on other continents can generally be at risk of the effects of reduced genetic diversity. Traditionally, these risks can be managed through constraining relationships in breeding populations based on known pedigree of both parental animals. In honeybees, this approach is often difficult to impossible due to queen mating behaviour. Mating on the wing as well as extreme polyandry with 6 to 25 drones creates a situation in which known pedigree is often limited to the maternal side. Even if drone contribution to a mating can be managed (e.g. via artificial insemination), it is still impossible to determine exact paternity. Since polyandrous mating serves to create a single colony with a flexible workforce, limiting effective paternity to a single drone or even drones from a single drone source/mother can be detrimental to colony viability. While this risk can be worth taking in a breeding population where colonies can be managed intensively, it is not a feasible strategy on a larger scale.

Due to these limitations, honeybee relationships can really only be fully determined with the use of DNA sequence. Conventional methods for the determination of relationships based on sequencing data can however fall short due to haplodiploid pedigree structure, which creates half-sisters (related by 0.25), full sisters (related by 0.5) and supersisters (related by 0.75). With the use autosomal genetic information (e.g. whole genome sequence or reduced representation genotyping such as Genotyping-by-

Sequencing), putative family relationships can be confirmed, but unknown relationships can often not be fully determined. Taking into account of both *csd* alleles and autosomal sequence however can potentially provide the necessary resolution to determine unknown relationships.

### **W619: Honeybee Genomics**

#### **Reproductive Variation and Admixture in Honeybee Populations of the United States**

**Brock Harpur**, Purdue University, West Lafayette, IN

Beekeepers are faced with severe challenges that affect their colonies' health. Among the most highly-reported of those challenges is queen failure. Queens can fail because they have mated with drones that have a low sperm count. Here, I present our multi-omic, field-based approach to quantify the genetic features associated with poor sperm production in drones within the United States.

### **W620: Honeybee Genomics**

#### **Facing the War between Honey Bee and Mites: Genomic Insights into Varroa Global Success**

**Maeva Angelique Techer**, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

In the last decades, Western honey bee populations had to face and continuously fight against tenacious Asian enemies: ectoparasitic *Varroa* mites. The Western honey bee (*A. mellifera*) has been introduced into the native range of its sister species, the Eastern honey bee (*A. cerana*). This new, sudden contact between both species allowed parasite spillover. About 70 years ago, *Varroa destructor*, took this opportunity and successfully switched hosts from the Eastern to the Western honey bee. This parasite rapidly spread quasi-worldwide and became the main driver of honey bee colony losses. Concerns arise as ten years ago, an emerging threat, *V. jacobsoni*, also jumped onto the Western honey bee in Papua New Guinea. Despite a reported loss of genetic diversity and quasi-clonality in invasive populations, mites are persisting and evolving new traits such as acaricide resistance.

To help turning the tide in the *Varroa*-honey bees-keeper war, we recently developed new genomic resources to identify the mechanisms behind the mites' success. Here, I will present i) the first genomic insights into the demographic origins of *V. destructor* and *V. jacobsoni* jumps. Using mite species collected in their native ranges from both the ancestral and novel hosts, we were able to correctly reconstruct the temporal dynamics of the switch. We further found that hundreds of haplotypes were involved in the initial host switch, and, while greatly reduced, some gene flow between mites adapted to different hosts remains. Finally, I will also briefly introduce ii) the development of a cost-effective targeted genotyping approach to leverage the genetic variability to identify and detect mite populations. This method, combined with a world collection, will aim to retrace the global spread and evolution of *V. destructor*.

### **W621: Honeybee Genomics**

#### **Genome-Wide Patterns of Genetic Differentiation of U.S. Commercial Honey Bee Stocks**

**Perot Saelao**, USDA-ARS Honey Bee Breeding, Genetics, and Physiology lab, DAVIS, CA; USDA-ARS, Baton Rouge, LA

Identification of genomic signatures across seven US commercial honey bee lines using whole genome pooled sequencing. The genetic diversity of honey bee stocks throughout the United States is poorly characterized. An improved understanding of genetic diversity would support developing markers that can be used to identify and breed bee populations of interest. Genomic signatures and molecular markers serve to easily identify samples and classify specific groups with a high degree of discriminatory power. The identification of these markers enables a more informed selection of genes and markers that can aid in developing a broader-based approach to stock identification and selection. In this study, seven stocks of honey bees (Russian, Minnesota Hygienic, Pol-line/VSH, Italian from three commercial

sources, and Carniolan from one commercial source) were whole genome sequenced in order to detect unique stock-based SNPs, indels, and structural variants. Population differentiation based on allele frequencies demonstrated unique genetic structure of these stocks. Many of these stock specific genetic variants can help to characterize the genetic foundation of important traits for selection, such as Varroa sensitive hygiene. This sequence information will help support future marker-assisted selection efforts.

## **W622: Honeybee Genomics**

### **Statistical Imputation of Queen Genotype from Pool Sequencing of Workers**

Bertrand Servin, GenPhySE, INRAE, ENVT, ENSAT, Toulouse, France, **Sonia Eynard**, GenPhySe, INRAE, ENVT, ENSAT, 31326, Castanet Tolosan, France, Fanny Mondet, Abeilles et Environnement, INRAE, France, Lucie Genestout, EVOLUTION, NOYAL SUR VILAINE CEDEX, France, Olivier Bouchez, INRAE GeT Genomics Facility, Castanet-Tolosan, France and Alain Vignal, GenPhySE, INRAE, ENVT, ENSAT, 31326, Castanet Tolosan, France

Characterising the genetic diversity of populations allows to better understand their demographic history and their adaptation to selective pressures. In honey bees, this characterisation is facilitated by a relatively small genome size, but is hindered by the fact that often the unit of observation and sampling is the colony rather than a single individual. Moreover, performing large scale genetic analyses of honey bees is a real challenge, due to the specific reproduction mechanism including multi-male insemination, making the genotype of a bee colony a mixture of contribution from the queen and the mating drones. In this work we propose an approach to characterise the genotype of a colony based on pool sequencing of worker bees. We introduce statistical models for the analysis of pool sequence data allowing to reconstruct jointly individual queen genotypes of colonies and allele frequencies in bee populations. We demonstrate the performance of our approach using data on 1500 colonies collected throughout three years within the FranceAgriMer funded, BeeStrong project. Population admixture, in terms of queen subspecies composition, validation was accomplished using information on geographical and sociological organisation of the beekeepers. In addition to a better understanding of the population dynamics of honey bees, our approach to genotyping bee colonies promises to facilitate the genetic analysis of complex traits, and can be used for genome wide association studies on phenotypes of interests, for instance to assess Varroa resistance in honey bee populations.

## **W623: Host-Microbe Interactions**

### **Are All *Pseudomonas syringae* infections the Same? What Plant Responses Can Tell Us about Taxonomy and Diversity within a Single Microbial 'species'.**

**David A. Baltrus**, School of Plant Sciences, University of Arizona, Tucson, AZ

Ongoing improvement of modern DNA sequencing technology are continuing to revolutionize genomic analyses in phytopathogens. Long read sequencing of bacterial chromosomes and plasmids enables the assembly of complete genomes, and therefore provides structural information that is otherwise unavailable in draft genomes. Cost efficient short read sequencing enables experiments like dual host-pathogen RNAseq. Here I explore the possibilities of both types of tech for the investigation of pathogenesis in the well-studied phytopathogen *Pseudomonas syringae*. Specifically, I show how complete genome sequences coupled with comparative genomics enables the identification of previously undescribed pathways for the production of a well-

studied toxin, tagetotoxin, and how dual host-pathogen RNAseq highlights divergent infection strategies for two strains of *P. syringae* on a common host.

#### **W624: Host-Microbe Interactions**

##### **Genome Wide Association Mapping of Co-Transcriptome Variation in the Interaction of *Botrytis cinerea* with Eudicots**

**Daniel Kliebenstien**, University of California, Davis, Davis, CA

Host/pathogen studies often focus on large effect genes in epidemic pathogens. This has developed a general molecular model where the interaction of proteins or metabolites from the host and pathogen trigger an “immune” response to create qualitative resistance. It is not clear how this model translates to broad host range endemic quantitative pathogens. To test this, we sequenced the genome of 96 diverse *Botrytis cinerea* isolates. The pathogen contains high levels of genetic with evidence for diversifying selection at known virulence loci such as toxin metabolite clusters and cell wall degrading genes. However, most selected regions had loci previously not linked to virulence and are likely identifying new virulence mechanisms. We used this genetic variation in the pathogen to study the host by infecting all 96 isolates on *Arabidopsis thaliana* and measuring both species transcriptomes. This identified extensive interaction between the two species centering on pathogen toxin production and host photosynthesis. Genome wide association mapping of virulence and the two transcriptomes showed a highly polygenic architecture controlling the interaction. This included hotspots in the pathogens genome that controlled virulence as well as a large number of transcripts in either the host or pathogen. We are extending this analysis to other dicots including Tomato, Lettuce, Chicory, Sunflower, Soybean and Brassica. Virulence on these species highly polygenic with a minimal effect of domestication on the host/pathogen interaction. Most plant species showed an increased resistance in the domesticated germplasm and a similar range of variation between wild and domestic germplasm.

#### **W625: Host-Microbe Interactions**

##### **Rapid Metagenomics for Livestock Antimicrobial Use Decisions**

**Anatoliy M Trokhymchuk**, Prairie Diagnostic Services, Saskatoon, SK, Canada

Genomic ASSETS (Antimicrobial Stewardship Systems for Evidence-based Treatment Strategies) for Livestock is a revolutionary application of rapid metagenomics to inform antimicrobial use in livestock. Antimicrobial resistance (AMR) is a global threat to public and animal health, placing unprecedented pressure on agriculture to reduce antimicrobial use. Current diagnostic AMR tests in livestock take 5-7 days. A metagenomics-based diagnostic system will provide results to practicing veterinarians within hours. Bovine Respiratory Disease (BRD) is the most common cause of morbidity/mortality and reason

for parenteral antimicrobial use in feedlot cattle in North America and is a prime target for antimicrobial stewardship. The objective of Genomic ASSETS is to develop a diagnostic system that uses a point-of-need metagenomic workflow to inform antimicrobial use decisions for BRD in feedlot cattle at the herd (e.g., pen) level.

The scope of project work includes 5 phases. Phase one includes “wet bench” process optimization (nucleic acid extraction, library preparation, sequencing, primary bioinformatic analysis). In phase two diagnostic sample selection and pen-level sampling strategies will be developed and bioinformatic pipelines optimized so diagnostic data can inform clinical decisions. Phase three will include epidemiologic analysis to link genomic data to calf health outcomes through dynamic risk assessment model. Phase four is a field roll-out of the genomic testing strategy to commercial feedlots in Alberta and Saskatchewan. Phases 1-4 include qualitative studies that engage the beef production chain to inform the design of the diagnostic system. Phase 5 will be the capstone of the project and conduct economic analysis of the system to ensure optimization for the beef feedlot industry.

Genomic ASSETS will change the way we make antimicrobial use decisions for cattle and other livestock by providing rapid, robust diagnostic information to end users.

### **W626: Hybridization, heterosis and balancing selection**

#### **Is the Magnitude of Heterosis Correlated with Transcriptome Size?**

Hua Yang<sup>1</sup>, Xiaowen Shi<sup>1</sup>, Tieming Ji<sup>1</sup>, Jianlin Cheng<sup>2</sup> and **James A. Birchler**<sup>1</sup>, (1)University of Missouri, Columbia, MO, (2)University of Missouri \_ Columbia, Columbia, MO

It has long been known that cell size correlates with the level of ploidy of a cell. More recent studies indicate that the transcriptome size correlates with ploidy although not necessarily strictly linearly. In example cases, the transcriptome size tracks more closely with cell size. In an analysis of various *Zea* and *Tripsacum* species and their hybrids at various ploidy levels, it was revealed that the size of mature leaf epidermal cells was positively correlated with ploidy but also with the degree of biomass heterosis in diverse hybrids. To pursue this observation further, different diploid hybrids previously shown to exhibit a gradient of biomass heterosis showed a cell size correlation with the magnitude of heterosis. Further, tetraploid homozygous lines were compared to single and double cross tetraploid hybrids, the latter of which exhibits greater heterosis. Again the cell size correlated with ploidy but further with the magnitude of heterosis. If cell size is a reflection of overall transcriptome size, these results suggest that hybrid genotypes that trigger heterosis cause a generalized increase in transcription. Such an effect is likely not fully uniform across all genes given that RNASeq studies from numerous labs find some level of relative changes in hybrids. The lesser strict correlation between ploidy and transcriptome size is potentially affected by varying degrees of the heterosis effect. These considerations would need to be accommodated into genetic and molecular models to explain heterosis.

### **W627: Hybridization, heterosis and balancing selection**

#### **Heterosis Dissection**

**Xuehui Huang**, Shanghai Normal University, Shanghai, China

Since approximate a century ago, many hybrid crops have been continually developed by crossing two inbred varieties. Owing to heterosis (hybrid vigor) in plants, these hybrids often have superior agricultural performances in yield or disease resistance succeeding their inbred parental lines. Several classical hypotheses have been proposed to explain the genetic causes of heterosis. During recent years, many new genetics and genomics strategies have been developed and used for the identifications of heterotic genes in plants. In hybrid rice, 17 populations containing totally 10,074 F<sub>2</sub> lines were used for searching the key heterosis-related gene loci, and similar approaches were also used in hybrid maize. Heterotic effects of the heterotic loci and molecular functions of the heterotic genes are being investigated in many important crops. More and more data and knowledge coming from the molecular studies of heterotic loci and genes will serve as a valuable resource for hybrid breeding by molecular design in future. In this

workshop, we aim to address recent advances in our understanding of the genetic and molecular mechanisms of heterosis in rice and maize. The remaining scientific questions on the molecular basis of heterosis and the potential applications in breeding will be discussed as well.

### **W628: Hybridization, heterosis and balancing selection**

#### **Changes in Total and Mitochondrial Proteomes Associated with Different Levels of Heterosis in Maize**

**Kathleen Newton**<sup>1</sup>, Diwakar Dahal<sup>2</sup> and Brian P Mooney<sup>2</sup>, (1)University of Missouri, Columbia, MO, (2)University of Missouri

Maize hybrids exhibiting heterosis were generated from inbred parents with increasing genetic distance. B73 was used as the common female parent in crosses with N192 (low heterosis), MO17 (high-heterosis 1), and NC350 (high-heterosis 2). Total and mitochondria-enriched proteomes were analyzed from ear shoots of field-grown hybrids and their inbred parents. Two-dimensional gel electrophoresis and GeLCMS (1D SDS-PAGE fractionation, trypsin digestion, LTQ Orbitrap nano-RP-LC MS/MS) were used to analyze changes in protein abundance. Although the differentially abundant proteins associated with increased heterosis belong to various pathways, protein and RNA metabolism as well as stress responsive proteins were the major classes changed in the higher heterosis hybrids. Additionally, the GeLCMS data revealed a significant change in abundance of a mitochondria-associated target of rapamycin (TOR) protein. The changes observed were consistent with a stress pre-priming model for increased vigor along with a role for the TOR kinase in heterosis.

### **W629: Hybridization, heterosis and balancing selection**

#### **Gene Expression and Heterosis**

**Samantha Snodgrass** and Matthew B. Hufford, Iowa State University, Ames, IA

The genetic mechanism of heterosis has been under serious study and debate for over 100 years, resulting in several possible and overlapping models. The complementation or dominance model of heterosis has the most empirical support, the other models being overdominance and epistasis. The complementation model posits heterosis is due to the accumulated masking of recessive, deleterious alleles. This implies that more distantly related populations with more accumulated drift effects will produce hybrids with higher heterosis. However, outbreeding depression can occur from hybridization of populations too different from each other or have adapted to different environments. Theoretically then, the amount of heterosis in the hybrid is distributed by the genetic distance of the parents, where heterosis is low at the extremes of genetic distance but high at some intermediate genetic distance. Most studies have focused on genic SNP variants between parents to test mechanistic models of heterosis. Recent genomic research has demonstrated the importance of regulatory and structural variation on phenotype. While there is strong theoretical support for incorporating these genetic variants into heterosis models, only a handful of studies have demonstrated the variants' potential impact.

Studying single parent expression (SPE) offers a chance to understand the importance of regulatory and structural variation and directly test the complementation model. SPE is a gene expression pattern in which the hybrid and one parent express a gene while the other parent does not. This particular expression pattern has been observed in wheat and maize and likely occurs more broadly. While expression complementation fits the complementation model, the underlying genetic sequence or mechanism driving this pattern has not been studied. We have developed a full diallel population with reciprocals from thirteen diverse maize lines to address these questions and have taken phenotypic data from a pilot study this past summer. Phenotypic data of this diallel may define the relationship between heterosis and genetic distance of the parents as well as tie incidence of SPE to heterosis. In this talk, I will explain how this experimental design can leverage information at the genomic scale to directly test classical genetics models.

### **W630: Hybridization, heterosis and balancing selection**

## **Towards Better Exploitation of Heterosis in Commercial Hybrid Grain Sorghum**

**David R Jordan**<sup>1</sup>, Alan W Cruickshank<sup>2</sup>, Colleen H Hunt<sup>2</sup>, Adrian M Hathorn<sup>3</sup>, Yongfu Tao<sup>1</sup> and Emma S Mace<sup>1</sup>, (1)The University of Queensland, Warwick, QLD, Australia, (2)Department of Agriculture and Fisheries, Queensland Government, Warwick, QLD, Australia, (3)The University of Queensland, Brisbane, Australia

Cytoplasmic-nuclear male sterility (CMS) is used to facilitate the production of hybrid seed in sorghum. The system relies on maternally inherited mitochondrial genes that interfere with critical metabolic processes involved in the formation of viable pollen, combined with dominant nuclear restoration genes that are able to mitigate the impacts of the mitochondrial genes allowing production of fertile pollen. This system imposes constraints on the exploitation of heterosis in sorghum by restricting genetic diversity, particularly in the female heterotic pool. In this presentation we will highlight the molecular basis of these constraints as well as opportunities to use this knowledge to improve hybrid performance.

## **W631: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**

### **The Breeding Management System - Development and Deployment**

**Graham McLaren**, Integrated Breeding Platform, Texcoco, Mexico

The Breeding Management System (BMS) of the Integrated Breeding Platform (IBP) is now in its tenth year of development and fifth year of deployment as a multi-user institutional breeding information management system.

The BMS has transitioned from a stand-alone application into a modern multi-user browser based application which can be deployed on institutional LANs or in the Cloud, making it ideal for corporate use within the confines of a company and for collaborative use in public research where research teams are widely dispersed. The BMS is an ontology-driven system, which scales well in both these situations supporting multiple crops, hundreds of users with distinct roles and permissions, large nurseries with thousands of entries and large trials with hundreds of environments. It is an open system with an exposed Application Programming Interface, which includes the BrAPI standards for interoperability.

To date the BMS is deployed in numerous stand-alone situations around the world and in 34 institutions from international research organizations, to national research programs and in private companies. The IBP will continue to deploy the BMS with subsidies for national programs in Africa and on a cost-recovery basis for advanced institutes and commercial enterprises.

## **W632: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**

### **Modernizing Plant Breeding Programs in Ethiopia**

**David R Jordan**<sup>1</sup>, Emma Mace<sup>2</sup>, Amare Seyoum<sup>3</sup>, Nigussie Girma<sup>3</sup>, Berhanu Fenta<sup>3</sup>, Girum Azmach<sup>3</sup>, Lealem Tilahun<sup>3</sup>, Demissew Ababulgu<sup>3</sup>, Habte Zegeye<sup>3</sup>, Mekuria Dejene<sup>3</sup>, Taye Tadesse<sup>3</sup>, David Rodgers<sup>2</sup>, Alison Kelly<sup>4</sup>, Errol Corsan<sup>4</sup> and Alemayehu Assefa<sup>3</sup>, (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, Australia, (2)The University of Queensland, St Lucia, QLD, Australia, (3)Ethiopian Institute of Agricultural Research, Ethiopia, (4)The University of Queensland, Australia

Public plant breeding programs in many parts of Africa have suffered from a deficit of long term funding and limited exposure to modern plant breeding methods. In this presentation, we report on the experiences of the partnership between the Ethiopian Institute for Agricultural Research and the University of Queensland over a period of 7 years funded by the Bill and Melinda Gates Foundation. This partnership has involved comprehensive modernization of 8 field crop breeding programs based on quantitative genetics principles to improve genetic gain including benchmarking of existing programs,

planning (including product profile development and breeding pipeline redesign), crossing strategies, digitisation of data collection and storage, implementation of modern statistical methods and relevant mechanisation. These changes have been implemented with a strong focus on continuous improvement and resource use efficiency. The successes and challenges will be discussed.

### **W633: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding Informatics Support for Plant Breeding Programs at ICARDA**

**Zakaria Kehel**, International Center for Agriculture Research in the Dry Areas ICARDA, Rabat, Morocco

ICARDA provides innovative and science-based solutions for communities across the non-tropical dry areas. ICARDA has a global mandate for the crop development of barley, lentil, and faba bean and a shared mandate for wheat and chickpea respectively with CIMMYT and ICRISAT. The biodiversity and crop improvement program invested in modernization of its breeding programs including use of barcoding, digital data collection and Breeding Management System (BMS). The later was even of a necessity in the context of ICARDA decentralization strategy. Breeding programs at ICRDA adopted BMS since 2016 and it has allowed ICARDA to safeguard historical data, share breeding activities and plans between the main breeding stations in Morocco, Lebanon and India. ICARDA is structuring its BMS instance in multiple databases, one database per crop. Each crop is also structured in several programs including prebreeding, breeding and international nurseries.

Optimal breeding data management is a key component in breeding programs to allow data driven breeding decisions. Access to different data (phenotypic, pedigree and trials) and trial analysis including genotype's selection should be easy and timely. In addition, access to historical data is essential in breeding programs to allow breeders comparing patterns of GxE and draw conclusions on target population environments, predicting additive variance, and finally extract information to run genomic selection.

Even though BMS includes a comprehensive data analytic component that is used by ICARDA breeding teams, ICARDA in collaboration with CIRAD and IBP is developing an R package to query BMS database using BrAPI for a full integration with the molecular data management system (GIGWA) and optimal access for more complex analysis such as GxE, GWAS and GS.

ICARDA collaborates with IBP and contributes to resolve the technical challenges of SSL certificate installation and setup on BMS server, improve the backup batch script on Amazon Web Services (AWS) to perform remote backup on S3 bucket, and help in prioritization of adopting BrAPI calls in BMS API upon needs of integration with more complex analysis.

### **W634: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding Modern Analytical Tools for Plant Breeding and their Integration into Information Management Systems**

**Fred A. van Eeuwijk**, Wageningen University & Research - Biometris, Wageningen, Netherlands

### **W635: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding Experiences of the Integrated Genotyping Service and Support (IGSS) in Modernizing Plant Breeding in Africa**

**Cathrine Ziyomo**, BECA-ILRI hub, Nairobi, Kenya

The establishment of the IGSS platform in 2014 with funding from the Bill & Melinda Gates foundation at the Biosciences eastern and central Africa (Beca)-ILRI hub in Kenya has significantly contributed to building a critical mass of African breeders technically competent in the use of DNA markers for crop improvement. In addition, the platform has increased accessibility and utilization of genotypic data for research and breeding. The IGSS is jointly operated by Diversity Array Technology Pty Ltd (DArT), an Australian based genotyping and information technology company – and the (Beca) – ILRI Hub – a shared agricultural biosciences facility in Nairobi. Beginning 2020, the platform will transition to a self-

sustaining operations as a non-profit organization called **SeqArt**. Samples genotyped have been aggregated from CGIAR centers, gene banks, small private seed companies, National Agricultural Research Institutions (NARIs) and academic institutions. Genotypic data generated from the platform has been useful in breeding programs in a diversity of ways including germplasm characterization and diversity studies, genome wide association studies, bi-parental mapping, genomic selection, monitoring of varietal adoption, quality control in seed production, and profiling of parental lines. However, these data have not fully translated into improved breeding methodologies or crop varieties due to limited resources and technical knowledge at many African breeding stations. Currently, most plant breeders in Africa rely on phenotypic selection alone, and lack the institutional capacity and resources required for successful implementation of marker assisted selection. There is a need for strengthening breeding teams within NARS institutions and establishing a sustained support service for breeding applications.

### **W636: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**

#### **10 Years to Digitize Breeding Programs in Sub-Saharan Africa: Achievements, Lessons Learned and Perspective**

**Jean-Marcel Ribaut**, Integrated Breeding Platform, Texcoco, Mexico and IBP Team

Innovation in plant breeding is imperative to meet the growing demand for staple food crops in developing countries. Modernizing breeding was therefore a major objective of the Generation Challenge Programme (GCP, <http://www.generationcp.org>). In this endeavor, the GCP created the Integrated Breeding Platform (IBP, <https://www.integratedbreeding.net>), to provide breeding material, knowledge and tools to assist researchers in their work, including custom-built software for reliable data management – the Breeding Management System (BMS Pro). These activities were sustained mainly through funding by the Bill & Melinda Gates Foundation, which ended this last September after 10 years of direct collaboration. The IBP has proven to be agile, adaptable and bold over the years, and is now applying the same spirit and resolve to find revenue from both public and private sources to continue serving its broad basis of stakeholders, among which national programs in Sub-Saharan Africa (SSA) remain front and center. BMS Pro – a professional-grade software package distributed through LAN or cloud – is being used by close to 700 users in over 30 organizations of different types around the world (17 in SSA). We have learned that digitizing breeding is less about technology than it is about changing mindsets; it requires proper support on the ground, and that management commits to empower adoption within institutions. Although there is still some way to go before reaching routine adoption, a solid basis has been established and continues to be supported by a new generation of African breeders. Breeding digitization in Africa is well underway.

### **W637: International Cotton Genome Initiative (ICGI)**

#### **Upland Cotton Genome Sequence**

**Z. Jeffrey Chen**, The University of Texas at Austin, Austin, TX

### **W638: International Cotton Genome Initiative (ICGI)**

#### **Neofunctionalization of lncRNAs Originated in Cotton Interspecific Hybridization and Polyploidization**

**Xueying Guan**, Zhejiang University, Hangzhou, China

Interspecific hybridization and whole genome duplication (WGD) are known as driving forces behind genomic and organism diversification. The effect of hybridization and WGD on non-coding region of the genome in particular remain largely unknown, however. The non-coding transcripts represent the unique pattern for each species both in plant and animal genomes. In this study, we examined the profile of long non-coding RNAs (lncRNAs), comparing them with that of coding genes in allotetraploid cotton (*Gossypium hirsutum*, AADD), its putative diploid ancestors (*G. aboreum*, AA; *G. ramondii*, DD) and an (*G. aboreum* x *G. ramondii*, AD) F<sub>1</sub> hybrid. We found most lncRNAs (80%) were allelic-expressed in the allotetraploid genome. Moreover, the genome shock of hybridization reprogrammed the non-coding

transcriptome in the F<sub>1</sub> hybrid. With a low throughput functional screening of 100 lncRNA gene candidates via VIGS technology, we found about 20% of lncRNAs were involved with biological functions in cotton seedling. A representative novel lncRNA XLOC\_409583 activated after polyploidization from a LINE in the A subgenome of allotetraploid cotton, was involved in control of cotton seedling height. We found new lncRNAs were hyper reactive to abiotic stress. Our results reveal that the processes of hybridization and polyploidization enable the neofunctionalization of lncRNA transcripts, acting as important sources of increased evolutionary plasticity for plants.

### **W639: International Cotton Genome Initiative (ICGI)**

#### **Genome-Wide Variations Provide Insight into the Genetic Architecture of Cotton Elite Lines**

**Avinash Sreedasyam**<sup>1</sup>, Jerry Jenkins<sup>1</sup>, Christopher B. Plott<sup>1</sup>, Christopher A. Saski<sup>2</sup>, Fred Bourland<sup>3</sup>, Warwick N. Stiller<sup>4</sup>, Peng Chee<sup>5</sup>, Jane Grimwood<sup>1</sup>, Jeremy Schmutz<sup>1</sup> and Don C. Jones<sup>6</sup>,

(1)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (2)Clemson University, Clemson, SC, (3)University of Arkansas, Fayetteville, AZ, (4)Commonwealth Scientific and Industrial Research Organisation (CSIRO), Narrabri, NSW, Australia, (5)University of Georgia-Tifton Campus, Tifton, GA, (6)Cotton Incorporated, Cary, NC

Allotetraploid cottons, despite their wide-geographic distribution retained collinearity, gene order and content with limited genotypic diversity. Public cotton breeding programs have been instrumental in developing elite cotton lines with likely narrow genetic backgrounds but varied phenotypic performance. In this study, we generated high-quality reference genomes for three such elite cotton lines using long read sequence technology. We identified genome-wide variations among these lines and the publicly available *Gossypium hirsutum* L. acc TM-1 genome. We will present the likely causal genetic elements involved in phenotypic differences between the elite lines with reference to TM-1. These findings may facilitate future genetic efforts to further enhance yield and sustainability of this economic crop. Furthermore, the genomic resources and analysis methods developed in this project will be of great use for cotton and other crop research communities conducting similar efforts.

### **W640: International Cotton Genome Initiative (ICGI)**

#### **Development and Characterization of a Reciprocal Set of Interspecific Near Isogenic Lines in Cotton**

**Jeevan Adhikari**<sup>1</sup>, Rahul Chandnani<sup>2</sup>, Ellen Regina E Skelton<sup>3</sup> and Andrew H. Paterson<sup>3</sup>, (1)The University of Georgia, Athens, GA, (2)University Of Georgia, Athens, GA, (3)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

Cotton (*Gossypium spp.*) is a major source of natural fiber for textile industries around the globe. Mature cotton fibers, produced from tiny seed trichomes during four overlapping stages of development (fiber initiation, cell elongation, secondary cell wall deposition and maturation), are valued for their quality as defined by length, strength, fineness, elongation and uniformity. A lot of physiological changes, driven by associated transcriptome alterations, occur during these four stages of growth and development of cotton fiber. Identification and understanding of these alterations are important to dissect the stages involved in transforming primitive trichomes to the economically important fibers of modern cotton cultivars. To better understand the stages involved in transformation of the epidermal cell into mature cotton fiber and identify associated transcriptomic alterations, we are constructing a reciprocal set of near-isogenic lines (NILs) using Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) as the parents, each of which contains one and only one introgressed segment from the donor genotype, but collectively covering the entire donor genome. These NILs, each consisting about 0.5% of the donor genome in a reference background, also provide a powerful tool for genetically dissecting complex traits, increasing the precision with which phenotypic changes can be mapped to transcriptomic and genetic alterations. Genotyping by sequencing of BC<sub>5</sub>F<sub>1</sub> lines revealed an average of 3.6 introgressed segments in the population with Acala Maxxa as the recurrent parent. Genotyping is ongoing for the other set of NILs with Pima S6 as the recurrent parent. Introgressions will be verified in the BC<sub>5</sub>F<sub>2</sub> progenies of each

line with SSR markers selected from the introgressed regions and verified NILs with single introgressions will be phenotyped and characterized for various fiber quality related traits.

#### **W641: International Cotton Genome Initiative (ICGI)**

##### **MAGIC Population from 11 Upland Cotton Varieties Reveals Genetic Loci Affecting Agronomic and Fiber Phenotypes Including Flame Retardancy of Textiles**

**Gregory Thyssen**, Cotton Fiber Bioscience & Cotton Chemistry and Utilization Units, New Orleans, LA, David D. Fang, Cotton Fiber Biosciences Research Unit, USDA-ARS-SRRC, New Orleans, LA, Johnie N Jenkins, USDA-ARS, Mississippi State, MS, Jack McCarty, Genetics & Sustainable Agriculture Research Unit, USDA-ARS, Mississippi State, MS, Linghe Zeng, USDA ARS, Stoneville, MS, B. Todd Campbell, USDA-ARS, Florence, SC, Doug J Hinchliffe, Cotton Chemistry and Utilization Unit, USDA-ARS-SRRC, New Orleans, LA, Brian D. Condon, Cotton Chemistry and Utilization Research Unit USDA-ARS-SRRC, New Orleans, LA, Chris D. Delhom, USDA ARS SRRC, New Orleans, LA, Md Sariful Islam, Sugarcane Production Research Unit, USDA-ARS, Canal Point, FL and Don C. Jones, Cotton Incorporated, Cary, NC

Multi-parent advanced generation intercross (MAGIC) populations are useful to breeders and researchers because novel combinations of beneficial alleles from diverse parents may generate transgressive segregation and extreme or novel traits. Based on whole genome sequencing of 550 MAGIC RILs, we present loci and candidate genes that are associated with cotton fiber traits including length, strength and maturity, from multiple location-years. We further identified specific genes that control herbicide tolerance and nematode resistance in the MAGIC population which we functionally confirmed by gene silencing. We also identified RILs with natural flame retardant properties that dramatically exceed the parental lines. Surprisingly, since this is not true of any of the parents or typical commercial cotton varieties, non-woven fabric made from several RILs can self-extinguish after exposure to fire.

#### **W642: International Cotton Genome Initiative (ICGI)**

##### **New Variants of CRISPR RNA-Guided Genome Editing System in Cotton**

**Shuangxia Jin**, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

As an allotetraploid, most genes have multiple copies that belong to At and Dt subgenomes in upland cotton. Different genome editing tools are desirable for the functional genomics research in cotton: a plant species with very complex genome (AADD,  $2n=4x=52$ ; genome size of 2.5 Gb). Recently, we successfully established an efficient and precise CRISPR-Cas9 system in cotton for functional genomics research. In addition, we established an efficient CRISPR/Cpf1 system to expand the scope of genome editing in cotton for the first time with a very high efficiency (87%) and no off-target effects were detected in the most potential off-target sites. More importantly, the mutated phenotype and genotype in T0 generation were faithfully inherited to their progeny and some homozygous mutants were obtained in T1 generation. In addition to Cpf1 (Cas 12a), C2c1 (Cas 12b) also belongs to the second class of CRISPR proteins, which has not been tested for plant genome editing yet. In this experiment, CRISPR/C2c1 vectors were constructed and an endogenous gene of GhCLA was selected as a target genes. Since C2c1 is a heat-induced endonuclease, different heat treatments were tested during cotton tissue culture and genetic transformation in order to explore the optimal temperature and time for genome editing. For the CRISPR/C2c1 system, 451 independent T0 plants were obtained from C2c1-CLA vector after the genetic transformation mediated by Agrobacterium. By analyzing the Hi-Tom results, we found that the calli treated at 42°C, 45°C and 48°C could be edited by CRISPR/C2c1. Most editions were the DNA deletions of larger fragment along with some base substitutions in the target genes. We revealed that the calli treated at 45 ° for 4d showed the highest editing efficiency and negligible effect on the survival of the calli.

Most recently, a *G. hirsutum*-Base Editor 3 (GhBE3) base editing system was developed to create single base mutations in the genome of cotton. A cytidine deaminase sequence (APOBEC) fused with nCas9 and uracil glycosylase inhibitor (UGI) was inserted into our CRISPR/Cas9 plasmid (pRGEB32-GhU6.7). Three target sites were chosen for two target genes, GhCLA and GhPEBP, to test the efficiency and accuracy of GhBE3 with the editing efficiency ranging from 26.67 to 57.78%. Targeted deep sequencing revealed that the C→T substitution efficiency within an 'editing window', approximately six-nucleotide windows of-17 to-12 bp from the PAM sequence, was up to 18.63% of the total sequences. Furthermore, whole-genome sequencing analyses on two GhCLA-edited and one wild-type plants with about 100x depth showed that no bona fide off-target mutations were detectable from 1500 predicted potential off-target sites across the genome. In addition, the edited bases were inherited to T1 progeny. These results demonstrate that GhBE3 has high specificity and accuracy for the generation of targeted point mutations in allotetraploid cotton.

Overall, we established a comprehensive genome editing system including CRISPR-Cas9, CRISPR-Cpf1, CRISPR-C2C1 as well as the base editor in cotton with high efficiency and accuracy, which builds a solid foundation for cotton functional genomics research in the future.

### **W643: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

#### **Phytomedomics and Nutriomics for Health Security**

**Chittaranjan Kole**, ICAR-National Institute for Plant Biotechnology, New Delhi, India

The incidence rate of several deadly diseases, particularly cancer and diabetes, is highly alarming. According to the WHO-IARC report, the global cancer burden is estimated to have risen to 18.1 million new cases and 9.6 million deaths in 2018. According to the IDF projections, approximately 425 million adults were living with diabetes in 2017. Utilization of medicinal plants and nutraceutical crops are the potential options for alternative and complimentary medicines to mitigate these problems. According to the Zion Market Research, the global herbal supplement market is expected to reach approximately USD 86.74 billion by 2022, growing at a CAGR of around 6.8% between 2017 and 2022. Presently, plant-based drugs contribute 50% to the clinical drugs. This commercial importance coupled with severe prevalence of the deadly diseases underscores the need for the generation of genetic, genomics and breeding resources in medicinal plants and functional food crops. The genome sequences of a large number of medicinal plants such as Bitter Gourd, Calotropis, Cannabis, Catharanthus, Ginkgo, Ginseng, Neem, Ocimum, Salvia, etc. are already available. Similarly, genome sequencing of a number of nutraceutical crops specifically among millets, oilseeds, fruits, nuts, and vegetables, etc. have been accomplished. Achievements made so far in the fields of genetics, structural and functional genomics, and 'improvised breeding' in the major health-related plants (HRPs) will be enumerated and the future road map with special focus on cloning of therapeutic (TH) genes/QTLs followed by their utilization through cisgenesis and molecular pharming will be depicted.

### **W644: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

#### **Diversity in Metabolites and Fruit Quality Traits in Blueberry Enables Ploidy and Species Differentiation and Establishes a Strategy for Genetic Studies on Bioactive Traits**

**Massimo Iorizzo**<sup>1</sup>, Molla F. Mengist<sup>2</sup>, Mary H. Grace<sup>2</sup>, Jia Xiong<sup>2</sup>, Colin D. Kay<sup>2</sup>, Nahla Bassil<sup>3</sup>, Kim Hummer<sup>3</sup>, Mario Ferruzzi<sup>2</sup> and Mary Ann Lila<sup>4</sup>, (1)Plants for Human Health Institute, Department of Horticultural Science, North Carolina State University, Kannapolis, NC, Madison, WI, (2)Plants for Human Health Institute, North Carolina State University, NCRC, 600 Laureate Way, Kannapolis, NC, 28081, USA, (3)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (4)Plants for Human Health Institute, Kannapolis, NC

Blueberry is well recognized as a rich source of health promoting phytochemicals such as flavonoids and phenolic acids. Despite the important roles blueberries have on health effects, information is limited about the levels of variation in bioactive compounds within and between ploidy level and species, and their association with fruit quality traits. Such information is crucial to elucidate the genetic mechanisms

controlling biosynthesis of these compounds in blueberry. Hence, our objective was to evaluate 33 phytochemicals belonging to four major groups of flavonoids and phenolic acids across 128 blueberry accessions over two years together with fruit quality traits, including fruit weight, titratable acidity, total soluble acids and pH. Highly significant variation among accessions, years, and accession by year interaction were identified for most of the traits. Broad sense heritability of traits ranged from 20% to 90%, with most traits revealing moderate to high broad sense heritability ( $H^2 > 40\%$ ). Cluster analysis grouped phytochemicals by their functional structure (eg. anthocyanins, flavanols and flavonols). Fruit weight showed a negative correlation with most of the metabolites. Multivariate analysis of the traits resulted in separation of diploid, tetraploid and hexaploid accessions, indicating that each ploidy group has a distinct metabolite profile and a discrete set of fruit quality traits. Overall, traits with high heritability were greatly discriminative, indicating that genotypic effects explain the extensive bioactive and fruit quality trait diversity identified within and between ploidy groups. These results provide a framework to uncover the genetic basis of bioactive and fruit quality traits and will be useful to advance blueberry-breeding programs focusing on these traits.

#### **W645: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

##### **The Bitter Gourd (*Momordica charantia*) Genome Reveals the Genomic Architecture of Domestication**

Hideo Matsumura<sup>1</sup>, Min-Chien Hsiao<sup>2</sup>, Atsushi Toyoda<sup>3</sup>, Naoki Taniai<sup>4</sup>, Kazuhiko Tarora<sup>4</sup>, Naoya Uraski<sup>4</sup>, Shashi S Anand<sup>2</sup>, Narinder Dhillon<sup>5</sup>, Roland Schafleitner<sup>6</sup> and **Cheng-Ruei Lee**<sup>2</sup>, (1)Shinshu University, Ueda, Japan, (2)National Taiwan University, Taipei, Taiwan, (3)Advanced Genomics Center, NIG, ROIS, Shizuoka, Japan, (4)Okinawa Prefectural Agricultural Research Center, Itoman, Japan, (5)AVRDC - Regional Office for East and Southeast Asia, Nakhon Pathom, Taiwan, (6)World Vegetable Center, Tainan, Taiwan

While many studies have investigated the process of plant domestication, most "classic" cases, such as rice and maize, focused on artificial directional selection on novel mutations on Mendelian genes, leaving strong signatures of detectable selective sweeps in the genome. Here we report the long-read genome assembly of bitter gourd (*Momordica charantia*), an Asian vegetable with medical value, containing substances with potential antidiabetic effects. We obtained high contig N50 (close to 10 Mb) and proportion of sequences placed on chromosomes (96%). The well-assembled repetitive region allowed us to investigate the distribution of repetitive elements throughout the genome, which was not possible using short-read technologies. Population genomics revealed the wild and cultivar groups separated at about 6000 years ago, followed by the separation of cultivars into the South Asia and Southeast Asia groups at about 800 years ago. Genome-wide association study (GWAS) revealed the polygenic architecture of traits associated with fruit characteristics, and one important GWAS peak for fruit size resides within a region missing from a previous short-read based reference genome. We also identified a chromosomal region with low overall divergence between but high polymorphism within wild and cultivar groups. This region colocalizes with a known locus controlling flower sex ratios, providing one of the rare example where strong artificial selection for gynoecey could not overwhelm the forces of balancing selection in nature maintaining proper sex ratio.

#### **W646: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

##### **Harness Carotenoid Metabolism for the Development of Nutrient-Enriched Crops**

Hongbo Cao, Cornell University and **Li Li**, Cornell University, Ithaca, NY

Carotenoids are important natural products to humans as they are both provitamin A compounds and dietary antioxidants. Despite significant progress on the understanding of carotenoid metabolism, factors that affect the final carotenoid content in crops remain to be fully understood. Phytoene synthase (PSY) is widely accepted as a major flux-controlling enzyme in the carotenoid biosynthetic pathway and its activity defines carotenoid pool size in crops. Thus, PSY is used extensively for metabolic engineering of carotenoids in crops. Previously, we discover that post-translational regulation of carotenoid biosynthesis

includes PSY enzyme stabilization by OR proteins and degradation by Clp protease to maintain carotenogenic enzyme proteostasis in modulating carotenoid biosynthesis in plants. Our recent study of two tomato PSY isoforms, SIPSY1 and SIPSY2, reveals their activity divergence. We found that the fruit-specific PSY1 was less effective in promoting carotenoid biosynthesis than the green tissue-specific PSY2. Through PSY protein 3-D structure modeling and site-directed mutagenesis analysis, we identified the key amino acid residues responsible for high PSY enzyme activity. By examining the evolutionary features of these key residues, it highlights the potential to rationally design the rate-limiting enzyme for targeted metabolic engineering or breeding of carotenoid-enriched crops.

#### **W647: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

##### **Whole Genome Sequencing and Genetic Diversity Analysis in *Atractylodes lancea* and *Ephedra sinica***

Hajime Hiyama<sup>1</sup>, Takahiro Tsusaka<sup>1</sup>, Kenta Shirasawa<sup>2</sup>, Hideki Hirakawa<sup>3</sup> and **Sachiko Iso**<sup>2</sup>, (1)Tsumura & Co., Inashiki, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Japan, (3)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan

*Atractylodes lancea* and *Ephedra sinica* has been used as medical plants especially in East Asia. In order to investigate genome wide diversity analysis and breeding, reference genomes have been tried to construct by using Illumina and PacBio sequences. Genome sizes of *A. lancea* and *E. sinica* were estimated as 4.8Gb and 15Gb, respectively, based on kmer frequency analysis using Illumina reads. A total of 34 cell PacBio (Sequel) sequences were assembled to 31,976 primary sequences by FALCON unzip with total length of 6.87 Gb. Assembled sequences were polished by PacBio reads by Arrow, and complete BUSCOs increased from 82.7% to 93.6%. Meanwhile, Illumina scaffolds were constructed with paired-end and mate pare reads by SOAP denovo2, and further scaffolding was performed with Pacbio reads by OPERA-LG. The total length of scaffolds was 30.2Gb consisting with 50,074,622 sequences.

The assembled sequences in *A. lancea* and *E. sinica* were used as reference genomes in variant call in genetic diversity analysis of the two species. dd-RAD-Seq reads identified candidate 751,463 variants within 288 *A. lancea* accessions. Meanwhile, dd-RAD-Seq and GRAS-Di reads were obtained in 192 *E. sinica* accessions, suggested GRAS-Di was more robust system for variant call in species with large genome size. The candidate 836,702 and 531,320 variants identified with GRAS-Di and dd-RAD-Seq reads, respectively, reduced 15,567 and 1,602 variants when they were filtered with max-missing = 0.8. Both *A. lancea* and *E. sinica* showed highly heritability in medical compounds within the tested accessions. It is expected that the constructed reference genomes contribute molecular breeding in the two medical plants.

#### **W648: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

##### **Achievements and Prospects of Breeding for Enhanced Nutritional Quality in Cereal and Legumes Crops in West Asia and North Africa By Icarda**

**Michael Baum**<sup>1</sup>, Shiv Kumar Agrawal<sup>1</sup>, Aladdin Hamwiche<sup>2</sup>, Fuad Maalouf<sup>3</sup>, Sripada Udupa<sup>1</sup>, Ashutosh Sarker<sup>4</sup>, Wuletaw Tadesse<sup>1</sup>, Filippo Maria Bassi<sup>1</sup>, Miguel Sanchez-Garcia<sup>5</sup>, Andrea Visoni<sup>1</sup>, Mustapha El-Bouhssini<sup>5</sup>, Seid Ahmed<sup>1</sup>, Mariana Yazbek<sup>3</sup>, Zakaria Kehel<sup>5</sup> and Ahmed Amri<sup>1</sup>, (1)ICARDA, Rabat, Morocco, (2)ICARDA, Cairo, Egypt, (3)ICARDA, Lebanon, (4)ICARDA, India, (5)ICARDA, Morocco

Cereals and food legumes are the mainstay of healthy agri-food systems as they complement each other in providing protein of high biological value, rich in micronutrients such as Fe, Zn, Se, etc., prebiotics, and other bioactive compounds for a healthy and nutritious diets. The role of wheat, barley and food legumes in alleviating the hidden hunger caused by micronutrient deficiencies widely prevalent among two billion people, mainly in South Asia and Sub-Saharan Africa has widely been recognized. Following the health claims of beta glucans in barley, research is undertaken by ICARDA to develop varieties with high beta glucans and high micro-nutrients from interspecific crosses with *Hordeum spontaneum*. ICARDA is also developing durum wheat germplasm with carotenoids. Current evidence shows that dietary legumes are associated with reduced cardiovascular diseases, hypertension, and obesity

incidence (Viguiliouk *et al.* 2019). Recently, there is a drive to incorporate legumes as part of a cereal based diet to improve cardiometabolic health as their consumption may reduce cholesterol, support weight management via glycaemic responses and aid digestive health (Mudryj *et al.* 2014, Tilman *et al.* 2015). Some legumes like grass pea are also known for the presence of a plant toxin called ODAP which, if taken in large quantity for a long time, causes lathyrism, a paralysis of lower limb. Past studies have revealed that large genetic variation exists for Fe, Zn, and Se in wheat, barley, lentil, chickpea, faba bean and grass pea, for ODAP content in grass pea and tannin content in faba bean and offer scope for genetic improvement. Recent advances in genomic tools and technologies have facilitated generation of large-scale sequencing and genotyping datasets in these crops. Combined analysis of high-resolution phenotypic and genetic information is paving the way towards identification of genes/QTLs and biological pathways associated with these nutritional and antinutritional traits. This presentation highlights the current efforts made towards the genetic improvement in cereals and food legumes using classical breeding and modern genomics.

#### **W649: International Phytomedomics and Nutriomics Consortium (ICPN) 1**

##### **Plant Virus Expression Vectors for the Treatment of Some of the Most Challenging Diseases Known Today**

**Kathleen Hefferon**, University of Toronto, Toronto, ON, Canada

Plant Virus Nanoparticles: New Applications for Developing Countries

##### **Abstract:**

For over two decades now, plants have been explored for their potential to act as production platforms for biopharmaceuticals, such as vaccines and monoclonal antibodies. Without a doubt, the development of plant viruses as expression vectors for pharmaceutical production have played an integral role in the emergence of plants as inexpensive and facile systems for the generation of therapeutic proteins. More recently, plant viruses have been designed as non-toxic nanoparticles which can target a variety of cancers and thus empower the immune system to slow or even reverse tumor progression. The following presentation describes the employment of plant virus expression vectors for the treatment of some of the most challenging diseases known today. The presentation concludes with a projection of the multiple avenues by which virus nanoparticles could impact developing countries.

#### **W650: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

##### **A Genome Wide Association Approach Combined with Quantitative Trait Loci Analysis Reveals the Genetic Architecture of Glucosinolate Biosynthesis in *Brassica rapa* L. Leaves**

Xiaonan Li, Shenyang Agricultural University, Shenyang, China and **Yong Pyo Lim**, Chungnam National University, Daejeon, South Korea

Glucosinolates (GSLs) are sulfur-rich secondary metabolites, synthesized predominantly in the Brassicaceae, that play important roles in human health, defense against pathogens and insects, and flavor quality. The genetic architecture of GSL biosynthesis in *Brassica rapa* L. is poorly understood. We performed a conventional QTL analysis combined with a genome wide association approach to identify genomic regions and genes that regulate glucosinolate biosynthesis in *B. rapa* under two environments. Four consensus QTL clusters were identified for various glucosinolate compounds. Both QTL and association mapping revealed that paralogous *R2R3 MYB* transcription factors, *MAM1*, and *AOP2* gene family members on chromosomes A02, A03, A04, A07 and the A09 syntenic regions were associated with different glucosinolate indicating their functional divergence in regulating glucosinolate biosynthesis.

In addition, A common SNP of the *MAM1a* paralog anchored to A02 was significantly associated with aliphatic and indolic glucosinolates in a region no QTL was detected. Two SNPs of *MYB28a* formed four haplotypes that associated with gluconapin, aliphatic and total glucosinolate. Variation in expression of *GSL-OHa*, rather than sequence variation, explained the QTL associated with gluconapin and progoitrin accumulation in *B. rapa*. The promoter sequence analysis among natural population demonstrated that the lack of MYB binding site (MBS) in oil-type *B. rapa* possibly repressed the expression of *GSL-OHa*. This study demonstrated that genome wide association analysis is a useful complementary approach to traditional bi-parental QTL mapping for dissection of the genetics underlying glucosinolate variation in *B. rapa*, and to facilitate further molecular mechanism studies of glucosinolate biosynthesis.

## **W651: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

### **Haplotype Networking of GWAS Hits for Citrulline Variation Associated with the Domestication of Watermelon**

**Umesh K. Reddy**<sup>1</sup>, Joshi Vijay<sup>2</sup>, Suhas Shinde<sup>1</sup>, Padma Nimmakayala<sup>1</sup>, Lakshmi V. Abburi<sup>1</sup>, Suresh Alaparathi<sup>1</sup>, Carlos Ortiz<sup>1</sup>, Amnon Levi<sup>3</sup> and Girish Panicker<sup>4</sup>, (1)Department of Biology, West Virginia State University, Institute, WV, (2)Department of Horticultural Sciences, Texas A&M University, Uvalde, TX, (3)USDA, ARS, U.S. Vegetable Laboratory, Charleston, SC, (4)Center for Conservation Research, Alcorn State University, Lorman, MS

Watermelon is a good source of citrulline, a non-protein amino acid. Citrulline has several therapeutic and clinical implications as it produces nitric oxide via arginine. In plants, citrulline plays a pivotal role in nitrogen transport and osmoprotection. The purpose of this study was to identify single nucleotide polymorphism (SNP) markers associated with citrulline metabolism using a genome-wide association study (GWAS) and understand the role of citrulline in watermelon domestication. A watermelon collection consisting of 187 wild, landraces, and cultivated accessions was used to estimate citrulline content. An association analysis involved a total of 12,125 SNPs with a minor allele frequency (MAF) >0.05 in understanding the population structure and phylogeny in light of citrulline accumulation. Wild egusi types and landraces contained low to medium citrulline content, whereas cultivars had higher content, which suggests that obtaining higher content of citrulline is a domesticated trait. GWAS analysis identified candidate genes (*ferrochelatase* and *acetolactate synthase*) showing a significant association of SNPs with citrulline content. Haplotype networking indicated positive selection from wild to domesticated watermelon. To our knowledge, this is the first study showing genetic regulation of citrulline variation in plants by using a GWAS strategy. These results provide new insights into the citrulline metabolism in plants and the possibility of incorporating high citrulline as a trait in watermelon breeding programs.

## **W652: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

### **Progress and Applications of the 1000 Medicinal Plant Genomes Project (1KMPG)**

**Chang Liu**, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences&Peking Union Medical College, Beijing, China

**Background:** DNA barcoding technology has been widely used in research and real-world applications and has played a pivotal role in ensuring quality and safety of botanical materials. With the maturity of the single-marker based technology and the wide adoption of the next- and third-generation DNA sequencing technologies, it is time to consider extending the single-marker based identification to a genomic marker based approach. Since March 2017, the Institute of Medicinal Plant Development and Illumina Company have signed a collaboration agreement to start a “1000 Medicinal Plant Genomes” project. The initial strategy is to use a genome skimming strategy to obtain the plastome sequences of 1000 Medicinal Plant Genomes in three years. Here, we would like to describe the progress of this project. **Results:** We have collected and annotated ~500 medicinal plants species. The genomic DNA of ~500 medicinal plant species has been sequenced. The corresponding plastome sequences have been assembled and annotated. Several applications, such as bioinformatic tools for simultaneous determination of multiple components from biological mixtures, a block-chain based tracking system and an intellectual property registration system, have been developed. On the other hand, we have also run

into several problems, such as ensuring the quality of the voucher samples and fair intellectual property sharing. **Significance:** As the success of this project will require the participation of researchers, industrialists, and governmental officials alike around the globe, we would like to call for a close collaboration among all practitioners involved in using genomic approaches to study medicinal plants.

## **W653: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

### **Genome Evolution and Regulation of Oil Biosynthesis of a Model Oilseed Crop Sesame**

**Hongmei Miao**, Henan Sesame Reserach Center, Henan Academy of Agricultural Sciences, Zhengzhou, China and Haiyang Zhang, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, China, Zhengzhou, China

Sesame (*Sesamum indicum* L.,  $2n=26$ ) is an important and specific oilseed crop in the world. Here we report the chromosome-scale reference genome assembly for sesame, based on the complicated sequencing platforms and assembly strategies with a super high-density SNP genetic map and a high-density BAC-FISH cytogenetic map. The genome resources revealed the ancient evolution position and the evolution diversity of the *Sesamum* species at the Asterid branch. Chromosome translocation, as well as chromosome fission and fusion were detected and reflected the genome structure characters of *Sesamum*. Massive expansions in dozens gene families regulating fatty acid synthesis and metabolism were found in the cultivated sesame, compared with the wild species. GWAS analyses of oil content trait in 560 sesame core germplasm were also performed, in order to reveal the regulation of oil biosynthesis in sesame. Results indicated that specific gene family expansion and key gene diversity realted with the oil biosynthesis pathways contributed to the high oil content in sesame. The accurate genome of sesame will facilitate further genetic analysis of seed quality traits in sesame and accelerate the sesame variety improvement.

## **W654: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

### **A Whole Genome Association Study in *Salvia miltiorrhiza***

**Mei Jiang** and Chang Liu, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences&Peking Union Medical College, Beijing, China

*Salvia miltiorrhiza* is one of the most commonly used traditional Chinese medicine and recognized as a model medicinal plant, which has remarkable curative effect in the treatment of cardiovascular and cerebrovascular diseases and anti-oxidation. Plants produce a variety of metabolites having a critical role in treating disease due to the genomic variation. Here, we produced a whole genome association study of 383 *S. miltiorrhiza* accessions based on the metabolome. This project was composed of five parts. First, these diverse accessions were collected all over China and grown in the same place, each accession obtained multiple plants by root propagation. We collected leaves to extract DNA and root from three different plants per accession to detect metabolome. Second, we resequenced the whole genome of 383 *S. miltiorrhiza* accessions and obtained average 5 GB raw data for each accession, which comprised about 10-fold coverage of the *S. miltiorrhiza* genome. A total of about 8 million SNPs were obtained by both SAMtools and Genome Analysis Toolkit. The genetic structure of the 383 accessions was analyzed through the population structure and principal component analysis, showing these accessions were grouped into two distinct groups. Third, we obtained the metabolic profile of *S. miltiorrhiza* by using a liquid chromatography-tandem mass spectrometry (LC-MS/MS) based method. Hundreds of metabolites were detected, containing the tanshinone, salvianolic acids and compounds in their biosynthesis pathways which are the primarily medicinal constituents. In addition, the population structure was analyzed based on the metabolom. Fourth, a metabolic genome-wide association study was conducted, which obtained hundreds of common variants influencing numerous metabolites. To identify candidate genes related to tanshinone biosynthesis that have not been identified previously, we will look for the protein cluster that is related to the associated metabolic trait encoded at these loci and perform clusters analysis of candidate gene relative to homologous genes with known function. Finally, validating candidate genes by detecting the metabolites of plants were overexpressed and inhibited candidate genes. Our study provides insights into the genetic base of *S. miltiorrhiza* metabolome

variation and can facilitate the analysis of the tanshinone biosynthesis pathway and the selection of elite traits of *S. miltiorrhiza*.

## **W655: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

### **Towards Genomics-Enabled Breeding in Pearl Millet: a Climate Resilient and Nutrient Dense Cereal Crop**

**Desalegn D. Serba**<sup>1</sup>, Ghislain Kanfany<sup>2</sup>, M. Govindaraj<sup>3</sup>, Kebede T Muleta<sup>4</sup>, Paul St Amand<sup>5</sup>, Amy N. Bernardo<sup>6</sup>, Davina Rhodes<sup>7</sup> and Guihua Bai<sup>4</sup>, (1)Kansas State University, Hays, KS, (2)Institut Sénégalais de Recherches Agricoles (ISRA), Senegal, (3)ICRISAT, India, (4)Department of Agronomy, Kansas State University, Manhattan, KS, (5)USDA-ARS-Small Grain Genotyping Lab, Manhattan, KS, (6)Hard Winter Wheat Genetics Research Unit, USDA-ARS, Manhattan, KS, (7)Kansas State University, Manhattan, KS

Pearl millet [*Pennisetum glaucum* (L.)] is a climate resilient cereal crop, essential for food and nutritional security of more than 100 million people in the arid and semi-arid areas of Africa and India. However, its genetic improvement is lagging behind other major cereals. With the recent advances in next-generation genotyping technologies and the release of the pearl millet genome sequence, genomics-assisted breeding has the potential to accelerate its genetic improvement for agronomic and nutritional traits. With the aim of developing molecular resources for pearl millet breeding and population genomics studies, we characterized a large globally-sourced panel of pearl millet germplasm and breeding lines using genotyping-by-sequencing. High genetic diversity and population structure mostly overlapping with geographic origin or pedigree relationship of the populations were detected. Prevalence of rapid LD decay, which can be attributable to the long history of recombination among landraces, was observed. The results indicate that migration to various agroclimatic conditions had a significant evolutionary role in the formation of distinct and divergent forms of pearl millet. Selective sweeps were observed in the genome regions where major QTLs for high iron (Fe) and zinc (Zn) were colocalized. Higher levels of micronutrients (30-120 ppm Fe and 20-90 ppm Zn) and flowering time temperature tolerance (up to 42°C) will set the base for progressing toward identification of genetic sources for breeding highly resilient varieties to increase productivity and nutritional quality.

## **W656: International Phytomedomics and Nutriomics Consortium (ICPN) 2**

### **Concomitant Phytonutrient and Transcriptome Analysis of Mature Fruit and Leaf Tissues of Tomato Grown using Organic and Conventional Fertilizer**

Richard M. Sharpe<sup>1</sup>, Luke Gustafson<sup>2</sup>, Seanna Hewitt<sup>3</sup>, Benjamin Kilian<sup>3</sup>, James Crabb<sup>4</sup>, Christopher Hendrickson<sup>5</sup>, Derick Jiwan<sup>6</sup>, Preston Andrews<sup>3</sup> and **Amit Dhingra**<sup>7</sup>, (1)Department of Horticulture, Washington State University, Pullman, WA, (2)University of Maryland Extension, (3)Washington State University, Pullman, WA, (4)Department of Horticulture, Washington State University, (5)National University, San Diego, CA, (6)Hazel Technologies, (7)Molecular Plant Sciences, Washington State University, Pullman, WA

Enhanced levels of antioxidants, phenolic compounds, carotenoids and vitamin C have been reported for several crops grown under organic fertilizer, albeit with yield penalties. As organic agricultural practices continue to grow and find favor it is critical to gain an understanding of the molecular underpinnings of the factors that limit the yields in organically farmed crops. Concomitant phytochemical and transcriptomic analysis was performed on mature fruit and leaf tissues derived from *Solanum lycopersicum* L. 'Oregon Spring' grown under organic and conventional fertilizer conditions to evaluate the following hypotheses. 1. Organic soil fertilizer management results in greater allocation of photosynthetically derived resources to the synthesis of secondary metabolites than to plant growth, and 2. Genes involved in changes in the accumulation of phytonutrients under organic fertilizer regime will exhibit differential expression, and that the growth under different fertilizer treatments will elicit a differential response from the tomato genome. Both these hypotheses were supported, suggesting an adjustment of the metabolic and genomic activity of the plant in response to different fertilizers. Organic fertilizer treatment showed an activation of photoinhibitory processes through differential activation of

nitrogen transport and assimilation genes resulting in higher accumulation of phytonutrients. This information can be used to identify alleles for breeding crops that allow for efficient utilization of organic inputs.

## **W657: International Sheep Genomics Consortium**

### **International Sheep Genomics Consortium Update**

**Shannon M. Clarke**, AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand

The long-term goals of the International Sheep Genomics Consortium (ISGC) to develop underpinning resources for the sheep research community has resulted in continued improvement of the sheep genome assembly and development of low, medium and high-density Illumina SNP chips. The ISGC members have continued to make available whole genome sequence data to the community that has been captured via the Sheep Genomes Database, an initiative of the ISGC that extends the consortiums recent achievements. SheepGenomesDB is an electronic warehouse containing sequence variants called from the expanding collection of sheep genomes. Through the application of a single harmonised pipeline for read QC, mapping, variant detection and annotation, SheepGenomesDB makes available variant collections derived in a standardised manner. Run 2 has seen ~1000 animals analysed with variant collections positioned on the OAR V3.1. The consortium is now in the process of Run 3 that will include an additional ~300 animals utilising Rambouillet v1 genome assembly with the aim of providing users with tools to obtain variants defined by chromosomal location, SNP annotation results or via animals and breeds of interest. An update of the ISGC's activities will be presented.

## **W658: International Sheep Genomics Consortium**

### **An Update of the Ovine FAANG Project**

**Brenda M. Murdoch**<sup>1</sup>, Stephen N. White<sup>2</sup>, Michelle R. Mousel<sup>2</sup>, Alisha Massa<sup>3</sup>, Kim C. Worley<sup>4</sup>, Alan L. Archibald<sup>5</sup>, Emily L. Clark<sup>5</sup>, Brian Dalrymple<sup>6</sup>, James W. Kijas<sup>7</sup>, Shannon M. Clarke<sup>8</sup>, Rudiger Brauning<sup>9</sup>, Timothy P.L. Smith<sup>10</sup>, Tracy Hadfield<sup>11</sup> and Noelle Cockett<sup>11</sup>, (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (3)Washington State University, (4)Baylor College of Medicine, Houston, TX, (5)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (6)Institute of Agriculture, The University of Western Australia, CRAWLEY, Australia, (7)CSIRO Animal, Health and Food Science, St Lucia, Australia, (8)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand, (9)AgResearch Limited, Invermay Agricultural Centre, Mosgiel, New Zealand, (10)USDA, ARS, USMARC, Clay Center, NE, (11)Utah State University, Logan, UT

The goal of the Ovine FAANG project is to facilitate a greater understanding of the complex nature of gene regulation within this globally important food and fiber species. Compilations of functional assays have been used to characterize and define the multifaceted biological mechanisms that contribute to gene regulation. Specifically, we examined coding and non-coding transcript isoforms, alternative splicing, promoters and cis-acting regulatory elements, open chromatin, histone modifications, and DNA methylation across a wide range of sheep tissues. To date we have mRNA short read sequencing data on 60 tissues, long read data on 8 tissues, and micro RNA on 30 tissues. To complement this gene expression data, cap analysis of gene expression (CAGE) data for 56 of these same tissues was used to confirm transcription start sites across the ovine genome. Histone modification sequencing results using H3K4me3, H3K27ac, H3K4me1 and H3K27me3 marks are currently being analyzed for a subset of 47 of these tissues. Other assays currently underway include ATAC-Seq assays to assess chromatin accessibility, and whole genome and reduced representation bisulfite sequencing to determine DNA methylation status. Overall, this project will provide one of the highest resolution annotations of the reference genome of a livestock species. The resources we have generated are foundational. More specifically, this expands our understanding of how gene-regulation controls phenotypic plasticity in this economically important livestock species.

## **W659: International Sheep Genomics Consortium**

### **High-Quality Sheep Genome Assemblies for White Dorper and Romanov Breeds By F1 Trio Binning**

**Shavahn C. Loux**, University of Kentucky, Lexington, KY, Benjamin D. Rosen, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, Michael P. Heaton, USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, Brad Freking, US Meat Animal Research Center, Brian L. Vander Ley, University of Nebraska, Lincoln, Clay Center, NE, Derek M. Bickhart, Dairy Forage Research Center, USDA-ARS, Madison, WI, Shannon M. Clarke, AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand, Thomas Murphy, ARS-USDA, Ted Kalbfleisch, University of Louisville, Louisville, KY and Timothy P.L. Smith, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE

Trio-binning is a method of genome assembly that uses heterozygosity between parental alleles to separate long sequence reads into bins by parental origin. This enables the assembly of two fully phased haploid genomes from a single individual. In this study, a White Dorper ram was crossed with a Romanov ewe, as these breeds form the foundation of a composite population displaying superior characteristics in reproduction, health and welfare while maintaining adequate meat quality. Additionally, this cross has previously exhibited high levels of heterosis. Short-read data were generated from both the sire and dam and used to define parent-specific *kmers*. The list of *kmers* was then used by the Canu assembler to sort long-reads generated from the F<sub>1</sub> offspring into parent-of-origin bins. Genome assemblies of the maternal and paternal bins were performed both by Flye (a graph-based assembler) and Canu (an overlap-layout-consensus assembler). The NG50 averaged 49.8 and 51.2 Mb for the White Dorper and Romanov assemblies, respectively, with the largest contigs spanning over 150 Mb for both. Contig N50s increased from 2.6 Mb in the Oar\_rambouillet\_v1.0 to 63 Mb in both of the new assemblies, with a significant reduction in scaffold numbers as well. Our preliminary results confirm the feasibility of using trio binned reads for the creation of an accurate haploid genome for two distantly related individuals within the same species and provides a preview of the genome assemblies in progress for the White Dorper and Romanov sheep breeds.

\*Shavahn Loux and Ben Rosen contributed equally to this work

## **W660: International Sheep Genomics Consortium**

### **Genomic Predictions for Enteric Methane Production are improved by Metabolome and Microbiome Data in Sheep (*Ovis aries*)**

**Elizabeth Ross**, Queensland Alliance For Agriculture and Food Innovation - University of Queensland, Brisbane, QLD, Australia

Methane production from rumen methanogenesis contributes approximately 65% of greenhouse gas emissions from the agricultural sector. This study has performed genomic predictions for methane production from 99 sheep (*Ovis aries*) across three years using a residual methane phenotype that is corrected for live weight, rumen volume, and feed intake. Prediction accuracies ranged from 0.058 to 0.220 depending on the time point being predicted. When prediction accuracy was corrected based on a published heritability of 0.18, the prediction accuracies were between 0.138 and 0.519. The BLUP prediction algorithm was then applied to relationships between animals that were built on the rumen metabolite and microbiome of the animals. Prediction accuracies for the metabolome based relationships were 0.254 and 0.132; and the prediction accuracy for the first microbiome time point was 0.142. The second microbiome time point could not successfully predict residual methane production. In the three cases where residual methane phenotypes were successfully predicted from the metabolome or the microbiome, the addition of the genomic relationships was able to predict residual methane production with a greater accuracy than either of the relationship matrices alone (0.274, 0.158 and 0.182 for the two metabolome and the first microbiome based relationships respectively).

## **W661: International Sheep Genomics Consortium**

### **Across-Population Genomic Predictions in Composite Sheep Populations with Similar Development History**

**Luiz Brito**, Purdue University, West Lafayette, IN

The predictive performance of genomic breeding values (GEBVs) is dependent on various parameters, including the size of training population and genetic relationship between training and prediction populations. Various Norwegian (NOR) and New Zealand (NZ) sheep populations were developed by crossing multiple breeds, with overlapping founder breeds. The main goal of this study was to describe the genetic diversity and connectedness between NOR and NZ sheep populations, and validate GEBVs predicted for NOR animals using the NZ training population. A total of 47,056 NZ and 828 NOR animals were genotyped with a HD (606K SNPs) panel. A PCA-plot based on the HD genotypes showed no clear evidence of diverging clustering among animals, suggesting a genetic similarity between populations from both countries. Estimated breeding values (EBVs) for weaning weight ( $WW_{NOR}$ ), carcass weight ( $CW_{NOR}$ ), EUROP carcass classification and EUROP fat grading were predicted for NOR animals. In addition, GEBVs for  $WW_{NZ}$ , pre-slaughter weight,  $CW_{NZ}$ , x-ray  $CW_{NZ}$ , carcass fatness at the GR site, ultrasound fat depth and ultrasound eye muscle depth were predicted for NOR animals using the NZ training population. EBVs and GEBVs for NOR animals were compared based on the Pearson correlation coefficient adjusted for the average EBV reliability. Pearson correlation coefficients ranged from 0.19 (between  $WW_{NOR}$  and  $WW_{NZ}$ ) to 0.40 (between  $CW_{NOR}$  and x-ray  $CW_{NZ}$ ). Considering the differences in trait definition between countries, these findings indicate a promising opportunity to perform across country genomic predictions in composite sheep populations.

## **W662: International Sheep Genomics Consortium**

### **The Effects of Paternal Nutrition on Traits of the Next Generation in Sheep**

**Hasan Khatib**, University of Wisconsin Madison, Madison, WI

The aim of this study was to evaluate the impact of pre-pubertal diet in Polypay rams on complex traits, DNA methylation, and transmission to offspring. A total of ten twin pairs of F0 rams were divided so that one ram was fed a control diet and the other was fed the control diet with added methionine. Diet was associated with earlier age at puberty in treatment vs. control rams. Additionally, treatment rams had altered weight compared to control rams. A total of ten F0 rams were bred and the F1 generation was fed a control diet. F1 rams showed a difference in weight and scrotal circumference (SC) at puberty, but not in age at puberty. The DNA methylation of F0 ram sperm was assessed, and genes related to both sexual development and growth were prevalent in the data. These results provide novel information about the mechanisms through which the pre-pubertal paternal diet may alter growth and sexual development.

## **W663: International Sheep Genomics Consortium**

### **Whole Genome Neutral and Adaptive Diversity in Traditional Sheep Breeds: Case of Morocco**

**Badr Benjelloun**<sup>1,2</sup>, Abdessamad Ouhrouch<sup>3,4</sup>, Simon Boitard<sup>5</sup>, Abdelmajid Haddioui<sup>3</sup>, Frédéric Boyer<sup>6</sup>, Bertrand Servin<sup>7</sup>, François Pompanon<sup>8</sup> and Anne Blondeau Da Silva<sup>9</sup>, (1)LECA, UGA-CNRS-USMB, France, (2)INRA-Morocco, Beni Mellal, Morocco, (3)USMS Beni Mellal, Morocco, (4)INRA-Morocco, Morocco, (5)GenPhySE, INRA-INPT-ENVT-Université de Toulouse, France, (6)LECA/CNRS, Grenoble, France, (7)GenPhySE, INRAE, ENVT, ENSAT, Toulouse, France, (8)LECA - Université Grenoble Alpes, Grenoble, France, (9)GAMAA Univ. Limoges, INRA, France

Genomic selection signatures are very useful for understanding how environmental and anthropic constraints impact genome variation and the distribution of domestic species. Sheep farming plays a key role and is one of the main sources of milk and meat in the world. Unlike cosmopolitan breeds, traditional populations represent interesting hotspot of diversity and the understanding of their distribution would help facing the impact of environmental changes. Here, we characterized neutral genome diversity and demographic history as well as intra and inter-population selection signatures in the main sheep breeds

reared in Morocco using their entire genomes. The complete genome data from 87 individuals representing five predominant local sheep breeds in Morocco were used to infer demographic history, which has made it possible to estimate the evolution of the effective population size over time. Two methods were used to investigate selection signatures: one to detect putative regions under selection within each of these breeds and the second to detect selection signatures that differentiate the breeds one from the other. We identified several hundreds of regions/genes under selection from the studied breeds. We highlighted several biological processes involved in local adaptation as well as those linked to zootechnical performances characterizing each breed. Findings of this study increased our understanding on how genetic diversity is distributed in local breeds.

#### **W664: International Sheep Genomics Consortium**

##### **Chasing Colors: Identifying the Genetic Variants Responsible for Coat Color Variation in Sheep**

**Christian J Posbergh**<sup>1</sup>, Elizabeth A. Staiger<sup>1,2</sup> and Heather J Huson<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)Auburn University, Auburn, AL

White wool is the dominant product in the commercial wool market and commands a higher price than non-white wool. However, non-white wool can bring significantly higher prices than white wool in the hand spinning wool market, particularly in the Northeastern United States. Several studies have examined the phenotypic variation and genetic inheritance patterns at specific loci including the Agouti, Extension, and Brown loci in sheep. However, there has been little work to determine the molecular cause of these different phenotypes. The objective of this study was to identify variants associated with various coat color variation observed within and across sheep breeds, such as dilution, red head and legs, and white spotting. Using Illumina 150 base pair paired-end reads we generated whole genome sequences (approximately 20x coverage) from 18 sheep across the Romeldale, Romney, Jacob, and California Red breeds that were selected to represent a variety of coat color patterns. Reads were aligned to the Oar\_v4.0 genome assembly using the Burrows-Wheeler aligner, and variants called following Genome Analysis Toolkit's "Best Practices" workflow. To date, only genes known to be associated with color variants in other species have been investigated. We have identified a variant responsible for the lilac dilution in Jacob sheep within *MLPH*, and a variant within *TYRP1* which is associated with the red hair color observed within the California Red and Tunis sheep. Work is still ongoing to identify variants around the *ASIP* region due to the wide variety in phenotypes attributed to the *ASIP* locus.

#### **W665: International Sheep Genomics Consortium**

##### **Whole Genome Sequence Data and Analysis from Wild Thinhorn Sheep**

**Sarah Helen Dias dos Santos**<sup>1</sup>, Si Lok<sup>2</sup>, Timothy N. H. Lau<sup>2</sup> and David W. Coltman<sup>1</sup>, (1)University of Alberta, Edmonton, AB, Canada, (2)The Centre for Applied Genomics, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, ON, Canada

Thinhorn sheep, the pure white Dall (*Ovis dalli dalli*) and the dark pelage Stone subspecies (*Ovis dalli stonei*), are Canadian icons. The two subspecies are believed to have arisen through vicariance at the last ice age. Due to admixture following the retreat of the ice sheets, their complex evolutionary relationship is not fully understood. To provide a genomic underpinning to study their relationship, our group is embarking on whole genome sequencing and de novo assembly of wild thinhorn sheep. Here, we present the first draft genome assembly of the Stone sheep generated by short-read sequencing combined with microfluidic partitioning on the 10X Genomics platform. The resulting highly contiguous assembly has scaffold N50 length of over 8 Mbp and total reconstruction of 2.6 Gbp, representing 87% of the genome estimated by K-mer analysis. Ninety-two percent of BUSCO mammalian gene set is represented in its entirety. We have also assembled the complete mitochondrial genomes for Dall and Stone sheep and have performed phylogenetic analysis with other *Ovis* species. Genome assembly of the Dall sheep along with improved draft assemblies of the Stone and intermediate "Fannin" sheep are in progress. When completed, we first aim to identify

the genomic determinants linking pelage colour and other notable traits, with the longer-term goal to reconstruct the comprehensive evolutionary and phylogenomic history of Canadian wild sheep and their admixture in a cultural, geographical and environmental perspective.

### **W666: International Wheat Genome Sequencing Consortium (IWGSC)**

#### **IWGSC Workshop Welcome**

**Kellye Eversole**, IWGSC, Lee's Summit, MO and Etienne Paux, INRA GDEC, Clermont-Ferrand, France

The IWGSC has led the effort to develop a high quality, reference sequence of the large, complex (hexaploid), bread wheat genome (IWGSC RefSeq v1.0 and annotation v1.1); and, as part of its Phase II activities, continues to coordinate projects to close gaps in the reference sequence (IWGSC RefSeq v2.0), deploy traditional and innovative projects to manually and functionally annotate the bread wheat genome, ensure the availability of high quality sequences of genomes which represent the full breadth of wheat diversity (landraces and elite lines), and develop new genomic resources that will permit the full exploitation of the reference sequence. In this presentation, a brief overview of these activities will be provided.

### **W667: International Wheat Genome Sequencing Consortium (IWGSC)**

#### **Improved Reference Genome Sequence of 'Chinese Spring' Wheat Provides Insights into Structural Changes of the Wheat Genome**

**Tingting Zhu**<sup>1</sup>, Le Wang<sup>1</sup>, Karin R. Deal<sup>1</sup>, Gabriel Keeble-Gagnere<sup>2</sup>, Rudi Appels<sup>2</sup>, Pat McGuire<sup>1</sup>, Jan Dvorak<sup>1</sup> and Ming-Cheng Luo<sup>1</sup>, (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)AgriBio, Centre for AgriBioscience, Department of Economic Development, La Trobe University, Bundoora, VIC, Australia

Bread wheat (*Triticum aestivum*) is globally the most extensively cultivated crop. Its genome is large (> 15 Gb) and consists of over 80% repeated sequences. The successful whole-genome-shotgun genome assembly of the bread wheat cv. 'Chinese Spring' (CS, IWGSC RefSeq v1.0) was a milestone for wheat genomics, genetics, and breeding. To improve this assembly, we constructed an optical map of CS using Bionano Direct Label and Stain (DLS) technology. This map, spanning 14.4 Gb, was used to detect and resolve errors in the IWGSC RefSeq v1.0 assembly, as well as to super-scaffold the sequences. In addition, we produced 545 Gb raw reads of CS with the PacBio's single-molecule real-time (SMRT) sequencing technology and assembled them into 12.9 Gb contigs. We then used the contigs to close gaps, which improved the contiguity of the assembly, and ultimately produced a revised assembly, IWGSC RefSeq v2.0. Compared to the pseudomolecules in the IWGSC RefSeq v1.0 assembly, the new pseudomolecules include 283 scaffolds (75.0 Mb) that had not been previously anchored, have new locations for 244 scaffolds (396.9 Mb), have 357 scaffolds (469.8 Mb) re-oriented, and have 65% fewer gaps. We deployed the v2.0 pseudomolecules along with the DLS optical maps of bread wheat progenitor species (*T. urartu*, *T. turgidum* ssp. *diccocoides*, and two *Aegilops tauschii* accessions), and investigated whole-genome structural variation (SV) events that took place during the evolution of bread wheat. We showed that the characteristics of SVs and evolution rates differ among bread wheat subgenomes. This work is parts of projects supported by US National Science Foundation grants IOS-1929053 and IOS-1238231.

### **W668: International Wheat Genome Sequencing Consortium (IWGSC)**

#### **Adaptive Introgression from Wild Emmer into Bread Wheat Revealed by Sequencing 1000 Wheat Exomes**

**Fei He**<sup>1</sup>, Raj Pasam<sup>2</sup>, Fan Shi<sup>3</sup>, Surya Kant<sup>4</sup>, Gabriel Keeble-Gagnère<sup>5</sup>, Pippa Kay<sup>2</sup>, Kerrie Forrest<sup>2</sup>, Allan Fritz<sup>6</sup>, Pierre J. Hucl<sup>7</sup>, Krystalee Wiebe<sup>8</sup>, Ron E. Knox<sup>9</sup>, Richard D. Cuthbert<sup>9</sup>, Curtis Pozniak<sup>7</sup>,

Alina Akhunova<sup>10</sup>, Peter L. Morrell<sup>11</sup>, John Davies<sup>12</sup>, Steve R. Webb<sup>13</sup>, German C. Spangenberg<sup>14</sup>, Ben Hayes<sup>15</sup>, Hans Daetwyler<sup>16</sup>, Josquin Tibbits<sup>5</sup>, Matthew Hayden<sup>17,18</sup> and Eduard Akhunov<sup>1</sup>, (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Australia, (3)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Australia, (4)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Horsham, Australia, (5)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia, (6)Department of Agronomy, Kansas State University, Manhattan, KS, (7)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (8)Crop Development Centre, University of Saskatchewan, Canada, (9)Swift Current Research and Development Center, Swift Current, SK, Canada, (10)Integrated Genomics Facility, Kansas State University, Manhattan, KS, (11)Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN, (12)Corteva Agriscience, Agriculture Division of DowDuPont, Indianapolis, IN, (13)Corteva Agriscience, Agriculture Division of DowDuPont, (14)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, Australia, (15)Queensland Alliance For Agriculture and Food Innovation, Bundoora, VIC, Australia, (16)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Australia, (17)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Vic, Australia, (18)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia

Introgression from wild relatives is a potential source of beneficial diversity. The contribution of historic introgression to adaptive evolution and improvement of wheat following its origin and dissemination across the world remains unknown. Here, we used the IWGSC RefSeq v.1.0 to generate a haplotype map including 7.3 million SNPs based on targeted re-sequencing of 890 diverse wheat landraces and cultivars, and tetraploid wild and domesticated relatives to identify historic introgression from wild emmer. On average, introgressed genomic regions (IGRs) comprised about 11.8% and 11.4% of genome per accession in landraces and cultivars. Genomic regions associated with improvement selection and environmental adaptation overlapped with 20.4% and 8% of detected IGRs, respectively, suggesting that wild emmer contributed beneficial alleles used during the development of locally adapted cultivars. We showed that improvement selection, environmental adaptation and introgression significantly reduced deleterious mutation burden in modern wheat. The overlap between the signals of gene flow, GWAS, improvement and environmental adaptation was suggestive of adaptive introgression. We demonstrated that alleles introduced from wild emmer explained substantial proportion of phenotypic variance for harvest weight, drought susceptibility and plant height, indicating that historic gene flow from the wild relative played an important role in shaping the agronomic phenotypes in modern wheat. A detailed map of genome-wide introgression developed in our study can guide targeted deployment of wild relative diversity in wheat breeding programs. These efforts, besides introducing novel adaptive variation available for selection, have potential to reduce mutation burden in the wheat genome and further accelerate breeding.

## **W669: International Wheat Genome Sequencing Consortium (IWGSC)**

### **Extending the Frontiers of Genomics-Assisted Breeding for Grain Yield, Stress-Resilience and Quality of Bread Wheat using the Reference Sequence**

**Philomin Juliana**, International Maize and Wheat Improvement Center (CIMMYT), El Batan, Mexico

The International Wheat Genome Sequencing Consortium's reference sequence (RefSeq v1.0) of the bread wheat genome has created exciting opportunities for genomics-assisted breeding in bread wheat. We have aligned 78,606 genotyping-by-sequencing markers for 44,624 lines from the International Maize and Wheat Improvement Center's (CIMMYT) stage 1 yield trials evaluated between 2014 and 2018 to the RefSeq v.1.0. The anchored markers filtered for different levels of missing data were used in genomic prediction models for a subset of 3,485 lines phenotyped for 35 key traits, to understand the impact of genomic coverage and training populations on the genomic predictabilities of traits. The genomic coverage associated with the filtered marker sets clearly showed a decreasing trend towards the centromeric regions with stringent filtering for missing data, and thereby served as ideal sets for evaluating the effect of genomic coverage on the predictabilities of traits. We observed similar accuracies at the different levels of genomic coverage, with the high-coverage marker subset providing an average

increase of only  $0.02 \pm 0.02$  in accuracy compared to the low-coverage subset, in both cross-validations and prediction across panels, across all traits, implying that once the genomic resolution has been reached in a crop with high linkage disequilibrium like wheat, marker number is no longer a critical limiting factor for prediction accuracies. The traits grain color, seedling and field resistance to stem rust, mixing time, alveograph W, flour sedimentation, loaf volume, protein content, and thousand kernel weight had the highest genomic prediction abilities (0.60–0.85), whereas all of the other traits were moderately predictable in within panel cross-validations. In predictions across panels, the traits with high predictabilities were seedling resistance to stem rust, grain color, mixing time, alveograph W, field resistance to stem rust and flour sedimentation with an average decrease of only  $0.07 \pm 0.05$  in accuracy from the corresponding cross-validation accuracies. However, we observed low across-panel predictabilities of traits like grain yield, phenology, Septoria tritici blotch etc. with an average decrease of  $0.20 \pm 0.06$  from the corresponding within-panel accuracies. We also performed genome-wide association studies (GWAS) to dissect the genetic architecture of 50 traits, evaluated in South Asia, Africa and the Americas, and anchored the significant markers to a reference wheat genotype–phenotype map, aligned to the RefSeq. This map highlights the application of the RefSeq as a platform for comparing and validating GWAS results and will also serve as a community resource providing opportunities for accelerating genomics-assisted wheat breeding through the targeted selection of desired regions. Furthermore, we generated the genomic fingerprints of 44,624 wheat lines comprising several key varieties cultivated worldwide for trait-associated markers that provide an important leap in understanding the genetic basis of traits in superior varieties. Finally, we examined the allele frequency dynamics for key trait-associated markers to characterize the role of selection in shaping patterns of allelic variation over time. We conclude that the RefSeq has shifted the paradigm of genomics-assisted breeding in wheat, with numerous benefits to CIMMYT and its partners globally.

## **W670: International Wheat Genome Sequencing Consortium (IWGSC)**

### **Releasing Natural Variation in Bread Wheat by Modulating Meiotic Crossovers**

Stuart Desjardins<sup>1</sup>, Daisy Ogle<sup>2</sup>, Kim Osman<sup>3</sup>, Wei Jiang<sup>4</sup>, James Simmonds<sup>5</sup>, Mark O. Winfield<sup>6</sup>, Cristobal Uauy<sup>5</sup>, Eugenio Sanchez-Moran<sup>3</sup>, Chris Franklin<sup>3</sup>, Ian R Henderson<sup>4</sup>, Keith J. Edwards<sup>6</sup> and **James D. Higgins<sup>1</sup>**, (1)University of Leicester, Leicester, United Kingdom, (2)University of Leicester, LEICESTER, United Kingdom, (3)University of Birmingham, Birmingham, United Kingdom, (4)University of Cambridge, Cambridge, United Kingdom, (5)John Innes Centre, Norwich, United Kingdom, (6)University of Bristol, Bristol, United Kingdom

Meiotic recombination is a fundamentally important mechanism in wheat breeding as it generates novel variation and creates unique allelic combinations. The ability to control recombination in crops would revolutionize breeding by unlocking genetic diversity and thus making a significant contribution to much needed future increases in yield and quality. To tackle this problem, a BBSRC strategic LoLa team has been assembled consisting of researchers from Birmingham, Bristol, Cambridge, Leicester and the John Innes Centre. Our approach is organised into five work packages: 1) *Molecular cytogenetic tools for the analysis of meiotic progression*; 2) *Identification of putative gene targets*; 3) *Identification of mutant alleles of genes of interest*; 4) *Establishing a synthetic biology approach to modify recombination*; 5) *Development of the required genetic materials and markers*. The 5-year project is currently coming to the end of its fourth year, and results will be presented as an overview of progress from all work packages. This will include current data as well as future directions in the context of the aims of the IWGSC.

## **W671: International Wheat Genome Sequencing Consortium (IWGSC)**

### **Serpins: Genome-Wide Characterisation and Expression Analysis of the Serine Protease Inhibitor Family in *Triticum aestivum***

**Harriet Rose Benbow**, University College Dublin, Dublin, Ireland

Serpin genes are serine protease inhibitors that can inhibit endogenous or exogenous proteases. We have identified and annotated the entire 'Serpinome' of wheat from the IWGSC RefSeq v1.0 and constructed a high-quality phylogenetic tree of the gene family. The expression of these genes was

explored using published RNAseq data across a variety of tissues from the developing grain, spikelet and spike of wheat. We show that certain clades of the serpin family are highly expressed during grain development, and that there appears to be some functional redundancy in this gene family. Furthermore, using 13 RNAseq datasets of wheat tissues infected with biotic stressors, we identified serpins with a significant disease response. The majority of these disease-responsive serpins were upregulated by *Fusarium graminearum*, a devastating fungal pathogen that attacks the spike and developing grain of wheat. A subset of serpins displayed pleiotropic activity and therefore may contribute to our knowledge of yield:disease trade-offs in elite wheat breeding. By combining gene expression and phylogenetics, we observed a difference in the evolutionary history of the function of these genes: those involved in grain development are functionally well-conserved (although variable in sequence), whereas disease-responsive serpins appear more sporadically across the phylogenetic tree, indicating variability and rapid evolution of function. We hypothesise that these genes evolved to inhibit proteases from invading fungal pathogens thus contributing to effector-triggered immunity.

This study identified potential targets for wheat improvement and may shed light on plant: pathogen interactions and the evolution of disease resistance in wheat.

## **W672: International Wheat Genome Sequencing Consortium (IWGSC)**

### **Homeologous Epistasis in Wheat: The Search for an Immortal Hybrid**

**Nicholas Santantonio**<sup>1</sup>, Jean-Luc Jannink<sup>2</sup> and Mark E. Sorrells<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)USDA-ARS / Cornell University, Ithaca, NY

The three subgenomes of allohexaploid wheat each contain mostly complete, yet evolutionarily divergent, sets of genes. Like a diploid hybrid, allopolyploids will have multiple versions, or homeoalleles, for every gene. Functional redundancy between these homeologous genes should result in a deviation from additivity. Epistatic interactions between homeoalleles are analogous to dominance effects, but are fixed across subgenomes through self pollination. Wheat can therefore be viewed as an immortalized hybrid, with the opportunity to identify, select and fix favorable homeoallelic interactions within inbred varieties. Prior to sequencing of the wheat genome, estimation of functional redundancy between homeoallelic loci was restricted to a few well known genes, and the global contribution of all homeoallelic interactions to genetic variation was largely unknown. We demonstrate a statistical framework using the homeologous dwarfing genes of wheat, *Rht-1*, and extend this methodology to search for genome-wide patterns indicative of homeoallelic subfunctionalization in a breeding population. Using the IWGSC RefSeq v1.0 sequence, 23,796 homeoallelic gene sets were identified and anchored to the nearest DNA marker to form 10,172 homeologous marker sets. Interaction predictors constructed from products of marker scores were used to fit homeologous interaction effects, as well as predict whole-genome genetic values. Some traits displayed a pattern indicative of homeoallelic subfunctionalization, while other traits showed a less clear pattern or were not affected. Using genomic prediction accuracy to evaluate importance of marker interactions, we show that homeologous interactions explain a portion of the non-additive genetic signal, but are less important than other epistatic interactions across subgenomes.

## **W673: Interoperability and Federation Across Bioinformatic Platforms and Resources**

### **The Airborne Environmental Observations Laboratory for Unoccupied Systems (AEOLUS)**

**Tyson Swetnam**, CyVerse, Tucson, AZ, David LeBauer, University of Arizona, Jane Wyngaard, University of Notre Dame and David Durden, Battelle Inc

Small Unmanned Aerial Systems (sUAS) are now ubiquitous and increasingly valuable tools for observing biological, ecological, and geophysical phenomena. For example, linking *in situ* observations from sUAS of organism phenotype to their genomic information, and environmental factors, is critical for so-called 'GxE' research. The **Airborne Environmental Observations Laboratory for Unoccupied Systems (AEOLUS)** establishes a cloud native cyberinfrastructure for sUAS data analyses and related research on publically-funded cloud and high performance computing resources. AEOLUS has three integrated Specific Aims: 1) the data management lifecycle through ingestion of data, automation of

metadata collection, curation, and support of data publication with attribution; 2) enablement of high throughput computing for data processing, orchestration of computational resources, and enablement of interactive scientific analyses through cloud native computing, containers, and machine learning; and 3) training and adoption of these workflows via in person workshops, on-line tutorials and step-by-step instructional manuals and materials. AEOLUS utilizes CyVerse, a national academic and research cyberinfrastructure. CyVerse provides computational resources and federation to other state and national resources (XSEDE), for processing these research sUAS data, hosting data sets that researchers are required to publish, and supports open source software for locating and reusing any sUAS data from local to continental scale. This talk will focus on the first and second aims of AEOLUS, and provide information about how to access CyVerse resources.

#### **W674: Interoperability and Federation Across Bioinformatic Platforms and Resources**

##### **MaizeGDB Interoperability for Multiple Genomes**

**Margaret Woodhouse**, USDA/ARS/MaizeGDB, Ames, IA

Lower cost and improved sequencing technology have made sequencing and assembly of multiple plant genomes a reality for many labs. As a consequence, agricultural and other genomics databases must be proactive in preparing for an influx of high-quality genomic assemblies since hosting a single reference genome is no longer sufficient to keep pace with current genomics research. But the challenge in hosting multiple genomes is the ability to connect these genomes to each other and to various phenotypic data types crucial to making these data useful to the genetics, genomics and plant breeding communities. The USDA-ARS supported maize genetics and genomics database (MaizeGDB) is expecting to host up to fifty genomes by the end of 2020. This talk will demonstrate MaizeGDB's pan-genomic approach to interconnecting multiple genomes that syntenically link shared loci such as the protein-coding gene space, cis and trans-regulatory elements, SNPs, GWAS, and phenotypic trait data.

#### **W675: Interoperability and Federation Across Bioinformatic Platforms and Resources**

##### **BrAPI: RESTful Specifications to Enable Interoperability among Plant Breeding Databases Focused on Plant Phenotype/Genotype Databases**

**Pete Selby**, Cornell University, Ithaca, NY

#### **W676: Interoperability and Federation Across Bioinformatic Platforms and Resources**

**TBD**

**Brian D. Gregory**, University of Pennsylvania, Philadelphia, PA

#### **W677: Interoperability and Federation Across Bioinformatic Platforms and Resources**

##### **A Calibration Method for Accurate and Reproducible Two-Dimensional Plant Measurements from Consumer Cameras**

**Amy Tabb**, USDA-ARS-AFRS, KEARNEYSVILLE, WV

Images are used frequently in plant phenotyping to capture measurements. Establishing the relationship of image pixels to camera geometry and physical units is a process called geometric camera calibration. Traditionally in computer vision, a fixed-focus camera is calibrated using many views of a calibration pattern. In this talk, I will describe a different way to calibrate that allows for fast acquisition of samples using consumer cameras, such as DSLR or cell phone cameras. The method uses a printed calibration pattern under the sample, and code that is a companion to this paper <https://arxiv.org/abs/1904.13187>. This method allows teams to collect image samples from different cameras, regions, and environments with comparison on the same physical scale for reproducible and accurate two-dimensional plant phenotyping.

## **W678: Interoperability and Federation Across Bioinformatic Platforms and Resources**

### **A Status Update and Preliminary Valuation of the GEMS Platform**

**Bryan Runck**, University of Minnesota, Minneapolis, MN

Over the past three years, the GEMS Agroinformatics platform went from an idea to a full beta version servicing scientists across the public, nonprofit, and for-profit sectors. This talk briefly describes the trajectory of the GEMS platform; provides a rough valuation of components of the platform in terms of cost and time; and overviews new developments in the areas of platform federation, data de-identification, and distributed sensing and robotics. While valuation is still preliminary, results suggest that platform components generate greater than 10x improvements in terms of cost and time savings for agricultural informatics practice.

## **W679: Interoperability and Federation Across Bioinformatic Platforms and Resources**

### **COPO: A Metadata Platform for Brokering FAIR Data in the Life Sciences**

**Felix Shaw**, Earlham Institute, Norwich, United Kingdom

Scientific innovation is increasingly reliant on data and computational resources. Much of today's life science research involves generating, processing, and reusing heterogeneous datasets that are growing exponentially in size. Demand for technical experts (data scientists and bioinformaticians) to process these data is at an all-time high, but these are not typically trained in good data management practices. That said, we have come a long way in the last decade, with funders, publishers, and researchers themselves making the case for open, interoperable data as a key component of an open science philosophy. In response, recognition of the FAIR Principles (that data should be Findable, Accessible, Interoperable and Reusable) has become commonplace. However, both technical and cultural challenges for the implementation of these principles still exist when storing, managing, analysing and disseminating both legacy and new data.

COPO is a computational system that attempts to address some of these challenges by enabling scientists to describe their research objects (raw or processed data, publications, samples, images, etc.) using community-sanctioned metadata sets and vocabularies, and then use public or institutional repositories to share it with the wider scientific community. COPO encourages data generators to adhere to appropriate metadata standards when publishing research objects, using semantic terms to add meaning to them and specify relationships between them. This allows data consumers, be they people or machines, to find, aggregate, and analyse data which would otherwise be private or invisible. Building upon existing standards to push the state of the art in scientific data dissemination whilst minimising the burden of data publication and sharing.

## **W680: IRIC: Rice Informatics for the Global Community**

### **The International Oryza Map Alignment Project (IOMAP) from a Platinum Reference Genome Sequence (PSRefSeq) Perspective**

**Rod Wing**, Arizona Genomics Institute, University of Arizona, Tucson, AZ

## **W681: IRIC: Rice Informatics for the Global Community**

### **A Fitness Consequence Map for the Rice Genome**

**Michael D. Purugganan**, New York University, New York, NY

The extent to which sequence variation impacts plant fitness is poorly understood. High-resolution maps detailing the constraint acting on the genome, especially in regulatory sites, would be beneficial as functional annotation of noncoding sequences remains sparse. Here we present a fitness consequence map for rice (*Oryza sativa*). We inferred fitness consequence scores ( $\rho$ ) for 246 inferred genome classes derived from nine functional genomic and epigenomic datasets, including chromatin accessibility,

mRNA/sRNA transcription, DNA methylation, histone modifications, and engaged RNA polymerase activity. These were integrated with genome-wide polymorphism and divergence data from 1,477 rice accessions and 11 reference genome sequences in the *Oryzaeae*. We found  $p$  to be multimodal, with ~9% of the rice genome falling into classes where more than half of the bases would likely have a fitness consequence if mutated. Around 2% of the rice genome showed evidence of weak negative selection, frequently at candidate regulatory sites, including a novel set of 1,000 potentially active enhancer elements. This fitness consequence map provides perspective on the evolutionary forces associated with genome diversity, aids in genome annotation, and can guide crop breeding programs.

## **W682:IRIC: Rice Informatics for the Global Community**

### **South Green Resources to Manage Rice Big Genomics Data**

Clément Agret<sup>1</sup>, Céline Gottin<sup>2</sup>, Alexis Dereeper<sup>3</sup>, Christine Tranchant-Dubreuil<sup>4</sup>, Annie Chateau<sup>1</sup>, Anne Dievert<sup>2</sup>, Gautier Sarah<sup>5</sup>, Alban Mancheron<sup>1</sup>, Guilhem Sempéré<sup>6</sup>, **Manuel Ruiz**<sup>7</sup> and Gaëtan Droc<sup>7</sup>, (1)LIRMM, France, (2)CIRAD, UMR AGAP, France, (3)IRD, UMR IPME, Montpellier cedex 5, France, (4)IRD - UMR DIADE, Montpellier, France, (5)INRA, UMR AGAP, France, (6)CIRAD, UMR Intertryp, Montpellier, France, (7)CIRAD, UMR AGAP, Montpellier, France

We have developed the Rice Genome Hub, an integrative genome information system that allows centralized access to genomics and genetics data, and analytical tools to facilitate translational and applied research in rice. The hub is built using the Content Management System Drupal with the Tripal module that interacts with the Chado database. The Hub interface provides several functionalities (Blast, DotPlots, Gene Search, JBrowse, Primer Blaster, Primer Designer) to make it easy for querying, visualizing and downloading research data. We also plugged in-house tools developed by the South Green bioinformatics platform.

Among these tools, Gigwa is a Web-based tool which provides an easy and intuitive way to explore large amounts of genotyping data by filtering the latter based not only on variant features, including functional annotations, but also on genotype patterns.

We also developed RedOak, a reference-free and alignment-free software package that allows for the indexing of a large collection of similar genomes. RedOak can be applied to reads from unassembled genomes, and it provides a nucleotide sequence query function. This software is based on a k-mer approach and has been developed to be heavily parallelized and distributed on several nodes of a cluster. Analysis of presence-absence variation (PAV) of genes among different genomes is a classical output of pan-genomic approaches. RedOak has a nucleotide sequence query function, including reverse complements, that can be used to quickly analyze the PAV of a specific gene among a large collection of genomes.

## **W683:IRIC: Rice Informatics for the Global Community**

### **Planteome: Ontologies and Resources for Integrative Plant Genomics**

**Laurel Cooper**, Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, OR

The Planteome Project ([www.planteome.org](http://www.planteome.org)) provides semantic integration of plant genomics datasets and reference ontologies describing domains in plant biology (plant anatomy, development, phenotypes, traits, growing conditions and treatments, plant stresses) across 95 plant species. The Planteome also hosts external reference ontologies, as well as species-specific Crop Ontologies developed by plant breeding and research communities from around the world. The latest release, Version 3.0 of the project database includes more than two million bioentities (data objects), with more than 21 million associations between bioentities and ontology terms. The online database and APIs provide researchers with tools to access resources for plant traits, phenotypes, diseases, genomes, gene expression and genetic

diversity data across 95 plant species. In this presentation, we will demonstrate how to access the tools and database resources at the Planteome for integration into your own research. The Planteome project is supported by the National Science Foundation award IOS #1340112

## **W684: Its More than Just Genomics: Understanding How Societal, Economic and Regulatory Factors Influence the Application of Genomics in Plants and Animals**

### **How Regulatory Factors Influence the Application of Genomics in Animals**

**Alison Van Eenennaam**, University of California, Davis, CA

Selection for more productive and resilient plant and animal varieties has been an incredibly important component of improving yield while resulting in a decreased environmental footprint per unit of food production. Although conventional animal breeding falls under no formal regulatory framework, it has a long history of safe use. Regulatory agencies do not evaluate new conventionally-bred breeds and varieties for health and environmental safety prior to commercial release. Breeders routinely phenotype potential breeding stock for changes in productivity, reproductive efficiency, reactions to disease, and quality characteristics, but they are not routinely evaluated for unintended effects at the molecular level. In the words of one animal geneticist, “For millennia, animal breeders have performed what amounts to a mega-scale, phenotype-driven mutagenesis screen.” Certain technologies used in modern breeding programs such as artificial insemination, genomic selection, in vitro maturation and fertilization of oocytes, embryo transfer, and even cloning are not formally regulated; although the veterinary drugs that facilitate advanced reproductive technologies are subject to regulation by the United States Food and Drug Administration (FDA) as animal drugs. These technologies have been rapidly adopted into breeding program design and have been associated with accelerated rates of genetic improvement and have generated little controversy. However, since 2009 the products resulting from the use of modern molecular technologies such as genetic engineering (GE) in animal breeding programs have been regulated as “new animal drugs”. Technically it is the “genomic alteration” resulting from the use of molecular technologies that is regulated as a drug, and not the animal itself. However given it is not possible to separate an animal from its own genome, this point becomes moot. Regrettably, in part due to the uncertain regulatory path to market, only a single food animal that has an intentional genomic alteration resulting from the use of molecular technologies has ever reached the U.S market. The fast-growing AquAdvantage salmon, first produced by researchers at a Canadian University in 1989. The multigenerational new animal drug evaluation required to commercialize this GE fish was subject to unaccountable regulatory delay. In early 2017, the FDA released its updated draft “Guidance for Industry #187” stating their intent to regulate all intentional alterations of genomic DNA introduced by gene editing in animals as new animal drugs, irrespective of the novelty of the alteration. This includes knock-outs, SNPs and intraspecies allelic substitutions that could otherwise have been achieved by conventional cross breeding. The draft guidance suggests the need for genotypic and phenotypic durability studies over multiple generations, and recommends that at least two of the sampling points be from non-contiguous generations (e.g., F1 and F3). Further, the draft guidance recommends that all surplus investigational animals and their biological products be disposed of by incineration, burial, or composting. The costs and timeframes associated with the new animal drug regulatory approach in food animal species will make it difficult for public sector researchers and small companies to use gene editing methods to introduce useful beneficial alterations to solve zoonotic disease and animal welfare problems in the United States.

## **W685: Its More than Just Genomics: Understanding How Societal, Economic and Regulatory Factors Influence the Application of Genomics in Plants and Animals**

### **Opportunities, Risks, and Benefits of Genetic Modification: Sunflower as a Case Study**

**Loren Rieseberg**, University of British Columbia, Vancouver, BC, Canada

In the late 1990s and early 2000s, numerous transgenic traits were introduced into sunflower cultivars, ranging from herbicide resistance to increased nitrogen assimilation to enhanced protein quality. Despite the potential benefits of these traits, genetically modified (GM) sunflowers were not commercialized for two main reasons. First, sunflower is native to North America and high levels of gene flow have been

documented with its wild relatives, which impeded regulatory approval in the USA. Second, sunflower is widely grown in Europe, and there were concerns that the release of GM varieties would damage access to the edible oil market due to public and regulatory resistance to GM products. As a result of the decision to stay non-GMO, sunflower production soared in Europe, but declined in the Americas due to competition from GM crops such as soybean and canola. Another indirect consequence has been an increased reliance on sunflower wild relatives for genetic diversity in breeding. This approach is time-consuming and has resulted in the introduction of maladaptive alleles and structural variants into the sunflower crop gene pool. The advent of genome editing earlier this decade resulted in a burst of interest in both the public and private sector in harnessing this technology, both for traits previously targeted by transgenic approaches, as well as for new traits. However, recent regulatory decisions in Europe have dampened this initial enthusiasm. If genome editing is not deployed for sunflower improvement, then the current duality of the sunflower market is likely to be further exacerbated, with continuing production gains in Europe and Asia and potentially further declines in the Americas.

### **W686: Its More than Just Genomics: Understanding How Societal, Economic and Regulatory Factors Influence the Application of Genomics in Plants and Animals**

#### **Applying Genomic Knowledge to Forest Trees by Gene Editing and Transformation: A Case Study of Scientist Efforts to Resolve Market and Regulatory Obstacles to Field Research and Breeding**

**Steven H. Strauss**, Oregon State University, Corvallis, OR

The use of plant biotechnology methods that fall outside of conventional breeding and random mutagenesis are excluded by forest product certification brands and/or stringently regulated in most of the world. These rules effectively classify the method of direct (recombinant DNA-based) modification as inherently dangerous or socially/ethically unacceptable, in opposition to decades of advice from scientists worldwide that stress product over process as the focus for risk/benefit considerations. In forestry and woody plant horticulture there has also been sufficient applied research and field trials that the benefits and general reliability of direct modification methods are well established. In addition, the expectation of severe climate change and associated rapid biotic and abiotic stresses on trees has made the case for more aggressive innovation and breeding much stronger than at the times the restrictive policies were established. I will describe the broad context for this ethical and political debate with respect to forestry practices and policies, biological opportunities and obstacles to the use of direct modification methods, and the recent efforts of a group of scientists to try and remove the blanket prohibitions to the use of directly modified trees in field research or operations by certified forest product companies. I argue that scientists need to use their strong “brand credibility” to much more aggressively organize and advocate for policies that follow, rather than contradict, scientific findings and advice. Perhaps surprisingly “green” organizations and companies appear to present some of the largest obstacles to progress in this domain.

### **W687: Its More than Just Genomics: Understanding How Societal, Economic and Regulatory Factors Influence the Application of Genomics in Plants and Animals**

#### **Public and Stakeholder Engagement in the Governance of Emerging Biotechnologies**

**Jason Delborne**, Genetic Engineering and Society Center, North Carolina State University, Raleigh, NC

Emerging biotechnologies not only challenge existing regulatory regimes, but also raise new ethical, environmental, and social questions. For example, the genetically engineered American chestnut tree, which includes a gene from wheat to protect the tree from a fungal blight that drove the population to functional extinction, is designed to persist and spread in unmanaged environments. How will the public perceive the entry of a genetically engineered tree into a national park or forest? What competing understandings of “naturalness” or “wildness” might lead people to oppose, support, or question the introduction of a GE American chestnut on public or private land? Such questions are complex and deserving of public and stakeholder engagement. Drawing upon reports by two National Academies of Sciences, Engineering, and Medicine committees (on gene drives and on forest biotechnology), a recent

IUCN report on synthetic biology and biodiversity conservation, and research projects on the GE American chestnut and gene drive mice, this presentation describes the opportunities and challenges of engaging stakeholders and broader publics about the governance of emerging biotechnologies.

### **W688: Its More than Just Genomics: Understanding How Societal, Economic and Regulatory Factors Influence the Application of Genomics in Plants and Animals**

**Moderated Panel and Audience Discussion (Panellists: A Van Eenennaam, L Rieseberg, S Strauss, J Delborne)**

**Kate E. Harland**, Genome British Columbia, Vancouver, BC, Canada

### **W689: IWGSC – Wheat Genome Manual and Functional Annotation**

#### **Impact of Wheat Refseq v1.0: Translational Research to Increase Wheat Yield by Isolating Sink Controlling Genes**

**Alex Mahlandt**<sup>1</sup>, Nidhi Rawat<sup>1</sup>, Raju Datla<sup>2</sup>, Cristobal Uauy<sup>3</sup> and Vijay K. Tiwari<sup>1</sup>, (1)University of Maryland, College Park, MD, (2)Global Institute for Food Security, Saskatoon, SK, Canada, (3)John Innes Centre, Norwich, United Kingdom

Bread wheat is one of the most important sources of calories for the global human population. To feed more than 9 billion people by 2050, global wheat production must increase alongside other crop plants. Synergistic approaches that include sustainable agronomic practices and genetic improvement for increased yield and resistance against biotic and abiotic stresses are required to fulfill the expected global wheat demand. There have been tremendous advancements in developing genomics resources of bread wheat, including completion of the reference genome, opening up new vistas for gene discovery and their deployment in wheat breeding programs. Integration of genetic tools, genomics resources and germplasm is critical to perform translational research for wheat improvement. By combining genetic tools with the availability of a wheat reference genome, its transcriptional landscape, exome sequencing datasets, and an excellent set of germplasm, we completed map-based cloning of two agronomically important genes controlling the number of grains per spike and spike length. Map-based cloning of these two agronomically important genes will generate new knowledge and help the wheat genetics community to increase the overall goal of yield increase. Additionally, our work outlines importance of IWGSC reference sequence and its impact on applied research on wheat improvement.

### **W690: IWGSC – Wheat Genome Manual and Functional Annotation**

#### **Annotation of the Wheat Prolamins**

**Yong Q. Gu**<sup>1</sup>, Naxin Huo<sup>1,2</sup>, Angela Juhasz<sup>3</sup>, Tingting Zhu<sup>2</sup>, Ming-Cheng Luo<sup>2</sup>, Susan B. Altenbach<sup>1</sup>, Daowen Wang<sup>4</sup> and Rudi Appels<sup>5</sup>, (1)USDA ARS, Western Regional Research Center, Albany, CA, (2)Department of Plant Sciences, University of California, Davis, Davis, CA, (3)School of Science, Edith Cowan University, Australia, (4)Chinese Academy of Sciences, China, (5)University of Melbourne, Melbourne, VIC, Australia

Wheat is one of the food crops most consumed by humans worldwide. However, the molecular basis of wheat flour end-use quality is still only partially understood. A single wheat cultivar contains about 70 to 100 similar but distinct gluten proteins (prolamins) that determine its end-use quality. Wheat flour proteins also trigger human health problems, including food allergies (FA), celiac disease (CD) and non-celiac wheat sensitivities (NCWS). Previous studies indicated that the wheat prolamins are encoded by complex multiple gene families that are mapped to three major genomic regions. However, it has been challenging to sequence these prolamins regions due to the presence of large gene family members and high content of repetitive DNA. In this study, we reconstructed high-quality sequences of wheat prolamins locus regions using PacBio long reads and BioNano genome maps. The BioNano maps proved to be useful not only in validating the accuracy of sequence contigs, but also in ordering and reorienting sequence contigs to build large scaffolds. The validated sequences harboring the prolamins gene loci

from the wheat A, B, and D genomes were annotated manually to identify a complete set of wheat prolamin genes from a single wheat cultivar cv Chinese Spring, facilitating more accurate and robust studies to understand the expression and function of individual prolamin genes using the wheat transcription data. To improve the prolamin annotation in the polyploid wheat genome, we aligned the reconstructed prolamin sequences with the corresponding regions from the IWGSC RefSeq V2 sequence. We found that although the two sequences aligned well, ~30% of the prolamin gene sequences in the reference sequence contained gaps, particularly in those repetitive regions of the genes containing microsatellite structures. In addition, several prolamin genes were missing likely due to the complexity in assembling gene family members with high sequence identities. Strategies for improving prolamin gene sequencing and annotation will be discussed in the presentation.

### **W691: IWGSC – Wheat Genome Manual and Functional Annotation**

#### **Genebank Genomics: Paving Way to Identify Candidate Genes for Disease Resistance in Wheat**

**Sandip Mallikarjun Kale**, Albert Wilhelm Schulthess, Axel Himmelbach, Martin Mascher, Jochen Christoph Reif and Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

More than half a million wheat genetic resources are resting in genebanks worldwide. Unlocking their hidden favorable diversity for breeding is pivotal for climate-smart agriculture needed to confront future food shortage. The Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK) hosts' *ex-situ* collection of ~ 28K wheat accessions. In the first phase of Genebank2.0 project, the genotyping-by-sequencing (GBS) approach was utilized to characterize all the winter wheat accessions from (IPK) *ex-situ* collections. The SNP data provided insight on population structure present within the collection. The study also identified tremendous diversity within genetic resources as compared to elite lines. In order to mine for favorable variation for biotic stress resistances against yellow rust (*Puccinia striiformis*), leaf rust (*Puccinia triticina*) and powdery mildew (*Blumeria graminis*), a trait-customized core collection has been defined. The core collection was then sequenced using resistance gene enrichment sequencing (RenSeq) approach. The genomic and phenotypic data was integrated and Kmer based association mapping was carried out in order to identify novel sources of resistances. The novel variation/(s) can be utilized after validation for improving disease resistances in wheat.

### **W692: IWGSC – Wheat Genome Manual and Functional Annotation**

#### **IWGSC Refseq Annotation v2.0**

**Frédéric Choulet**, GDEC, INRA, UCA, Clermont-Ferrand, France

An annotated reference genome sequence of bread wheat cv. Chinese Spring (IWGSC RefSeq) was published in 2018, representing 15.5 Gb assembled into 21 pseudomolecules, with 107k predicted genes and ca. 3.9 million copies of transposable elements. An improved version of the genome assembly is now available: IWGSC RefSeq v2.0. It resolved mis-oriented and mis-ordered scaffolds along the chromosomes and it included a step of gap-filling. Here, we report on the work that we are currently performing in order to transfer accurately the annotation v1.1 onto assembly v2.0. It concerns the anchor-driven alignment of the two versions of the 21 chromosome sequences and the development of a bioinformatic tool which aims at transferring automatically the annotated features. Issues and solutions regarding this update will be discussed, and a plan to deliver an annotation v2.0 in 2020 will be presented.

### **W693: IWGSC – Wheat Genome Manual and Functional Annotation**

#### **Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio**

**Shawn Quinn**, Curio Genomics, Dexter, MI

The publication of the IWGSC wheat genome reference and related annotations has unleashed new opportunities for researchers seeking to address critical global challenges to develop more productive and resilient crops in a time of climate changes and further population growth. However, with traditional

tools, the bioinformatics analysis of NGS sample data sequenced from this genome proves challenging due to its large size, 85% repeating sequences, and polyploidy nature.

We will present how we overcame the challenges of read mapping (both for DNA-Seq and RNA-Seq libraries) and read alignment visualization when dealing with the "large chromosome" complexity of the wheat genome. Additionally, we will demonstrate a novel approach to variant calling, coverage analysis, and gene expression calculation in hexaploidy species, along with the dynamic incorporation of the IWGSC reference and annotation sets. We will leverage several research examples developed, using the [Curio Genomics](#) platform, in collaboration with other IWGSC members to highlight powerful interpretive results and data visualizations, including an approach for filtering by predicted biological consequence as part of a variant analysis of one or more samples.

#### **W694: IWGSC – Wheat Genome Manual and Functional Annotation**

##### **JBrowse & Apollo for Manual Annotation of Wheat**

**Anne-Francoise Adam-Blondon**, URGI, INRA, Université Paris-Saclay, Versailles, France

#### **W695: IWGSC – Wheat Genome Manual and Functional Annotation**

##### **Applying Network Biology and Deep Learning Approach for Large-Scale Characterization of Gene Function in Wheat**

**Xi Wang**, BASF, Diagem, Belgium

BASF Agriculture Solution is committed to seed & trait business and provide biological solution. One of the focuses is R&D pipeline and seed development in wheat. To produce crop seeds with better performance such as higher yield, discovering trait-associated genetic elements followed by engineering and validation using GM and non-GM approaches is crucial. Accurate large-scale characterization of gene function can significantly facilitate identification and prioritization of candidate genes and therefore is one of the key first steps in the discovery pipeline. However, complexity of genome structure and lack of prior knowledge hamper understanding wheat gene function in large-scale. Genetic function has mostly been translated via homology reasoning, as this reflects the true evolutionary relationship across the species tree. However, delineating correct homology is a difficult problem and prone to error. In addition, complex evolution in crops, such as wheat has led to complex many-to-many homology relationships between genes. These relationships require complementation of homology relations with e.g. expression data in order to identify the gene that is most likely to have the 'same' function. This has motivated us to initiate several projects, focusing on novel wheat data generation and wet-/dry lab technology/methodology development. Two ongoing projects are relying on system biology approach and deep learning methodology, applying integrative network analysis on combined wheat tissue-specific interactome data, in order to identify functional modules, characterize gene function, distinguish gene homoeologous copies and eventually identify and prioritize candidate genes that are associated with traits of interest. The biological network data include robust internal and public co-expression data, but also high-quality protein-protein interaction and gene regulation network, which will be generated in-house using Y2H and DAP-seq experiments. On the other hand, deep learning approach will be developed to create a holistic image to which genetics and genomics features contribute. Gene function prediction in wheat using deep learning will be made in different steps, using both public and internal data sets: 1) learning from gene coding sequences with domain and sequence features, 2) learning from interactome data and network with complex pattern. We demonstrated the pilot study of GO prediction, which has shown that deep learning with improved architecture is more powerful than naïve homology search and machine learning for sequenced-based gene function prediction. In summary, we aim to leverage internal and academic expertise and resources to exploit novel experimental technologies and data analysis approaches for large-scale characterization of gene function in wheat, in complementation of homology-based solution we already have in-house.

#### **W696: IWGSC – Wheat Genome Manual and Functional Annotation**

## **Durum Wheat Pan-Transcriptome as a Bridge to Unravel Tetraploid and Hexaploid Wheat Gene Function and Evolution**

Danara Ormanbekova<sup>1</sup>, Sven O. Twardziok<sup>2</sup>, **Marco Maccaferri**<sup>1</sup>, Davide Scaglione<sup>3</sup>, Vera Vendramin<sup>3</sup>, Simone Scalabrin<sup>3</sup>, Giuseppe Sciarra<sup>4</sup>, Simona Corneti<sup>5</sup>, Matteo Bozzoli<sup>5</sup>, Andrea Massi<sup>6</sup>, Klaus F.X. Mayer<sup>2</sup>, Michele Morgante<sup>7</sup>, Curtis J. Pozniak<sup>8</sup> and Roberto Tuberosa<sup>9</sup>, (1)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy, (2)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (3)IGA Technology Services, Udine, Italy, (4)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Italy, (5)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, (6)Produttori Sementi spa - Syngenta, Argelato, Italy, (7)Università di Udine, Udine, Italy, (8)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (9)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, Bologna, Italy

This study presents the transcriptome analysis of 13 elite durum wheat varieties representatives of the worldwide cultivated germplasm. cDNA libraries were produced from roots, seedling leaves and developing grains. Based on the reference genome sequence assembly of durum wheat cv. Svevo, 75.0, 70.5 and 74.5% of high-confidence Svevo genes were expressed in grain, leaf and root, respectively. Principal Component Analysis (PCA) showed a gene expression clustering led by tissues and varietal ancestry. Differentially up- and down-regulated gene clusters based on tissues and varieties were identified.

Functional enrichment analysis for three Gene Ontology terms showed that differentially expressed genes were significantly enriched in transport, kinase activity, binding, enzyme activity and protein metabolism. Variance expression analysis projected on the Svevo assembly revealed the chromosome regions that drove the major expression variation patterns. Clustering the gene expression profiles and the cultivar's expression profiles evidenced several gene expression patterns related to their co-ancestry, particularly for the grain. Towards a more complete assembly of a pan-transcriptome in durum, the cultivar-specific reads that could not be mapped on the Svevo genome (4-30% referred to Svevo Illumina sequencing data) are being *de novo*

assembled. Further, using the transcriptome of the 13 varieties in relation to bread wheat reference genome (cv. Chinese Spring IWGSC RefSeq) we are currently investigating the gene loss/deletion during the polyploidisation events. Moreover, the availability of the genome assemblies of the 10+ Wheat Genomes Project, which includes cultivars that represent genetic diversity, will allow us to infer strong allele fixation events (allopolyploidisation bottleneck).

#### **W697: IWGSC – Wheat Genome Manual and Functional Annotation**

##### **Applying Machine Learning to Plant Literature: Augmenting Human Curation**

**Tanya Z. Berardini**, Phoenix Bioinformatics, Fremont, CA

While plant genome sequencing and computational analyses of large plant gene datasets have become routine, major challenges remain in assigning a function to each sequenced gene. High quality predictions of function rely on a solid foundation of ground truth about gene function rooted in experimentally backed results. The work of capturing such experimental data from the literature continues to be done by highly trained biocurators but, absent massive amounts of funding, is difficult to scale. Only Arabidopsis has had a large-scale effort to capture gene function from the literature, yet only ~30% of all Arabidopsis articles with relevant gene function data have been exploited for this purpose. For plant species that lack a dedicated set of data curators, manual extraction and capture of such information is not feasible. The fields of machine learning and natural language processing can be harnessed to convert the valuable experimental gene function data in the plant biology literature into structured, reusable data and automate at least part of the process of literature curation for gene function assignment of plant genomes. This talk will describe a strategy for developing such a pipeline and how the resulting software and data could be applied to the wheat genome and result in a functional annotation set that is both broader and deeper in information content than what is currently available.

#### **W698: IWGSC – Wheat Genome Manual and Functional Annotation**

##### **IWGSC Phase II Activities: Moving from Structural to Functional Wheat Genomics**

**Kellye Eversole**, IWGSC, Lee's Summit, MO

Bread wheat, the staple food for 35% of the world's population, is the last major crop species to benefit from a reference genome sequence. The IWGSC, with 2,400 members in 68 countries, is an international, collaborative consortium, established in 2005 by public and private wheat growers, breeders, and scientists, with the aim of delivering genomic tools and resources for wheat improvement.

In 2018, the IWGSC completed Phase I when it published the first high quality reference sequence of the bread wheat variety *Chinese Spring* (IWGSC RefSeq v1.0). The IWGSC RefSeq v1.0 represents 94% of the hexaploid wheat genome in 21 chromosome pseudomolecules; and identifies 107,891 high confidence genes, along with 4.7 million molecular markers.

The IWGSC has now entered a new phase of its activities. In particular, an improved version of the reference sequence, IWGSC RefSeq v2.0, which closed a number of gaps and corrected position and orientations of scaffolds, was released to the community in July 2019 under the Toronto pre-publication access agreement. The development of annotation v2.0 integrating functional and manual annotation, as well as alignment with other genomic resources, is also in progress. The IWGSC has established a

collaboration with Arbor Biosciences and released version 1 of a wheat exome capture panel in the fall of 2019 and is developing a promoter capture module for wheat. Finally, a wheat diversity project has begun, aiming at completing high quality sequences of landraces that will provide a complete representation of wheat's worldwide genetic diversity. By expanding structural genomics and moving towards functional understanding of the wheat genome, these activities will lay the foundation for genomics-based wheat improvement in response to challenges imposed by population expansion and climate change.

An overview of the IWGSC recent achievements will be presented, along with an outline of its current activities.

## **W699: IWGSC – Wheat Genome Manual and Functional Annotation**

### **IWGSC Proteomics Project**

**Song Weining**, Northwest A&F University, Yangling, China

## **W700: JBrowse, JBrowse 2 and Apollo**

### **Comparing Genomes and Visualizing Synteny with JBrowse 2**

**Colin Diesh**, University of California Berkeley, Bioengineering, Berkeley, CA

JBrowse 2 is a ground-up rewrite of JBrowse using modern web technologies. The motivation is to address emerging applications in genomics, and to make it easier for third-party developers to extend. We will describe new functionality of JBrowse 2 including synteny views, split views, and dynamically-rotatable circularized views inspired by tools like Circos. JBrowse 2 preserves the capabilities of JBrowse 1 (including full backward-compatibility with JBrowse 1 and the ability to interface with common bioinformatics data formats like BED, BAM, CRAM, and tabix), but moves to an even more modular architecture that can have multiple views on the same screen, and supports the ability to implement new views, tracks, stores, etc. as plugins. JBrowse 2 can be downloaded as a desktop application (built using the Electron framework), which requires no web server setup at all, or it can be used as a static website that is easy to set up and cheap to maintain (for example, on cloud hosting platforms). We demonstrate visualizing synteny in a real-world use case.

## **W701: JBrowse, JBrowse 2 and Apollo**

### **Apollo Provides Collaborative Genome Annotation Editing with the Power of JBrowse**

**Nathan A Dunn**<sup>1</sup>, **Colin Diesh**<sup>2</sup>, **Helena Rasche**<sup>3</sup>, **Anthony Bretaudeau**<sup>4</sup>, **Robert Buels**<sup>2</sup>, **Nomi Harris**<sup>1</sup> and **Ian Holmes**<sup>5</sup>, (1)Lawrence Berkeley National Laboratory, Berkeley, CA, (2)UC Berkeley Bioengineering, Berkeley, CA, (3)University of Freiburg, Freiburg, Germany, (4)Plant Health and Environment Division of INRA, Paris, France, (5)University of California, Berkeley, Berkeley, CA

JBrowse is a powerful genome viewer that allows researchers to view and share their genomes as well as customize that experience. The visualization of genomic elements is an important step to be able to more precisely describe annotated genomes, which is vital for accurately modeling the biological function of genomic elements. Apollo (<https://github.com/GMOD/Apollo>) is a web-based genome annotation that editor that utilizes JBrowse to allow users to refine their genome annotations using JBrowse tracks as evidence, such as genomic and transcriptome elements and predictive models.

In addition to the ability to visually review diverse sets of information, Apollo is also a collaborative tool with many features that improve the efficiency of an annotation project including real-time collaborative editing, the ability to promote search results directly as evidence, a revertible and visual history of genomic edits, and many automated structural editing operations. Additionally, functional annotations including Gene Ontology (GO) annotations are supported, as well as the ability to populate arbitrary metadata that can be predefined to support a research group's workflow.

Here, we will show how to get a project going quickly using Docker. This includes adding users, uploading genomes (FASTA) and evidence (GFF3, VCF, BAM) to the user-interface, creating annotations and exporting those annotations with author attributions as GFF3 for structural annotations and GPAD2/GPI2 for GO annotations. Finally, we will show how to take advantage of Apollo's web services using the python-apollo library (<https://pypi.org/project/apollo/>).

Apollo is used in hundreds of genome annotation projects around the world, ranging from the annotation of a single species to lineage-specific efforts supporting the annotation of dozens of genomes.

## **W702: JBrowse, JBrowse 2 and Apollo**

### **JBrowse: A Hands on Tutorial**

**Scott Cain**, Ontario Institute for Cancer Research, San Diego, CA

#### **Tutorial Level**

Beginner to Intermediate. Students should be comfortable performing simple command line tasks like moving files and running scripts.

#### **Intended Audience**

JBrowse is sufficiently easy to install that a biologist can easily set up and configure a JBrowse server after the initial hurdles of learning about configuration options and file formats are overcome. This class is intended to help them over those hurdles.

#### **Prerequisite Software and Conference PCs**

Prerequisite software for JBrowse will be pre-installed on the conference PCs in the classroom area of the California Room. Participants using these PCs will be able to setup and configure JBrowse during the workshop.

After the workshop, a VirtualBox system image with JBrowse prerequisite software pre-installed will be made available on [GMOD @ PAG page](#) at [GMOD.org](#). You can use this image to walk through the material presented at this workshop.

## **W703: Legumes**

### **Symbiotic Stimulation of Root Development in *Medicago truncatula* through the Eyes of GWAS**

**Sandra Bensmihen**, Laboratoire des Interactions Plantes Micro-organismes (LIPM), Université de Toulouse, Castanet-Tolosan, France

*Medicago truncatula* is a model legume able to establish two types of root endosymbioses: root nodule symbiosis (nodulation) with nitrogen fixing rhizobia and arbuscular mycorrhizal symbiosis with some Glomeromycotina fungi. These two types of root endosymbioses are beneficial for nitrogen and phosphate nutrition of the plant, respectively. Interestingly, establishment of these two endosymbioses requires a common set of plant genes, so called "Common symbiosis signaling pathway" (CSSP), and these two types of symbionts produce similar symbiotic signals called Lipo-chitoooligosaccharides (LCOs). On top of their role in the establishment of root endosymbiosis, purified LCOs also stimulate root branching.

Exploiting natural diversity together with genome wide association study (GWAS) is a powerful approach to explore trait architecture to identify gene polymorphisms associated with plant adaptation. This strategy was previously applied to identify nodulation-related loci in *Medicago truncatula*. Here, we have

used GWAS together with a newly developed local score test to discover genomic regions including SNPs associated with rhizobial (Nod) or mycorrhizal (Myc) LCO stimulation of lateral root development in 173 natural accessions of *Medicago truncatula*. Heritability was higher for Nod-LCO than Myc-LCO. Using different phenotypic parameters and the local score approach, we could identify 123 loci for Nod-LCO and 71 for Myc-LCO stimulation. The loci identified suggest strikingly different mechanisms possibly underlying lateral root stimulation by these two closely related molecules that I will discuss here.

#### **W704: Legumes**

##### **Beyond the Sequencing of the Pea Genome: Opportunities for Genomics-Based Breeding and Translational Approaches**

**Nadim Tayeh**, INRAE, UMR1347 Agroécologie, Dijon, France

The tribe Fabeae comprises more than 300 legume species, including some of the most ancient and important crops like *Pisum sativum* (pea), *Lens culinaris* (lentil), and *Vicia faba* (faba bean) used for food and feed. The genome sequence of pea, released in 2019, is an important milestone for the community working on legumes and especially on Fabeae. It brings into light evidences related to the genome expansion that occurred after the divergence of Fabeae from their sister tribes and highlights different chromosomal rearrangement events specific or not to the *Pisum* lineage. The pea genome sequence also represents a valuable resource to accelerate our understanding of the molecular basis of agronomically important traits and support breeding efforts. Great opportunities to address current agriculture challenges through genomics-based breeding and translational research projects are currently available. They are crucial for improving legume production and maintaining agricultural sustainability in the context of current and future agricultural challenges.

#### **W705: Legumes**

##### **A Single Major Effect Locus Controls Seed Imbibition and Flooding Tolerance in Common Bean (*Phaseolus vulgaris* L.)**

**Ali Soltani**<sup>1</sup>, Katelynn Walter<sup>1</sup>, Andrew T. Wiersma<sup>1</sup>, James P. Santiago<sup>1</sup>, Michelle Quigley<sup>1</sup>, Dan Chitwood<sup>1</sup>, Thomas D. Sharkey<sup>1</sup>, Timothy Porch<sup>2</sup>, Phillip Miklas<sup>3</sup>, Phillip McClean<sup>4</sup>, Juan M. Osorno<sup>4</sup> and David Lowry<sup>1</sup>, (1)Michigan State University, (2)USDA-ARS Tropical Agriculture Research Station, Mayaguez, PR, (3)USDA-ARS, Prosser, WA, (4)North Dakota State University, Fargo, ND

Slow-imbibing beans better tolerate flooding at germination stages. However, they are scarce within “Andean” genotypes, potentially due to their undesirable longer cooking time. We identified a slow-imbibing landrace genotype (PI 163122) that originated from a tropical region of India. This genotype and its derivative, PR9920-171 were about 3.5X more tolerant to flooding at germination stage compared to other Andean beans. Whole genome sequencing bulked segregant analysis of PR9920-171 × TARS-HT1 RILs revealed a single significant peak at the end of Pv03 that controls for imbibition rate. To fine-map this QTL, we saturated this region with KASP/Indel markers and subsequently conducted whole-genome sequencing of selected 30 recombinant lines. In addition, to track the uptake of water into the seed and to improve the accuracy and throughput of phenotyping, we imaged different bean lines with a CT scanner. By employing these strategies, we narrowed down the causal region to a 118 kbp interval. This interval contains 11 gene models, of which, four genes were detected with a total of seven variants within their exonic regions. The allelic distribution of these variants were assessed among other Andean genotypes. Results indicate that ABSCISIC ACID 8'-HYDROXYLASE-1 (CYP707A1) is the potential causal gene for slow-imbibing phenotype. This gene is critical for controlling seed dormancy in *Arabidopsis* via ABA regulation. We are currently investigating the slow imbibing allele frequency within wild Andean beans to establish whether this variant was selected upon during the domestication of common bean.

#### **W706: Legumes**

##### **Cowpea Genome Information Resources**

**Timothy J. Close**<sup>1</sup>, Stefano Lonardi<sup>2</sup>, Shengqiang Shu<sup>3</sup>, Maria Munoz-Amatriain<sup>4</sup>, Sassoum Lo<sup>5</sup>, Pei Xu<sup>6</sup>, Andrew D. Farmer<sup>7</sup>, Samuel A. Hokin<sup>7</sup>, Qihua Liang<sup>2</sup>, Jodi L. Humann<sup>8</sup>, Steven B. Cannon<sup>9</sup>, Joseph W. Carlson<sup>10</sup>, James Crabb<sup>11</sup>, Patrick A. Davidson<sup>10</sup>, David M. Goodstein<sup>3</sup>, Yi-Ning Guo<sup>12</sup>, Richard D. Hayes<sup>13</sup>, Dorrie Main<sup>14</sup>, Jeremy Phillips<sup>3</sup>, Philip A. Roberts<sup>1</sup> and Xinyi Wu<sup>15</sup>, (1)University of California, Riverside, Riverside, CA, (2)Department of Computer Science & Engineering, University of California, Riverside, CA, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)Department of Soil and Crop Sciences, (5)Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA, (6)China Jiliang University, China, (7)National Center for Genome Resources (NCGR), Santa Fe, NM, (8)Washington State University, Pullman, WA, (9)USDA–Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA, (10)DOE Joint Genome Institute, Berkeley, CA, (11)Department of Horticulture, Washington State University, (12)Department of Botany & Plant Sciences, University of California - Riverside, Riverside, CA, (13)US Department of Energy Joint Genome Institute, Berkeley, CA, (14)Department of Horticulture, Washington State University, Pullman, WA, (15)Zhejiang Academy of Agricultural Sciences, China

Cowpea (*Vigna unguiculata* [L.] Walp.) is a diploid warm-season legume, also known as black-eyed pea, among many other names. Cowpea is relevant mainly as a grain legume in the USA, Europe and Latin America, and as a fresh vegetable (longbean) in China and elsewhere in Asia. It is also of major importance as food and fodder in sub-Saharan Africa. The Global Food Security Reauthorization Act of 2018 (H. R. 5129) identified twelve target countries, several of which are African nations that rely on cowpea. Here we summarize progress on cowpea pangenome sequencing and annotation, focused so far on cultivated accessions including one representative of each of five sub-populations (IT97K-499-35 from the IITA breeding program in Nigeria, CB5-2 bred in California, Suvita2 as a landrace from Burkina Faso, Sanzi as a landrace from Ghana, and UCR779 as a landrace from Botswana) and two longbean accessions from the Asian sub-population (elite TZ30, and landrace ZN016). Several online portals provide various levels of access to information (e.g. genes, gene expression, variants) related to the IT97K-499-35 genome, which was the first well-sequenced accession among these seven (Lonardi et al., 2019, *Plant Journal* 98:767-782), and to the others that have been sequenced and annotated more recently. These portals are Phytozome (phytozome.net), the Legume Information System including Legume Mine (legumeinfo.org), NCBI Genome (ncbi.nlm.nih.gov/genome), and the Pulse Crop Database (pulsedb.org). Work is underway to provide interactive visual representations of multiply aligned genome sequences to provide facile access to features of the cowpea pangenome. Several US federal funding sources have supported or are currently supporting the development of this information and the teams that are now building these cowpea genome resources: NSF, USAID, USDA, DOE and NIH. The National Natural Science Foundation of China and the National Ten-Thousand Talents Program of China have supported the longbean sequencing. Additional sources of support are also noted.

## **W707: Legumes**

### **Towards Precision and Scalable Genome Editing in Crops**

**Feng Zhang**, University of Minnesota, St Paul, MN

A critical challenge we face today is to produce sufficient food and plant-derived products for a growing population. This challenge is further complicated by an ever-changing and unstable climate. In the past 30 years, a substantial effort has been made to sequence, assemble and characterize the genomes of more than thousands of plant species and landraces. This effort led to significant advances in understanding the basic mechanisms of plant growth, development, and interaction with the environment. The knowledge gleaned from these studies has paved the way to design better plant varieties by modifying, editing and rewriting their genetic code. My research has been focused on developing highly efficient genome editing and synthetic biology approaches to improve plant productivity and quality. In this talk, new approaches will be presented by combining genome editing and protoplast engineering technologies to improve precise genome editing in plant species.

## **W708: Legumes**

## **Structural Variation Underlies Multiple Agronomically Important Soybean Traits**

**Matthew E. Hudson**, University of Illinois, Urbana, IL

The study of genomic diversity in plants has mostly focused on small genetic polymorphisms such as single nucleotide polymorphisms (SNPs). However, larger structural changes in plants are increasingly known to be important for a number of traits. This is particularly well established for the rapidly evolving genes that mediate resistance to pests and pathogens. Copy number variation of DNA segments (CNV), presence-absence variation (PAV) and structural rearrangements are particularly abundant in resistance gene clusters encoding proteins in the canonical plant resistance gene family, the nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins. More recently, other types of gene have also been implicated in resistance to pathogens conferred by structural polymorphisms, such as in the case of the soybean cyst nematode (SCN, *Heterodera glycines*) resistance gene *Rhg1*. Data will be presented that other loci in soybean that confer important agronomic traits are associated with complex structural polymorphisms.

## **W709: Linkage and Deletion Mapping**

### **High-Density Genetic Linkage Maps: Methods and Applications**

**Beth Rowan**, University of California, Davis, Davis, CA

Genetic linkage maps are useful for a wide range of biological studies, including identifying the genetic basis of traits, understanding the process of meiotic recombination, and anchoring/verifying genome assemblies. High-throughput sequencing approaches have facilitated both the discovery of markers needed for generating genetic maps and the methods for their construction. Experimental design – from the type of population used to the selection of the sequencing/computational approaches – is something that must be carefully evaluated in light of the downstream application(s) of the map. Genetic distances in linkage maps can be influenced by technical aspects of the design (coverage, marker selection, algorithm) and by variation in the crossover rate along the chromosomes. The latter is influenced by many factors, including differences in structure between the two homologous chromosomes, where rearrangements of sequence such as transpositions and inversions typically lead to suppression of crossovers. Although the suppression around large (cytologically visible) rearrangements across a taxonomically diverse group of organisms has been known for decades, it has remained unclear whether suppression occurs over all types and sizes of structural variants and for how great of a physical distance beyond the borders of the variants. To assess the impact of structural variation on meiotic recombination, we used the sequences of over 2000 recombinant F<sub>2</sub> individuals from a single cross between two well-studied strains of the model plant *Arabidopsis thaliana* to make an extremely dense crossover map. Combining this map with the precise knowledge of structural differences between the two homologous chromosomes, we determined that crossovers are generally suppressed across of variants across a large range of sizes, however over only a very narrow window spanning the variant region. This knowledge will inform the design of future genetic mapping experiments. Newer sequencing methodologies (long-read, linked-read, and single cell sequencing) offer the possibility for constructing genetic linkage maps directly from gametes. These will be discussed, along with a generalized set of strategies for tailoring the experimental design for the genetic map to the downstream application.

## **W710: Linkage and Deletion Mapping**

### **Assembly-Free, High-Density Linkage Map Construction from Whole Genome Sequencing Data**

**Kyle Fletcher** and Richard Michelmore, The Genome Center, University of California, Davis, CA

We have implemented a method for construction of high-density linkage maps from raw sequencing reads without the need for a genome assembly. Conventionally, reads are aligned to an assembly to identify polymorphic sites that are then used to call the genotypes of segregating progeny. Typically, read-mapping and variant calling are computational bottlenecks. Therefore, assembly-free linkage mapping would streamline map construction. Polymorphic markers can be identified through interrogation of whole genome shotgun reads of both parents that can be classified as single or multi-nucleotide variants. A genotype table can be rapidly generated by the presence or absence of these markers in whole genome shotgun data of progeny that can be analyzed by established tools for linkage

mapping. We have applied this approach to whole-genome sequencing data of F<sub>1</sub> progeny derived from the out-crossing oomycete *Bremia lactucae*. Analysis identified over 17,000 markers which segregated in a near 1:1 manner. Linkage analysis using MSTmap placed 99% of the markers into 15 linkage groups. These linkage groups were colinear with the chromosome-scale *B. lactucae* reference assembly. This pipeline fills a niche not currently filled by other software solutions.

### **W711: Linkage and Deletion Mapping**

#### **Needs for Future Genetic Mapping Analyses Learned after Five Years of the CottonSNP63K Array**

**Amanda M. Hulse-Kemp**, USDA-ARS and Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC

Standardized genotyping platforms, such as the CottonSNP63K array, have allowed research communities to complete many different research projects on the same platform. This is incredibly valuable as standardization allows for direct comparison between studies. Utilizing a standardized platform means that the amount of knowledge increases fairly quickly and leads to a specialized set of needs in regards to combining studies and performing meta analyses. This talk will discuss what has been learned in five years and take a look into the needs of the future for the cotton research community following the consortium produced genotyping array, with particular focus on genetic mapping studies.

### **W712: Linkage and Deletion Mapping**

#### **Using Probabilistic Genotypes in Linkage Analysis of Polyploids**

**Yanlin Liao**, Wageningen University and Research, Wageningen, Netherlands

In linkage analysis of polyploids using genetic markers, different approaches are used to assign discrete allele dosages to polyploid samples based on Single Nucleotide Polymorphism (SNP) array signal intensities. Unlike the situation in diploids, for polyploids SNPs can have more than two allele dosages. Current tools developed for linkage analysis and QTL analysis in polyploids mainly make use of estimated discrete dosages. However, the estimation of discrete dosages can be problematic in cases where different dosage classes are not well separated, leading to errors and possibly large numbers of missing values in the genotyping, or in discarding of markers; also, information about the uncertainty in estimating the dosage classes is lost. Therefore, we developed an approach for linkage analysis based on probabilistic allele dosages instead of discrete dosages. In this study, we evaluated the approaches using dosage probabilities or discrete dosages by simulating SNP array data with varying levels of data quality. We also re-analyzed data from a potato SNP array from an experimental study to evaluate the consequences in a practical situation of genetic mapping. Apart from the application in genotyping populations with SNP arrays, the use of dosage probabilities is also an attractive way to deal with Genotyping by Sequencing (GBS) data, where there is often more uncertainty about the allele dosages than in SNP array data, especially in polyploids.

### **W713: Linkage and Deletion Mapping**

#### **Bayesian Genotype Calling in Polyploids, and a Novel Statistic to Infer Marker Quality from Read Depth Distribution I**

**Joyce Njuguna**, Dept of Crop Sciences, University of Illinois, Urbana, IL

Polyploidy presents multiple challenges to accurate genotype calling from sequencing data. In autopolyploids, estimation of allele dosage at low to moderate read depth is confounded by sampling error and overdispersion. Bayesian genotyping methods utilize population information, in the form of expected genotype frequencies, to more accurately call genotypes in individuals. Moreover, for biallelic markers, posterior genotype probabilities generated by Bayesian methods can be used to estimate posterior mean genotypes, which are more accurate than discrete genotypes and retain power to detect additive associations between genotype and phenotype as well as additive relationships among

individuals. Our R package polyRAD is the only Bayesian genotype caller to use population structure and linkage disequilibrium to improve genotyping accuracy. In a study evaluating the impact of genotype calling on GWAS of polyploid species we demonstrate that the use of posterior mean genotypes increases power for GWAS to detect trait associations. In both recent and ancient allopolyploids, a major challenge is filtering or splitting paralogs so that all markers behave in a Mendelian fashion. We present a novel statistic, *Hind/HE*, that uses read depth to identify collapsed paralogs and other non-Mendelian markers, as well as individuals that are not the expected ploidy. This statistic can be estimated prior to genotype calling, allowing us to utilize *Hind/HE* in an algorithm that sorts GBS or RAD tags into their correct alignment locations within highly duplicated genomes. In a study evaluating the use of the *Hind/HE* statistic-based algorithm, we demonstrate significant results with the addition of a higher confidence in the alignment location of each tag. Functions and scripts using *Hind/HE* are available in polyRAD v1.2.

#### **W714: Linkage and Deletion Mapping**

##### **Bayesian Genotype Calling in Polyploids, and a Novel Statistic to Infer Marker Quality from Read Depth Distribution II**

**Wittney D. Mays**, University of Illinois, Dept of Crop Sciences, Urbana, IL

#### **W715: Linkage and Deletion Mapping**

##### **Holistic Genotyping of Amplicon Panels with Spades and Blastn**

**Charles F. Crane**, USDA-ARS and Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN

Amplicon panels fall into three categories, the first two of which have been used widely as genetic markers: microsatellites, RADseq/ddRADseq, and targeted panels based on exons or transposon insertion junctions. Traditional genotyping of these panels has rested on detection and scoring of a single measure of variation, such as microsatellite length or SNP base within the amplicon. The focus on a single measure has limited the amount of allelic variation that can be reliably scored at a single locus. Here I discuss another, more holistic approach, in which sequenced amplicons are assembled with SPAdes and then scored on the basis of blastn bitscore. This tactic allows various combinations of individual SNPs and/or repeat-count variants to be scored as distinct alleles. Simulation results indicate that distinct alleles can be recognized with as little as 1% variation among them, given typical base-calling error frequencies in Illumina reads. Application of the method to a set of eight population samples of Hessian fly (*Mayetiola destructor* (Say)) confirmed that the method was robust even though up to half of the reads within any individual fly did not map to the primary alleles because of sequencing errors. Problems with the method include very uneven sampling among competing amplicons in multiplexed PCR, allele-specific differences in amplification rate, difficulty proving that a given amplicon arises from one locus, and stutter if the amplicons contain a dinucleotide-based microsatellite. The ability to recognize  $n$  alleles in a heterozygous  $n$ -ploid parent should greatly facilitate linkage and deletion mapping in polyploid species, since the allele-phasing problem mostly vanishes and allele dosage can be estimated from sufficient depth of read coverage.

#### **W716: Machine Learning and Artificial Intelligence in Genomics and Phenomics**

##### **Deep Learning for Image-Based Plant Phenotyping**

**Ian Stavness**, University of Saskatchewan, Saskatoon, SK, Canada

Similar to many application domains in computer vision, deep learning is fast becoming the tool of choice for image-based plant phenotyping. Deep learning has potential to meet the challenges of estimating complex plant phenotypes from highly variable imaging conditions. However, care must be taken when applying deep learning architectures to relatively small sized and low variation datasets common in plant phenotyping. In this talk, I will provide an introduction to convolutional neural networks for plant scientists and discuss the canonical architectures used for computer vision tasks involving classification,

regression, detection, segmentation, et al. I will present recent work from our group on the application of these architectures to plant phenotyping tasks, and discuss opportunities (and challenges) for the future of this emerging field.

### **W717: Machine Learning and Artificial Intelligence in Genomics and Phenomics**

#### **EASEL: An Integrated and Accessible Framework for the Annotation of Eukaryotic Reference Genomes**

**Jill L. Wegrzyn**<sup>1</sup>, Jeremy Bennett<sup>2</sup>, Susan McEvoy<sup>1</sup>, Sumaira Zaman<sup>3</sup> and Peter Richter<sup>1</sup>, (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Connecticut, Storrs, CT, (3)Department of Computer Science and Engineering, University of Connecticut, Storrs, CT

High throughput technologies have increased both the number and quality of eukaryote genomes. This increase in reference genomes, and their associated contiguity, is not yet met with efficient and intelligent workflows for the accurate detection of protein coding genes. The prediction of a true gene, and the correct translation initiation start site (TIS), remains a tremendous challenge, especially for non-models with minimal genomic resources. The majority of pipelines utilize RNA-seq reads or pre-assembled transcripts to train (apply supervised or semi-supervised methods) Hidden Markov Models to predict gene structures for species with minimal or substantial existing genomic resources. These programs struggle with predicting less common gene structures (long introns, micro-exons), finding the preferred TIS location, and distinguishing pseudogenes. We present EASEL (Efficient, Accurate, Scalable Eukaryotic modelS), a genome annotation tool that leverages deep learning, RNA folding, and functional annotations to enhance gene prediction accuracy. EASEL features a deep LSTM network that has the capability to learn species-specific patterns to predict non-canonical gene structures and train in reasonable time. Existing high quality alignments from BUSCO and related protein sources train the network on intron/exon parameters. The implicated genomic regions are further refined via unsupervised training with RNA folding and traditional consensus patterns to improve TIS detection. Predicted proteins are subject to filtering via functional analysis information from the program EnTAP (gene family and protein domain signatures). The pipeline is benchmarked for user friendliness, efficiency, completeness, and accuracy in the detection of protein coding gene models.

### **W718: Machine Learning and Artificial Intelligence in Genomics and Phenomics**

#### **An Empirical Analysis of Deep Active Learning Methods for Image-Based Plant Phenotyping**

**Koushik Nagasubramanian**<sup>1</sup>, Talukder Zaki Jubery<sup>2</sup>, Seyed Vahid Mirnezami<sup>3</sup>, Asheesh K. Singh<sup>2</sup>, Arti Singh<sup>2</sup>, Soumik Sarkar<sup>3</sup> and Baskar Ganapathysubramanian<sup>3</sup>, (1)Department of Electrical Engineering, Iowa State University, Ames, IA, (2)Department of Agronomy, Iowa State University, Ames, IA, (3)Department of Mechanical Engineering, Iowa State University, Ames, IA

Deep learning models have been highly successful in the image-based plant phenotyping applications such as disease detection and classification. However, one of the main challenges in achieving this success is the requirement of large amount of labeled data. Data annotation could be costly, time consuming and hard for many plant phenotyping tasks. To overcome this challenge, recently many active learning algorithms have been proposed to reduce the amount of labeling needed by deep learning models for achieving high performance. We empirically evaluate the performance of these active learning methods for image-based plant phenotyping applications.

### **W719: Machine Learning and Artificial Intelligence in Genomics and Phenomics**

#### **A Differential Expression Computational Workflow for Time-Series Transcriptomic Studies**

**Yueyao Gao**<sup>1</sup>, Cole Younginer<sup>2</sup>, William L. Poehlman<sup>3</sup>, Elise Schnabel<sup>1</sup>, Julia Frugoli<sup>1</sup> and F. Alex Feltus<sup>1</sup>, (1)Clemson University, Department of Genetics & Biochemistry, Clemson, SC, (2)Clemson

University, Department of Electrical and Computer Engineering, Clemson University, Clemson, SC,  
(3)Sage Bionetworks, Seattle, WA

A fundamental research problem in a transcriptomic study is to identify differentially expressed genes (DEGs) across groups of samples. Time-series transcriptomic data plays a pivotal role in understanding dynamic biological processes. To date, most time-series transcriptomic studies use well-known software packages to detect DEGs, such as DeSeq2, edgeR, and limma, that were designed for single time point comparison, which is not appropriate for sequential time-series data. Here, we used DP\_GP ([https://github.com/PrincetonUniversity/DP\\_GP\\_cluster/tree/master/DP\\_GP](https://github.com/PrincetonUniversity/DP_GP_cluster/tree/master/DP_GP)) for identifying DEGs for time-series data. We then applied neural networks to validate the distinct clusters of responsive transcripts identified by DP\_GP. We tested our algorithm using published time-series RNA-seq data from root maturation zone samples of a model legume *Medicago truncatula* comprising different conditions and genotypes. In a comparison with the DeSeq2 package, DP\_GP discovered more genes with differential expression patterns. Our findings provide additional insight to time-course differential expression not previously described.

## **W720: Machine Learning and Artificial Intelligence in Genomics and Phenomics**

### **Determining Effective Dimensionality Reduction Tools for Genetic Data used in Deep Learning Models**

**Therin Young**<sup>1</sup>, Zihao Zheng<sup>2</sup>, Koushik Nagasubramanian<sup>3</sup>, Patrick S. Schnable<sup>2</sup> and Baskar Ganapathysubramanian<sup>4</sup>, (1)Iowa State University, Ames, IA, (2)Department of Agronomy, Iowa State University, Ames, IA, (3)Department of Electrical Engineering, Iowa State University, Ames, IA, (4)Department of Mechanical Engineering, Iowa State University, Ames, IA

The high dimensionality of genome data poses significant challenges (overfitting, high correlation among input data, need for multiple genotypes for training well, etc.) for applying machine learning models. Hence, there is increasing interest in reducing the dimensionality of the input data (i.e. SNP data) before feeding into deep learning models. In this study, we compared a variety of data representation methods for deep learning-based dimensionality reduction of genomic data. The reconstruction error of the genomic data was assessed to determine the best data representation for deep learning-based dimensionality reduction. In addition to autoencoders, the effectiveness of the data representations methods was assessed for classical dimensionality reduction methods such as PCA. This study can serve as a practitioner's guide to effective genomic data representation for deep learning applications.

## **W721: Machine Learning and Artificial Intelligence in Genomics and Phenomics**

### **GSForge: A Package for Condition-Specific Gene Set Selection, Comparison, Quality Assessment and Visualization for Traditional and Machine Learning Approaches**

**Tyler Biggs** and Stephen P. Ficklin, Dept of Horticulture, Washington State University, Pullman, WA

GSForge is a Python software package that assists researchers in the selection of gene sets with potential association to an experimental condition or phenotypic trait, which offers new potential hypotheses for gene-trait causality. Candidate gene sets are first identified outside of GSForge using Differential Gene Expression (DGE) analysis, Machine-Learning (ML) approaches, aggregated via knowledge from literature, network analysis or other approach. One or more gene sets from any number of approaches as well as gene expression data are provided to GSForge to explore the predictability of each gene set for a given trait. ML-based Quality Assurance (QA) functions provide scores for these predictions, and multiple gene sets from different sources can be compared. Additionally, to assist researchers unfamiliar with ML-based gene set selection, GSForge provides a gene set identification tool that uses Boruta with Random Forests for robust feature selection. To explore model parameter space, a Nextflow workflow can be used to scale to high-performance computing infrastructure. The GSForge Python library provides helpful data objects that house data using the NetCDF file type, allowing for import into other analytical packages such as R. Finally, GSForge provides tutorials and interactive notebooks. All source-code and resources are available at <https://github.com/SystemsGenetics/GSForge>.

## **W722: Maize**

### **Long-Read A188 Maize Genome Assembly and Short-mer Guided Error Estimation**

**Sanzhen Liu**, Kansas State University, Manhattan, KS

Genome editing tools provide great potential for the elucidation of gene functions and crop improvements. In maize, however, current genetic and genomic resources as well as transformation capacity are insufficient for the full utilization of editing tools. We generated Nanopore long sequencing reads, Illumina paired-end reads, BioNano physical mapping data, and a genetic map to produce a chromosome-level genome assembly of a highly transformable maize inbred line A188. An approach based on quantification of k-mer abundance in both Illumina reads and the assembly was developed to estimate base errors and the assembly completeness. Additionally, the *de novo* genome annotation was performed using data from long cDNA direct sequencing through the Nanopore platform and Illumina RNA-Seq of diverse tissues. Comparison between A188 and the maize reference B73 genome identified both conserved regions between the two genomes and prevalent copy number variation (CNV). At the whitecap (*wc*) locus, A188 contained approximately 10 times copies more than B73. High copies of the *wc* locus caused the elevation of gene expression, presumably altering the pigment of seeds therefore. The study demonstrated the effectiveness of Nanopore long reads in the assembly and annotation of a maize genome, which would be highly valuable for further understanding complexity of large genomes and gene regulation modulated by genome structure.

## **W723: Maize**

### **Pangenome of the Maize NAM Founder Inbreds**

**R. Kelly Dawe**, University of Georgia, Athens, GA, Matthew B. Hufford, Iowa State University, Ames, IA, Doreen Ware, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and Candice N. Hirsch, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

De novo genome assemblies of the 25 maize NAM founder inbreds and a new B73 (version 5) genome will have been publicly released. The genomes were assembled as a part of a single large project using a combined PacBio/Bionano approach, with the intent of making all data directly comparable. mRNA was sequenced from ten tissues for each inbred, and the annotations will have been released with browser support at MaizeGDB. The NAM assemblies have contig N50s ranging from ~8-40 Mb and scaffold N50s over 100 Mb. We achieved even better results in an inbred containing Abnormal chromosome 10 - including truly gapless chromosome assemblies - by combining PacBio, Nanopore and Bionano assemblies. Our early interpretations of the data, including the prevalence and impact of structural variation will be presented.

## **W724: Meiotic Recombination**

### **ASY1 acts as a Dosage-Dependent Antagonist of Telomere-Led Recombination *via* Crossover Interference in Arabidopsis**

**Ian Henderson**, University of Cambridge, Cambridge, United Kingdom

During meiosis, interhomolog recombination produces crossovers and non-crossovers and creates genetic diversity. Meiotic recombination frequency varies at multiple scales, with high telomeric recombination and suppressed centromeric recombination typical in many eukaryote genomes. During recombination, chromosomes are tethered as loops to a polymerized axis, which includes the ASY1 HORMA protein. Using chromatin immunoprecipitation we show an ascending telomere-centromere gradient of ASY1 that correlates strongly with REC8-cohesin. We mapped crossovers genome-wide in the absence of ASY1 and observe that telomere-led recombination becomes dominant, although with reduced efficiency. Surprisingly, *asy1/+* heterozygotes remodel crossovers towards telomere-proximal regions, at the expense of the pericentromeres. Telomeric recombination increased in *asy1/+* in distal regions where ASY1 and REC8 enrichment are lowest in wild type, explaining their sensitivity to ASY1 gene dosage. Interestingly, crossover interference is normal in *asy1/+*, whereas it is undetectable in *asy1* mutants. This reveals that ASY1 functions to antagonise telomere-led recombination and promote

spaced crossover formation via interference. Together this provides new insights into the role of the meiotic axis in patterning recombination frequency along telomere-centromere axes in plant genomes.

### **W725: Meiotic Recombination**

#### **Natural Variation in Recombination Frequency Reveals *SNI1* as a New Regulator of Meiotic Crossovers**

Longfei Zhu<sup>1</sup>, Maja Szymańska-Lejman<sup>1</sup>, Julia Dłużewska<sup>1</sup>, Alexandre Pele<sup>1</sup>, Charles Underwood<sup>2</sup>, Nadia Fernández-Jiménez<sup>3</sup>, Monica Pradillo<sup>3</sup>, Ian Henderson<sup>2</sup> and **Piotr A. Ziolkowski**<sup>1</sup>, (1)Adam Mickiewicz University, Poznan, Poland, (2)University of Cambridge, Cambridge, United Kingdom, (3)Universidad Complutense de Madrid, Madrid, Spain

The frequency and distribution of meiotic crossovers (COs) are tightly controlled, however variation in this process could be observed within both different organisms and populations of the same species. This variation can be used to identify genetic factors important for crossover formation, especially for its regulation. Previous mapping using two different *Arabidopsis* accessions, Col and Ler, allowed us to localize two QTL peaks indicating the presence of genes which can affect crossover frequency. One of them has been identified as *HEI10*, gene encoding E3 ligase involved in ZMM crossover control. Now, using extensive mapping, we found out that the second locus corresponds to *SNI1*. Indeed, *sni1* mutant demonstrates elevated CO rate as measured using fluorescent reporter lines, suggesting that it has an anti-recombinational role. Moreover, complementation of the mutant with both Col and Ler alleles of *SNI1* restores wild-type crossover levels. *SNI1* is known as a negative regulator of systemic acquired resistance (SAR) in plants. However, experiments involving a mutant of *EDS1*, a gene encoding an upstream factor for *SNI1*, as well as plant treatment with salicylic acid (major trigger in SAR) showed that its meiotic recombination phenotype is not related to SAR pathway. *SNI1* was characterized as a counterpart of NSE6, a component of SMC5/6 complex. It has been recently shown that SMC5/6 complex, through its SUMO ligase activity, regulates SGS1-TOP3A-RMI1 activity in budding yeast. In plants, inactivation of SGS1 homolog, RECQ4, leads to a strong elevation in class II crossover numbers. Therefore, we investigated genetic interference in the *sni1* mutant and observed its significant decrease, which supports a view that the protein is involved in class II crossover control. The meiotic phenotype of *sni1* was further characterized with the aid of cytological approaches. A number of abnormalities in both meiosis I and II were observed, which is consistent with SMC5/6 function during different stages of cell division. We propose that *SNI1* regulates the activity of SMC5/6 complex, which in turn affects RECQ/BLM1/SGS1 anti-recombinational function leading to modified crossover landscape.

This work is supported by NCN grants 2016/21/B/NZ2/01757 and 2016/22/E/NZ2/00455 to PAZ.

### **W726: Meiotic Recombination**

#### **Inter-Homolog Mitotic Recombination is Enhanced in Diatoms and Contributes to High Haplotype Diversity in Clonal Populations**

**Petra Bulankova**<sup>1</sup>, Ilse Vercauteren<sup>2</sup>, Cristina Osuna Cruz<sup>1</sup>, Emmelien Vancaester<sup>1</sup>, Denis Jallet<sup>3</sup>, Gust Bilcke<sup>1</sup>, Cock van Oosterhout<sup>4</sup>, Fayza Daboussi<sup>5</sup>, Koen Sabbe<sup>6</sup>, Wim Vyverman<sup>6</sup>, Klaas Vandepoele<sup>1</sup> and Lieven De Veylder<sup>1</sup>, (1)VIB-UGent Center for Plant Systems Biology, Ghent, Belgium, (2)VIB-UGent Center for Plant Systems Biology, Ghent, Belgium, (3)Toulouse Biotechnology Institute, Bio & Chemical Engineering, Toulouse, France, (4)School of Environmental Sciences University of East Anglia, Norwich, United Kingdom, (5)LISBP - INSA de Toulouse, Toulouse, France, (6)Department of Biology Research group Protistology & Aquatic Ecology, Ghent University, Ghent, Belgium

Homologous recombination (HR) entails the exchange of genetic material between two strands of DNA. In meiotic cells, HR involves the homologous chromosomes and is a source of new allele combinations. In contrast, the preferred substrate for HR in vegetative cells are sister chromatids. As mitotic recombination between homologous chromosomes can result in harmful loss of heterozygosity (LOH) and chromosome rearrangements, it is usually strongly suppressed in diploid organisms. Strikingly,

through quantifying haplotype diversity by NGS, we observed a rapid accumulation of multiple haplotypes in diatom (*Seminavis robusta* and *Phaeodactylum tricornutum*) cell cultures clonally propagated from single founder cell, suggesting a high frequency of recombination between homologous chromosomes. Next, by comparing the genome of mother and daughter *Phaeodactylum* cultures started from a single cell and separated by 30 mitotic divisions we could detect 9 LOH tracts in four out of nine daughter cell cultures, with size ranging from 300 bp to 42 kb. Based on these data we made a preliminary estimation that the mitotic recombination rate per cell division is approximately 11 – 16 times higher in diatoms than in *S. cerevisiae*.

Although harmful for multicellular organisms, we suppose that enhanced mitotic recombination can lead to rapid fixation of new advantageous mutations in single cellular organisms with a rare occurrence of sexual reproduction like diatoms. To test this hypothesis, we are currently using strains mutant for PtUMPS gene as a LOH read-out system to precisely measure the rate of mitotic recombination and to test the effect of different stresses.

We believe that diatoms can become a suitable model to answer questions about the mechanism of mitotic recombination. In the future we want to focus on understanding how mitotic recombination contributes to the adaptation, identify the preferred molecular pathway and get new insights on the recognition of homologous chromosomes in vegetative cells.

### **W727: Meiotic Recombination**

#### **Generating Haplotype-Resolved, Chromosome-Level Genome Assemblies using Single-Cell Genome Sequencing of Recombined Gametes**

**Korbinian Schneeberger**, Max Planck Institute for Plant Breeding Research, Cologne, Germany

The generation of genome assemblies of diploid organisms is challenged by the presence of two highly similar, but not identical chromosome pairs. An efficient solution for this problem is separating the whole-genome sequencing data into two read sets where one set includes reads from the paternal genome and one set includes reads from the maternal genome (using similarity to the respective genomes). Once this is achieved, the two haploid genomes can be assembled independently. This however requires knowledge on these genomes. In my presentation I will show how we solved this problem even without knowledge of the parental genomes. Sequencing the genomes of hundreds of pollen genomes derived from the focal individual helped to first generate a genetic map and then helped to assemble both chromosome sets independently on chromosome level. I will finish the presentation by showing how the generation of such high quality genome assemblies can help to analysis of heterozygous somatic mutations.

### **W728: Meiotic Recombination**

#### **A Genotype-Phenotype Association for Autopolyploid Meiosis Stabilisation**

Paul J. Seear<sup>1</sup>, Martin France<sup>1</sup>, Catherine Gregory<sup>1</sup>, Darren Heavens<sup>2</sup>, Roswitha Schmickl<sup>3</sup>, Levi Yant<sup>4</sup> and **James David Higgins<sup>1</sup>**, (1)University of Leicester, Leicester, United Kingdom, (2)Earlham Institute, Norwich, United Kingdom, (3)Charles University, Prague, Czech Republic, (4)University of Nottingham, Nottingham, United Kingdom

Whole genome duplication (WGD) is often associated with increased ecological fitness and adaptation to new biological niches. However, the doubled set of chromosomes can lead to complex meiotic configurations at meiotic metaphase I, thus causing sterility. We have previously shown that eight meiosis genes are under selection in the outbreeding autotetraploids *Arabidopsis arenosa* and *Arabidopsis lyrata*, as well as extensive interspecies bidirectional gene flow between them. Using a genotype-phenotype approach our analysis has revealed that the meiotic chromosome axis protein *Asynapsis 3* (*ASY3*) is the major determinant of meiotic stability in the autotetraploids. The beneficial *ASY3* protein possesses a tandem duplication (TD) of a serine-rich region upstream of the coiled coil that may destabilize the protein thus creating a hypomorph that is associated with a reduction of

multivalents and distalization of chiasmata in stable bivalents. In addition, 320 alleles were identified from the eight genes, harboring evidence of gene conversion and widespread allelic chimerism, likely generated through meiotic recombination in *ASY1*, *PDS5b*, *PRD3*, and *ZYP1a/b* between alleles originating from both species, as well as between ploidies. We therefore suggest that these rapidly evolving genes provide precise control over meiotic recombination in the autopolyploids, the very process that gave rise to them.

## **W729: NCBI Genome Resources**

### **NCBI Wants Your Sequence Data! How Do I Get It There?**

**Ilene Mizrachi**, National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD and Karen Clark, National Center for Biotechnology Information (NCBI/NLM/NIH)

GenBank and the Sequence Read Archive (SRA) at NCBI are open, comprehensive archives of nucleotide sequence data that are used to make scientific discoveries. Deposition of sequence data to these archives is often required as part of the publication process and allows for data reuse. Recent enhancements to data submission and processing have made it easier to submit reads, assemblies and annotations to NCBI. You can now perform simple updates to your SRA submission directly. As part of a continuing effort to make submissions easier, we are enhancing our genome and transcriptome submission wizards. This talk will discuss some of the recent improvements to submissions and detail some of the validations that are performed before genomes are accepted.

## **W730: NCBI Genome Resources**

### **Annotation of Eukaryote Genomes at NCBI**

**Jinna C. Hoffman**<sup>1</sup>, Francoise Thibaud-Nissen<sup>1</sup>, Terence D. Murphy<sup>1</sup> and the Eukaryotic Genome Annotation Team<sup>2</sup>, (1)NCBI/NLM/NIH, Bethesda, MD, (2)National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

The NCBI Eukaryotic Genome Annotation Pipeline has been used to annotate over 575 organisms, including plants and animals of agricultural importance. The pipeline uses a modular framework for the execution of all annotation tasks from the retrieval and alignment of experimental evidence to the prediction of genes, and loading of RefSeq-accessioned annotation products to public databases. The annotation of an assembly takes one to two weeks to complete, depending on the size of the genome and the amount of same-species or close cross-species data in public databases.

This talk will provide a high-level view of the annotation pipeline. We will demonstrate how the quality of an annotation is assessed, and how the genome assembly and the availability of short and long transcriptomics reads affects the accuracy and the richness of the annotation results. New features of the annotation process will also be discussed, such as the use of a different aligner (minimap2) for long transcriptomics reads to better handle low accuracy sequences. We will show how to download and visualize the annotated genomes along with data from other sources in the Genome Data Viewer.

To see all eukaryotes annotated by the NCBI Eukaryotic Annotation Pipeline, and request the annotation of your next assembly, visit: [http://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/all/](http://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/)

## **W731: NCBI Genome Resources**

### **Accessing Homologous Gene Datasets at NCBI**

**Nuala A. O'Leary**<sup>1</sup>, Peter Meric<sup>1</sup>, Greg Schuler<sup>1</sup>, Vichet Hem<sup>1</sup>, Xuan Zhang<sup>1</sup>, Robert Falk<sup>1</sup>, William Anderson<sup>1</sup>, Kelly McGarvey<sup>1</sup>, Kurtis Haro<sup>1</sup>, Anne Ketter<sup>1</sup>, Wratko Hlavina<sup>1</sup>, Terence D. Murphy<sup>1</sup>, Valerie A. Schneider<sup>1</sup> and the Eukaryotic Genome Annotation Group, (1)NCBI/NLM/NIH, Bethesda, MD

NCBI recently added a new way for users to find evolutionarily related genes within and across organisms represented in the NCBI RefSeq dataset. The goal of this new service is to facilitate comparative genomic research by allowing users to easily access sequence data as well as visualization and analysis tools for homologous gene sets from the increasing number of annotated eukaryotic reference genomes. This new service returns two types of related genes sets. The first is a set of vertebrate orthologous genes calculated by NCBI Gene based on a combination of protein sequence similarity, local synteny information, and manual assertion. The second provides a set of genes that share protein architecture with the orthologous gene set and includes genes from all metazoans as well as from selected plant, fungal and protist species. Both resources display genes in rows that can be expanded to reveal more detailed information on transcripts and proteins, as well as links to the NCBI genome browser and InterPro protein families. This presentation will describe how to access these homologous gene sets and introduce NCBI's newest efforts to improve downloading of whole genome sequence and annotation datasets.

### **W732: NCBI Genome Resources**

#### **The New Pubmed Is Here!**

**Kathi Canese**, NCBI/NLM/NIH, Bethesda, MD

Please join us for a presentation on the updated version of PubMed. The new PubMed is richly featured, including Advanced Search, Search details, Search history, filters, My NCBI, links from MeSH, and more. The new PubMed boasts some great new display, navigation and output features in a truly responsive design that facilitates mobile access. Find highly relevant articles more easily using the Best Match sort, now the default sort order in PubMed. Best Match uses a state-of-the-art machine learning algorithm that is trained on aggregated user searches. Additionally, improvements to retrieval include enhanced synonymy, addition of plural forms, better British/American translations, and unlimited truncation. The new PubMed will be the default system in early 2020.

### **W733: NCBI Genome Resources**

#### **Taxonomy Lookup; Data Retrieval: How to Find and Stream Genomic Data in the Cloud!**

**Ben Busby**, NCBI/NLM/NIH, Bethesda, MD

The thousand plant genomes (<https://www.nature.com/articles/s41586-019-1693-2>) project has recently analyzed several hundred whole genomes and over a thousand plant transcriptomes, making all of the data accessible in the NCBI SRA, and adding processed reference data to appropriate repositories CyVerse and Zenodo. Seminal findings have been made in expansion of transcription factor gene families. We'll show you how to use a variety of NCBI databases and tools to further analyze this data on the cloud in an efficient manner! As an example, we'll look at presence and expression of family members involved in stress responses in major agricultural crops. Lessons learned here should be exemplary for leveraging SRA in the cloud across a wide variety of biological questions.

### **W734: New Approaches for Developing Disease Resistance in Cereals**

#### **Exploring Leaf Rust Resistance in *Aegilops tauschii* using Association Genetics Coupled with Resistance Gene Capture**

**Sreya Ghosh**<sup>1</sup>, Sanu Arora<sup>1</sup>, Kumar Gaurav<sup>1</sup>, Burkhard Steuernagel<sup>2</sup>, Oadi Matny<sup>3</sup>, Dragan Perovic<sup>4</sup>, Albrecht Serfling<sup>5</sup>, Macarena Forner-Martinez<sup>6</sup>, Frank Ordon<sup>4</sup>, James A Kolmer<sup>7</sup>, Brian J. Steffenson<sup>8</sup> and Brande B.H. Wulff<sup>1</sup>, (1)John Innes Centre, Norwich, United Kingdom, (2)The Sainsbury Laboratory and John Innes Centre, Norwich, United Kingdom, (3)Department of Plant Pathology, University of Minnesota Twin Cities, St. Paul, MN, (4)JKI, Julius Kuehn Institute, Quedlinburg, Germany, (5)Institute for Resistance Research and Stress Tolerance, Julius Kühn-Institut, Germany, (6)John Innes Centre, United Kingdom, (7)Cereal Disease Lab, United States Department of Agriculture - Agriculture Research Service, (8)Department of Plant Pathology, University of Minnesota, Saint Paul, MN

Leaf rust of wheat, caused by the obligate biotrophic pathogen *Puccinia triticina*, is the most common among the three rusts of wheat worldwide. Wheat leaf rust epidemics can cause significant yield losses, and the most economical and environmentally sound method to control them is breeding for resistance to this pathogen. There are more than 70 leaf rust resistance genes that have been mapped and designated in wheat, but the sequences of only a handful of genes have been identified so far. Many of these genes, including some of the cloned ones, have been transferred into cultivated wheat from the diploid wild wheat progenitor *Aegilops tauschii*. Access to the sequences of such resistance genes benefit breeders by allowing them to design gene-specific markers to quickly and cheaply track them in their breeding programs. It also helps researchers understand the genetic architecture of leaf rust resistance, and how it can be engineered to stay one step ahead of a rapidly evolving pathogen. Traditional map-based or mutational genomics-based gene cloning approaches are time and labour intensive, as they involve creating structured populations. A novel method for resistance gene cloning, which combines association genetics with resistance gene enrichment sequencing (AgRenSeq), can be used instead to rapidly obtain candidate genes for resistance in a diversity panel of wild wheat. Here, it will be described how AgRenSeq has been used to identify candidate genes for leaf rust resistance on chromosomes 1D and 2D using an *Aegilops tauschii* panel. Furthermore, it will be discussed how this data has been used to explore the candidate genes' frequencies in the panel to dissect the genetic architecture of leaf rust resistance against several races from North America and Europe.

### **W735: New Approaches for Developing Disease Resistance in Cereals**

#### **Field Transcriptomics Identifies a Novel Wheat Susceptibility Factor which Modulates Resistance to Rust**

**Pilar Corredor-Moreno**, Phoebe E Davey and Diane GO Saunders, John Innes Centre, Norwich, United Kingdom

Wheat yellow rust, caused by the obligate biotrophic fungus *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a major threat to wheat production worldwide and can lead to total crop loss when untreated. Identifying host pathways targeted by the pathogen is crucial to get a better understanding of the host mechanisms involved in the defence response. We used transcriptomics data obtained from *Pst*-infected field samples collected from wheat varieties showing different susceptibilities to *Pst*. Differential expression analysis identified changes in amino acid metabolism that led to the identification of *TaBCAT1*, a gene encoding a wheat branch-chain aminotransferase. *TaBCAT1* expression peaks at 24 hours post-inoculation in a susceptible interaction with modern *Pst* UK isolates while the expression was reduced in a resistant interaction. To further explore the function of this protein in the defence response, we developed *Tabcat1* tetraploid double mutant lines using the wheat TILLING (Targeting Induced Local Lesions in Genomes) population. *Tabcat1* mutant lines showed a dramatic reduction in susceptibility to *Pst*, suggesting a potential role as a wheat susceptibility factor. Mutant lines exhibited constitutive upregulation of pathogenesis-related (PR) genes in the absence of infection and this observation was replicated in hexaploid wheat when *TaBCAT1* was silenced using the barley stripe mosaic virus (BSMV). *Tabcat1* mutant lines also accumulated salicylic acid (SA) in the absence of pathogen infection, indicating that mutant lines have a constitutively activated defence machinery. These results suggest that *TaBCAT1* coordinates the activation of SA-mediated defence responses in wheat and could be exploited in resistance breeding to eliminate known susceptibility from commercial cultivars.

### **W736: New Approaches for Developing Disease Resistance in Cereals**

#### **Use of a High-Quality Wheat Genome Assembly Allows Cloning of a Novel Non-NLR Leaf Rust Resistance Gene**

**Markus C. Kolodziej**<sup>1</sup>, Jyoti Singla<sup>2</sup>, Thomas Wicker<sup>1</sup>, Javier Sanchez-Martin<sup>3</sup>, Helen Zbinden<sup>2</sup>, Julien Gronnier<sup>2</sup>, Simon G. Krattinger<sup>4</sup>, Hana Šimková<sup>5</sup>, Miroslava Karafiátová<sup>5</sup>, Jaroslav Doležel<sup>5</sup>, Curtis J. Pozniak<sup>6</sup> and Beat Keller<sup>1</sup>, (1)University of Zurich, Zurich, Switzerland, (2)University of Zurich, Switzerland, (3)University of Zurich, Zollikerstrasse 107, Switzerland, (4)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, (5)Institute of Experimental Botany, Centre of the

Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (6)University of Saskatchewan, Saskatoon, SK, Canada

Leaf rust (Lr) is a severe disease of bread wheat. Only a few of the 79 Lr resistance genes currently described in the gene pool of wheat and its progenitors have been cloned so far. All until now cloned race-specific Lr resistance genes (Lr1, Lr10, Lr21) thus encode for nucleotide binding site-leucine rich repeat immune receptor (NLRs). Here, we report the cloning of the race-specific seedling stage resistance gene *Lr14a* from wheat. We used ethyl methane sulfonate (EMS) to mutagenize the DNA of the *Lr14a* carrying winter wheat line *ArinaLrFor*. Three mutant lines, phenotypic susceptible for Lr, were selected and subjected to the MutChromSeq approach to identify mutated candidates for the resistance mediating gene. In this approach, the flow sorted chromosomes 7B of the three mutants were short read sequenced. We aligned the reads to a synthetically fragmented assembly of the 7BL pseudomolecule of a new *ArinaLrFor* wheat reference genome, which is part of the 10+ Wheat Genome Project. Sequence comparisons between the reference and the three mutants lead to the identification of single nucleotide polymorphisms (SNPs) in the coding sequence of one gene in all three mutants. This *Lr14a* candidate gene encodes a protein, which is, in contrary to the other race-specific resistance genes so far cloned, not an NLR. We used Virus Induced Gene Silencing (VIGS) to validate that the resistance phenotype is caused by the identified candidate gene. When targeting the gene via VIGS, infected *ArinaLrFor* plants show a loss of the resistance against leaf rust. We envision that functional characterization of Lr14a as well as its interaction with other proteins will uncover new mechanisms of immunity in wheat and other crops.

### **W737: New Approaches for Developing Disease Resistance in Cereals**

#### ***Lr67* Resistance Beyond Wheat: A Broad-Spectrum Disease Resistance Gene Functions in Multiple Crop Species**

**Katherine Dibley**<sup>1</sup>, Ricky Milne<sup>2</sup>, Wendelin Schnippenkoetter<sup>2</sup>, Jayakumar Bose<sup>3</sup>, Srinivas Belide<sup>1</sup>, David Jordan<sup>4</sup>, Anthony Ashton<sup>1</sup>, Peter Ryan<sup>2</sup>, Bo Zhou<sup>5</sup>, Steve Tyerman<sup>3</sup> and Evans Lagudah<sup>2</sup>, (1)CSIRO Agriculture & Food, Australia, (2)CSIRO Agriculture & Food, Canberra, Australia, (3)School of Agriculture, Food and Wine, The University of Adelaide, Australia, (4)The University of Queensland, St Lucia, QLD, Australia, (5)International Rice Research Institute, Los Banos, Philippines

Plant diseases resulting from fungal pathogens are a major constraint on global crop production. Plant pathogen resistance genes are a useful tool to limit reliance on agrochemicals for disease control. The majority of resistance genes cloned so far are pathogen or race-specific, giving effective protection that is however prone to being broken down by pathogen evolution. In contrast, a smaller group of genes confer partial, durable, broad-spectrum resistance to multiple plant pathogens. The wheat *Lr67* is one such gene, providing partial resistance to multiple biotrophic pathogens - stem, leaf and yellow rusts and powdery mildew - along with a leaf tip necrosis (Ltn) phenotype. *Lr67* is encoded by a hexose transporter gene from the STP13 family, with two amino acid substitutions differentiating the susceptible and resistant alleles. Transformation of *Lr67* into other cereal crops, including barley, rice and sorghum, provides partial disease resistance against the recipient crop species own adapted pathogens. However, negative pleiotropic effects, including extensive leaf tip necrosis, early leaf senescence and reduced vigour, are also observed. Here, we consider the trade-off between disease resistance and these negative effects and explore strategies to help uncouple these. Recent data examining the molecular function of *Lr67* will also be discussed, along with an *in vitro* screening system developed to rapidly identify novel promising “*Lr67*-like” gene variants for further *in planta* studies.

### **W738: New breeding technologies: Prospects and regulatory hurdles**

#### **Introduction to the Topic**

**Janina Metje**, Julius Kuehn-Institut, Quedlinburg, Germany

New Plant Breeding Techniques are on the rise, here we give a short overview.

## **W739: New breeding technologies: Prospects and regulatory hurdles**

### **Genome Editing; A Global Perspective**

**Thorben Sprink**, Julius Kühn Institut, Quedlinburg, Germany

Creating high value cultivars is one of the main challenges for breeders these days. For this purpose a whole toolbox of techniques can be used. These tools include conventional breeding techniques like simple crossing which has been used since centuries but frequently reached its limits to access sufficient natural variation. Mutation breeding using radiation or chemicals enabled breeders to induce undirected novel but randomly distributed mutations. The drawback is its need of extended backcrossing. Genetic engineering using transgenesis opened the field for specific changes in the genome conferring traits of interest even between species. But these changes are still undirected integrated into genomes and the social acceptance of cultivating transgenic plants is low in many countries. Recent developments in genome editing e.g. especially using clustered regularly interspaced short palindromic repeats (CRISPR) fused with CRISPR associated proteins (e.g. Cas9), enable to change the genome of plants in a directed, trait AND site specific way. The approach is time saving and in some cases allows to avoid creating (intermediate) transgenic lines for breeding.

These new techniques of "Genome Editing" have already been successfully applied to more than 60 different crops and model plants worldwide. Many studies have been basic research, testing and developing the technologies, but there are also plenty of applications to improve agronomical relevant traits of crops e.g. agronomical value, increasing biotic- and abiotic stress tolerance, food and feed quality. Additionally, first varieties created by using genome editing have already been released to national and international markets.

However, regulation of these new genome editing techniques is not harmonized between continents and countries and may hinder international trade in the future. Many American countries established a straight forward regulatory system and application procedure, which defines such genome editing products as non-GMOs if a transgene is not present in the genome. In contrast, the European union has a restrictive regulatory system for GMOs in place which also applies for genome editing as verified by the European court of justice. In this presentation, the newest products of genome editing in agriculture will be presented and a genome editing atlas will be presented in which developments in the Genome Editing field are projected on a world map to identify the global development of GE.

## **W740: New breeding technologies: Prospects and regulatory hurdles**

### **Novel Developments in Gene Drive Technologie and Regulation**

**Werner Schenkel**, Federal Office of Consumer Protection and Food Safety Germany, Berlin, Germany

It is clear that organisms containing engineered gene drives are genetically modified organisms (GMO) as they inevitably contain genetic elements foreign to the species. These organisms are subject to international agreements like the Cartagena Protocol on Biosafety to the Convention on Biological Diversity and to the national legislation of all Parties to the Protocol and Non-Parties having GMO legislation. Thus development and handling of organisms containing an engineered gene drive will have to be risk assessed. This is especially true for any intentional release of such organisms but also for preceding work done in laboratories.

The distinctive feature of the gene drive technology, namely the intentional spread of a genetic element into wild populations, set it apart from classical GMO that have been developed and approved for the intentional release up to now.

The question arises, if current regulatory frameworks and risk assessment schemes are suited to handle this new class of GMO. The potential for persistence and invasiveness in natural ecosystems has always been a crucial part of the risk assessment of GMO. Therefore, it is principally possible to handle the distinctive feature of gene drive organisms, the increased potential for spread of a genetic element to

whole populations, within current risk assessment schemes. However, the special mode of inheritance as well as fact, that generally populations in natural ecosystems are targeted will require the taking into account of different and probably additional data compared to classical GMO. As gene drives work as a system where the receiving population is an integral part, detailed information the genetic make-up of this population as well as its role in the ecosystem will be crucial for an exhaustive risk assessment. Approval or deregulation of GMO for experimental release or placing on the market is always based on a case by case assessment. As additional information can be requested as necessary during the approval process, organisms containing engineered gene drives can be safely regulated under current legal frameworks.

#### **W741: New breeding technologies: Prospects and regulatory hurdles**

##### **New Biotechnology and Governance Issues - Regulatory Clarification of Gene-Edited Products in Japan**

**Makiko Matsuo**, the University of Tokyo, Tokyo, Japan

With the advent of new biotechnology, namely the gene-editing, research and development of this technology is flourishing. Various types of application in the field of food and agricultural sector are now rapidly advancing in Japan – examples of the products under development include, high yielding rice, myostatin knocked out fish, and tomato etc. Gene-edited techniques can produce variety of products, from the so-called SDN1, which is similar to conventional breeding, to SDN3, equivalent to conventional GM. Since many of the products currently under development are SDN1, there was a need for clarification of the regulatory handling of such products.

The examination of the regulatory status of gene-edited product accelerated when the Japanese government released “Integrated Innovation Strategy” in 2018 to instruct relevant ministries to work on this matter. In Japan, environmental and biodiversity impacts are mainly managed by the Ministry of the Environment (MOE) under the (Japanese domestic) Cartagena Act, and food safety aspect of such product is managed by the Ministry of Health, Labor and Welfare (MHLW) under the Food Sanitation Act.

In 2019, after intensive discussion, the status of the products derived from gene-editing technology under the both laws, Cartagena Act and Food Sanitation Act, was clarified but there were slight differences in the regulatory scope and handling based on different grounds (MOE and MHLW). Of note is that both ministries established reporting mechanisms for the products that are out of the scope of regulation – MOE set up information provision mechanism and MHLW established prior consultation and notification procedure. The summary information of such products will be posted on the government website. In addition, the Consumer Affairs Agency also clarified the handling of such products for labeling.

The presentation will introduce the details of the way how regulatory status was clarified in Japan, the governance issues as well as the lessons learnt from Japanese experience. It also discusses the implication of the issues in the international context.

#### **W742: New breeding technologies: Prospects and regulatory hurdles**

##### **Genome Editing -Current Situation in Norway-**

**Trine Hvoslef-Eide**, Norwegian University of Life Sciences, As, Norway

#### **W743: Next Generation Genome Annotation and Analysis**

##### **Making Sense of a New Assembled Genome: Finding and Fixing Genome Annotation Errors through Community Curation Approaches**

David Micklos, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY and **Cristina F. Marco**, DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Without high quality gene models, scientists won't be able to design experiments targeting their genes of interest. Traditionally validating those gene models require manual curation. This is a labor-intensive and time-consuming process in which one or a few individuals evaluate and correct the computational predictions by using all the available evidence they can find.

Before the manual curation takes place, automated gene finders identify the parts of a sequence that encodes genes, regulatory sequences and repetitive elements. These pipelines are becoming faster as the algorithms improve and more accurate as they are able to incorporate more biological data to use in their predictions. But they still make errors.

We tested two distinct approaches to identify mispredictions from automated gene finders: the quality values generated from a gene annotation pipeline: MAKER-P, and the alignments between the translated protein sequences and its homologs across species by the Gramene gene tree visualizer.

We selected a subset of genes from the most recent maize reference genome annotation that was analyzed by a group of students. Our results showed that all the gene models analyzed using these two methods had errors, and gave the students the opportunity to correct them and support the community curation of a very relevant eukaryotic genome. Ideally this method could be used as a means to let students and even citizen scientists participate in the annotation of any sequenced eukaryotic genome.

## **W744: Next Generation Genome Annotation and Analysis**

### **Using Open-Source Software at SIMRbase to Share Genomic Data with the Sea Lamprey Community**

**Sofia Robb**, Stowers Institute for Medical Research, Kansas City, MO

SIMRbase (<https://simrbase.stowers.org>) is a genome database and browser that houses genomic and transcriptomic data for a variety of organisms used in the research programs at Stowers Institute for Medical Research (SIMR). While much of the data is unpublished and only accessible by SIMR members and their collaborators the data for one organism is open to the world, the sea lamprey, *Petromyzon marinus*. The drive to create SIMRbase was two-fold: the first priority was to have intuitive tools available for researchers to browse and search their genomic data as it is generated, and secondly, to have an expandable infrastructure in place for those maintaining the system to easily add and maintain organisms as they are sequenced and annotated. This has worked well for the sea lamprey community who have been using SIMRbase for browsing their newly sequenced germline genome assembly and experimental data aligned to the genome like RNAseq reads, for performing gene searches, for managing manual gene annotation efforts, and for sequence similarity searches using BLAST.

SIMRbase is constructed using a collection of open-source tools, Tripal, CHADO, JBrowse, and Apollo. Tripal, a Drupal, module is used by SIMRbase to access data stored in CHADO. CHADO is a relational database schema that stores genes and associated data such as genomic location, BLAST hits, publications, sequence data, and GO terms. Tripal creates gene pages which includes gene accessions and name, sequences, homology information, protein domains, and other related data. The layout and content of the pages are customizable. SIMRbase uses Tripal extension modules for loading MAKER gene annotations and precomputed BLAST output to CHADO, a NCBI-BLAST interface for BLAST searches against SIMRbase sequence databases, and gene keyword search tools. SIMRbase employs JBrowse for the genome browsers and Apollo for manual gene curation. The sea lamprey community is the first of many to utilize SIMRbase.

## **W745: Next Generation Genome Annotation and Analysis**

### **Approaches and Tools to Improve the Identification of Protein-Coding Regions in Large and Complex Genomes**

**Jill L. Wegrzyn**<sup>1</sup>, Sumaira Zaman<sup>2</sup>, Madison Caballero<sup>1</sup>, Jeremy Bennett<sup>3</sup>, Peter Richter<sup>1</sup>, Alyssa M. Ferreira<sup>1</sup> and Susan McEvoy<sup>1</sup>, (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)Department of Computer Science and Engineering, University of Connecticut, Storrs, CT, (3)University of Connecticut, Storrs, CT

Initiatives such as the Earth BioGenome Project and the Open Green Genomes will sequence thousands of species in the next few years. Recent improvements in sequencing technology, optical mapping, and chromosome conformation capture resulted in improved accuracy and contiguity of genome assemblies across the tree of life. These technological advancements enabled the completion of increasingly large and complex genomes, several exceeding 10 Gbp in length. These achievements have unfortunately not been met with the same improvements on the genome annotation side which remains a challenging and error-prone process for large genomes. The challenges associated with the annotation of protein-coding regions are numerous, but include repetitive content, associated large number of pseudogenes, and non-canonical gene structures. We will present two packages that assist with these challenges that used downstream of existing genome annotation packages, such as MAKER or BRAKER. The first, Gene Filtering, Analysis, and Conversion (gFACs), is developed to filter, analyze, and convert predicted gene models and alignments. The software operates across a wide range of alignment, analysis, and gene prediction files with a flexible framework for defining gene models with reliable structural and functional attributes. gFACs supports common downstream applications, including genome browsers, and generates extensive details on the filtering process. The second, EnTAP (Eukaryotic Non-Model Transcriptome Annotation Pipeline), was designed to improve the accuracy, speed, and flexibility of functional gene annotation for de novo assembled transcriptomes or predicted gene models. This package focuses on providing information for species without extensive genomic resources by fully integrating gene family comparisons. EnTAP and gFACs can be used together to assist with filtering the excessive number of genes that are often predicted against large and complex genomes. Finally, we will introduce a new package, EASEL, which is integrating machine learning approaches to provide more robust genome annotations.

## **W746: Next Generation Genome Annotation and Analysis**

### **High-Quality Annotation of the Tufted Duck genome**

**Ralf Christian Mueller**<sup>1,2</sup>, Lel Eory<sup>3</sup>, Richard Kuo<sup>3</sup>, Amanda Warr<sup>3</sup>, Jacqueline Smith<sup>3</sup>, Patrik Ellström<sup>4</sup>, Josef D. Järhult<sup>4</sup>, Björn Olsen<sup>4</sup>, Mahmoud M. Naguib<sup>5</sup>, Marcela Uliano-Silva<sup>6</sup>, Erich D. Jarvis<sup>7</sup>, Olivier Fedrigo<sup>8</sup> and Robert H. S. Kraus<sup>1,2</sup>, (1)Max Planck Institute of Animal Behavior, Radolfzell, Germany, (2)Department of Biology, University of Konstanz, Konstanz, Germany, (3)The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, (4)Department of Medical Sciences, Zoonosis Science Center, Uppsala University, Uppsala, Sweden, (5)Department of Medical Biochemistry and Microbiology, Zoonosis Science Center, Uppsala University, Uppsala, Sweden, (6)Wellcome Sanger Institute, Hinxton, United Kingdom, (7)The Rockefeller University, Vertebrate Genome Laboratory, and HHMI, New York, NY, (8)Vertebrate Genome Laboratory, The Rockefeller University, New York, NY

Wild and domesticated birds are exposed to many pathogens, and diseases can easily be carried over great distances by migrating species. Waterfowl, most prominently ducks, are not only economically important livestock but also serve as a reservoir for avian influenza virus (AIV). However, not all species of ducks are affected the same during an AIV outbreak, and this effect also differs between seasonal virus strains. Mallards and tufted ducks showed markedly different susceptibility during an outbreak of highly pathogenic AIV (H5N8) in Europe in 2016. While mallards can carry the virus without getting seriously ill, tufted ducks are very sensitive to infection and usually die within hours. In this study, we aim to understand if there is a genetic basis to resistance, and therefore produced a fully annotated, high-quality reference genome of the tufted duck to achieve this goal.

Often, immune genes are poorly annotated, due to lack of information on the correct splice structure due to their high variability, and recent gene family expansions are commonplace. While Illumina RNA-Seq is known to produce high-quality reads (error rate < 1%) with high sequencing depth, it also suffers from

the lack of the ability to sequence regions of low complexity or high GC content. The assembly of short reads can thus introduce ambiguities or fail to identify very similar genes and gene duplications. Full-length transcriptomes may help to identify the correct isoforms of these genes: Pacific Biosciences Iso-Seq reads hardly suffer from assembly errors or ambiguity, although the error rate is higher (~ 15%) as compared to RNA-Seq.

We sequenced RNA from six different tufted duck tissues on two different platforms: Illumina and PacBio. The genome was sequenced and assembled using the VGP 1.6 pipeline (<https://vertebrategenomesproject.org/>). In the current phase of this project, we are contrasting PacBio's full-length transcriptome with Illumina's short-read RNA-Seq technology. Additionally, we compare the results with a hybrid pipeline in which low-quality but highly contiguous long reads are corrected with high-quality but fragmented short reads.

## **W747: Next Generation Genome Annotation and Analysis**

### **Fastqforward: Whole Genome Sequencing Analysis in Under 10 Minutes**

**Carson Holt**, Department of Human Genetics, University of Utah, Salt Lake City, UT

Advances in sequencing technology are making whole genome analysis commonplace in research and clinical settings. However, data processing is still a major hurdle overwhelming many groups and stalling research progress.

To address this issue, we developed FastQForward a rapid analysis pipeline that streamlines the processing of data returned from sequencing centers. It takes FASTQ files as input, and produces aligned BAM and VCF files suitable for downstream statistical analysis. FastQForward distributes sequencing alignment, data polishing steps, and variant calling in a highly parallelized manner both within a single machine as well as across multiple machines in the cloud or on academic computer clusters. The pipeline is designed to be easy-to-use and can process a 30x coverage human whole genome sample in as little as 8 minutes using 56 nodes and 1232 CPU cores on a compute cluster compared to 24 hours or more for similar software analysis pipelines that run on single multicore servers.

Current applications of FastQForward have included prioritizing individual samples for quick diagnostics in clinical settings, exploring population structure in model organisms, and even more technical uses like rapidly exploring alternate processing settings for troublesome samples (optimal settings can be discovered in minutes rather than days).

FastQForward's ability to scale across all available resources makes it suitable for projects of all sizes. By rapidly generating results, FastQForward allows researchers to push past technical barriers and move projects forward with greater efficiency.

## **W748: Non-coding RNA**

### **Identification of Conserved and Novel Characteristics of Salt-Responsive Long Non-Coding RNAs in Brassicaceae**

**Kyle R Palos**, University of Arizona, Tucson, AZ

Soil salinization, leading to the loss of arable land, is a significant issue impacting approximately 20% of cropland. The development of technologies to maintain high crop yields in the face of soil salinization, climatic shifts, and the increasing human population is an urgent priority for plant scientists. Recently, many groups have identified long, non-protein coding RNAs (lncRNAs) as important genomic factors mediating stress responses in animal and plant systems. To better understand the role of lncRNAs in mediating prolonged salt stress, we developed a salt stress protocol and implemented it across three model species in the plant family Brassicaceae: 1) *Arabidopsis thaliana*, a salt sensitive species; 2) *Eutrema salsugineum*, a stress tolerant halophyte; and 3) *Camelina sativa*, an emerging stress tolerant biofuel crop. We hypothesized that *Eutrema salsugineum* and *Camelina sativa* might either encode

unique suites of lncRNAs or exhibit novel lncRNA expression patterns underlying their native stress tolerance. Currently we are generating long read RNA sequences in the three model organisms for plants grown under control and salt stress conditions. These data will provide a more complete picture of the high confidence, salt-responsive lncRNAs encoded in these species, along with more accurate documentation of splice variants. Moreover, long read RNAseq data allow the expression characteristics of highly similar paralogous lncRNAs to be more readily distinguished, an extremely challenging task with short read sequencing data, and an important consideration when assessing lncRNAs from the hexaploid *C. sativa*. Altogether, our approach provides insight into the contributions of both novel and conserved nuclear lncRNA expression patterns during salt stress in Brassicaceae, an important step in developing technologies capable of mitigating stress in plants cultivated in salinized cropland.

## **W749: Non-coding RNA**

### **Accurate Characterization of Expression and Alternative Splicing in Arabidopsis for Protein Coding and Long Non-Coding RNAs**

**Runxuan Zhang**<sup>1</sup>, Wenbin Guo<sup>2</sup>, Nikoleta Tzioutziou<sup>2</sup>, Juan Carlos Entizne<sup>2</sup>, Cristiane Calixto<sup>2</sup>, Allan James<sup>3</sup>, Csaba Hornyik<sup>1</sup>, Hugh Nimmo<sup>3</sup> and John Brown<sup>2</sup>, (1)The James Hutton Institute, Dundee, United Kingdom, (2)University of Dundee, Dundee, United Kingdom, (3)University of Glasgow, Glasgow, United Kingdom

Understanding the current limitations of RNA-seq is crucial for reliable analysis. We have developed several computational resources, methods and tools to address the challenges of RNA-seq data analysis, with the emphasis on plant species. We have carried out:

- A BBSRC funded project to construct an automated pipeline with multiple assemblers to capture the diversity of transcripts from different sources and technologies and stringent filters to construct a comprehensive Reference Transcript Dataset (RTD) for plants. Extensive experimental validation showed that RTDs constructed using our method outperform other available transcriptomes in RNA-seq analysis in quantification accuracy [1][2].
- A cutting-edge pipeline (*3D RNA-seq*) [3] for differential gene expression and alternative splicing analysis. *3D RNA-seq* incorporates the state-of-the-art methodologies while remaining simple and rapid. It allows (lab) biologists with no programming skills to perform a complete differential expression analysis of RNA-seq data in 3 days.

These tools/methods enabled the discovery of massive and rapid expression and AS responses to cold in Arabidopsis and identification of hundreds of genes with very early changes in expression/AS, including numerous novel cold-responsive transcription factors and splicing factors/RNA binding protein genes [4]. We also demonstrated cold induced changes in expression as well as AS of pri-miRNAs and lncRNAs [5].

[1] Zhang et al. (2017) Nucleic Acid Research, 45 (9): 5061-5073; [2] Flores et al (2019) BioRxiv <https://doi.org/10.1101/638106>; [3] Guo at al. (2019) BioRxiv. <https://doi.org/10.1101/656686>; [4] Calixto et al. (2018) Plant Cell, 30(7):1424-1444; [5] Calixto et al. (2019) Front. Plant Sci. 10:235

## **W750: Non-coding RNA**

### **Functional Annotation of Non-Coding Genomic Regions of Non-Human Primates**

**Naoki Hirose**, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

To understand functional non-protein-coding regions in genomes, the FANTOM5 Consortium has produced a series of genome-wide maps of human and mouse promoters, enhancers, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) expressed in various samples including the tissues (Forrest et al. 2014; Andersson et al. 2014; Arner et al. 2015; Hon et al. 2017; De Rie et al. 2017). CAGE (Cap Analysis of Gene Expression) was used to detect capped-5' ends of RNAs in a nucleotide resolution, in addition to RNA-seq and small RNA-seq. Despite such functional annotations in the both species, non-

coding regions are less conserved in the human and the mouse genomes. Therefore, it has been obstructive to thoroughly explore evolutionary landscapes of the functional non-coding regions of mammalian genomes.

To change this situation, we are producing an atlas of functional genomics of non-human primates, crab-eating macaques (*Macaca fascicularis*) and common marmosets (*Callithrix jacchus*). We applied CAGE, RNA-seq and small RNA-seq on various tissues of the macaques and the marmosets, respectively, majority of which matched the human's tissues investigated by FANTOM5. Here we determine >100,000 promoters, >10,000 enhancers, >20,000 lncRNAs and >3,000 miRNAs of the macaques and the marmosets, respectively. The human promoters, enhancers, lncRNAs and miRNAs are conserved 41~90% in the macaque, 29~84% in the marmoset and 12~68% in the mouse, respectively. Our annotations on a genome of the non-human primates may deepen insights into evolution of the functional non-coding regions of mammalian genomes.

## **W751: Non-coding RNA**

### **Plant Telomerase RNA Genes: Conserved Motifs, Pol-III Transcription, and Scenario of Telomere Divergence**

**Petr Fajkus**<sup>1</sup>, Vratislav Peska<sup>1</sup>, Eva Sýkorová<sup>1</sup>, Michal Závodník<sup>2</sup> and Jiri Fajkus<sup>1,2</sup>, (1)Institute of Biophysics, CAS, v.v.i., Brno, Czech Republic, (2)Masaryk University, CEITEC and Faculty of Science, Brno, Czech Republic

Telomerase RNA (TR) provides a template for telomere synthesis by telomerase reverse transcriptase (TERT). The sequence of telomere repeat is dictated by a template region (usually 1.5 telomere repeat long) of TR subunit, and is thus complementary to telomere sequence.

With recent characterisation of telomerase RNA genes in plants we described putative TRs from 75 representatives across land plant phylogeny. Four of them were experimentally validated as genuine telomerase RNAs in *Arabidopsis thaliana* (Brassicales), *Nicotiana sylvestris* (Solanales), *Allium cepa* and *Scilla peruviana* (both Asparagales). Despite of huge sequence variability between so far described 75 putative plant TRs, these sequences showed shared architecture of conserved motifs and promoter structure, typical for RNA polymerase III class III transcripts like U6, U3, 7SL, MRP RNAs involving Upstream Sequence Element (USE) and TATA box. Considering of extensive richness of Plantae kingdom, recently described TR sequences in 75 representatives (3 from Gymnosperms, 72 from Angiosperms) is just drop in the ocean.

Here we challenge Telomerase RNA genes in evolutionary ancestral plant clades outside seed plants (Spermatophyta), and clades where we previously failed with TR identification using Blast - these include clades Zingiberales and core Caryophyllales, involving important model and crop species.

We present a fresh comprehensive comparison of new RNA subunits from seed plants, ferns and lycophytes. Based on the upgraded comprehensive set of RNA subunits from seed plants, ferns and lycophytes we re-examine conserved TR motifs, which can reflect some functional significance in TR biogenesis.

In contrast to animals or yeasts, plant genomes frequently contain more TR-like genes. We hypothesize, that these TR paralogs can serve as a source material for telomere sequence evolution. Indeed, new screen for plant telomerase RNA subunits revealed several TRs harbouring template regions, which are not complementary to typical plant telomere motif (thus unable to synthesize typical plant telomere DNA). Two options can be considered – either some TR candidates identified *in silico* are not genuine telomerase RNAs or we find other species with unusual telomeric sequences. Fortunately, here we experimentally support the second option. Thus, we demonstrate the applicability of TR prediction for identification of evolutionary changes in plant telomeres.

Funding: Czech Science Foundation [17-09644S], ERDF [project SYMBIT, reg. no. CZ.02.1.01/0.0/0.0/15 003/0000477]

## **W752: Non-coding RNA**

### **Origins of Long Intergenic Non-Coding RNAs in Angiosperms**

**Keith Adams** and Haomin Lyu, University of British Columbia, Vancouver, BC, Canada

TBA

## **W753: Non-Seed Plants**

### **Sex Determination in Volvocine Algae**

Sa Geng<sup>1</sup>, Takashi Hamaji<sup>2</sup>, Minglu Gao<sup>1</sup>, Patrick Ferris<sup>1</sup> and **James Umen**<sup>3</sup>, (1)Donald Danforth Plant Science Ctr, (2)Kyoto University, Japan, (3)Donald Danforth Plant Science Center, St. Louis, MO

Sexual reproduction is ubiquitous in green algae, but in most groups is cryptic or poorly understood. Volvocine green algae have well-studied facultative haplontic sexual cycles that are typical for chlorophytes, with indefinite mitotic reproduction in the haploid phase punctuated by infrequent mating to produce dormant diploid zygospores. Meiosis occurs during zygospore germination to produce new haploid vegetative progeny. In all heterothallic (dioicous) volvocine algae the conserved gene *MID* (minus dominance) governs sexual differentiation and is found exclusively in the *MT*- haplotype or male sex-determining region. In isogamous unicellular *Chlamydomonas reinhardtii* (*Chlamydomonas*) and oogamous multicellular *Volvox carteri* (*Volvox*) the expression of *MID* is necessary and sufficient to induce *minus* or male gametogenesis, respectively, while in its absence *plus* or female differentiation occurs as a default program. Two important but unanswered questions are how the default program of *plus* or female gametogenesis is specified, and how that program is modified by expression of *MID*. To address these questions, we developed a comparative transcriptomics approach to find conserved candidate transcription factors that are expressed specifically during *Chlamydomonas* or *Volvox* gametogenesis. One candidate, *VSR1* (volvocine sex regulator), was identified and characterized. Using CRISPR/Cas9 genome editing we generated a *Volvox* male *vsr1* predicted null mutant whose novel phenotype was blockage of spermatogenesis and the conversion of sperm cell precursors to vegetative stem cells termed gonidia. The *vsr1* mutant was rescued with an epitope-tagged transgene whose protein product was nuclear localized in sperm precursors and sperm cells. The rescued *vsr1* strain was crossed to a wild-type female, and *vsr1* female progeny were obtained. Analogous to male *vsr1* mutants, female *vsr1* mutants made sterile egg-like cells which could not be fertilized, but instead differentiated back into vegetative gonidia. In parallel, we characterized a mutation in the *Chlamydomonas VSR1* ortholog, *CrVSR1*. Like its *Volvox* mutant counterpart, a *Chlamydomonas vsr1* predicted null mutation was sterile in both mating types, but could be rescued by a *CrVSR1* transgene. Thus, *VSR1* is a key missing component of the sexual differentiation pathway in volvocine algae that is required for gametogenesis in both mating types or sexes. A simple model for mating type or sexual differentiation involves *VSR1* serving as an activator of *plus* or female gene expression, with interaction between *MID* and *VSR1* in *minus* or males resulting in repression of *plus* or female genes and activation of *minus* or male genes. We began testing this model and found association between corresponding *VSR1* and *MID* proteins of both species by co-immunoprecipitation (Co-IP). Moreover, the interaction between *VSR1* and *MID* is likely to be direct as it could be reconstructed using a yeast two-hybrid assay. Our data fill a major gap in understanding green algal sex determination and provide a paradigm for understanding the evolution of sexual differentiation in other green lineage systems through interactions between conserved transcription factors.

## **W754: Non-Seed Plants**

### **Remarkable Genome Stability in Autosomes Contrasts with Dynamic Sex Chromosomes in Mosses**

**Sarah Carey**, University of Florida, Gainesville, FL

Karyotypic variation has long been recognized both for its phylogenetic value and its contribution to reproductive isolation. The evolutionary forces driving most changes in genome architecture, however, remain largely unknown. Genomic conflict over meiotic transmission is one plausible hypothesis to explain rapid karyotypic turnover in sexually reproducing eukaryotes. If this is broadly true, sex chromosomes may be overrepresented in chromosomal translocations. Here we reconstructed the evolution of the bryophyte karyotype by comparing the genomes of the hermaphroditic species *Physcomitrella patens* and the dioecious *Ceratodon purpureus*, mosses which diverged over 200 million years ago. We used synteny-based analyses to identify collinear chromosomes within *C. purpureus* (i.e., resulting from duplications) and between *C. purpureus* and *P. patens*. We found the ancestral chromosomes are clearly recognizable in the modern genomes, based on their gene content (using 5 collinear genes), although rearrangements have eroded much of the longer-range synteny (>7 genes) within these chromosomes. In stark contrast, even short-range synteny (<3 genes) was absent among genes on the *C. purpureus* sex chromosomes. We used phylogenomic analyses of transcriptomic data from 40 additional moss species to identify the ancestral chromosomes that comprise the modern *C. purpureus* sex chromosomes. Of the 3 reconstructed translocation events since the divergence from the *P. patens* lineage, 2 involved the sex chromosome. Together these results highlight the remarkable stability of the moss genome, with the notable exception of the sex chromosomes, and point to a major role for genomic conflict in the evolution of the karyotype.

## **W755: Non-Seed Plants**

### **The Genome of the Desiccation Tolerant Moss *Syntrichia caninervis* and its Response to Abiotic Stress**

**Kirsten Fisher**<sup>1</sup>, Anderson T. Silva<sup>2</sup> and Melvin J. Oliver<sup>2</sup>, (1)California State University, Los Angeles, Los Angeles, CA, (2)Division of Plant Sciences and Interdisciplinary Plant Group, Columbia, MO

Using Dovetail's proprietary proximity ligation technologies, we sequenced the genome of a single *Syntrichia caninervis* female derived from the Mojave Desert of the United States. The final reference genome assembly is at chromosome scale with 13 assembled scaffolds (chromosomes) ranging from 14.9 Mb to 37.3 Mb in length. The genome annotation revealed 16,545 nuclear encoded protein-coding genes with an average coding-sequence length of 1,513 kb. Our genome annotation recovered 403 of the 430 highly conserved core proteins in the Viridiplantae. Among the 16,545 genes predicted in *S. caninervis*, a total of 15,793 (95.4%) demonstrated at least partial sequence similarities (BLASTP E-value < 1e-5) to *Physcomitrella patens* proteins and all 16,545 genes could be annotated with InterProScan domain/family information. The intra-genomic synteny detected between the genomes *S. caninervis* and *P. patens* identifies a total of 1003 syntenic blocks with 5,412 gene pairs. In addition, we detected in each *S. caninervis* gene a maximum of four syntenic genes in the *P. patens* genome. This might relate to the two successive whole genome duplication events (Pp-WGD2) reported for *P. patens* and suggests chromosomal or segmental losses in the evolutionary history of *S. caninervis*. We generated a complete transcriptome for *S. caninervis*, and identified 3,972 transcripts that were differentially accumulated in response to environmental stress: heat, cold, dehydration/desiccation, and rehydration. A dominant pattern (DP) analysis script, used to analyze the patterns of changes in transcript abundance for each stress, identified 11 DPs. These patterns in transcript abundance suggested the occurrence of processes related to the plasticity of the stress responses of the moss exposed to altered environments. Functional analysis of each DP gave insight into processes important in each stress response. For example, DP2 transcripts, a possible signature transcript set for dehydration tolerance, revealed a significant enrichment of transcripts involved in membrane function and stability, responses to desiccation, and cysteine biosynthesis. DP2 contained over 240 transcripts that are involved in the response to desiccation, and we are currently investigating the control process that leads to the change in transcript abundance: transcription or sequestration (demonstrated to be of major importance in the response of the moss to desiccation). Assessed transcriptomal changes during dehydration of *S. caninervis* will greatly facilitate map-based cloning of important genes for crop improvement, particularly those involved in abiotic stress tolerance.

## **W756: Non-Seed Plants**

## **The Sphagnum Project: Sphagnum as a Model Organism for Carbon Sequestration and Niche Ecosystem Evolution**

**Adam Healey**<sup>1</sup>, John Lovell<sup>2</sup>, Christopher B. Plott<sup>2</sup>, Jerry Jenkins<sup>2</sup>, Bryan Piatkowski<sup>3</sup>, Travis Lawrence<sup>4</sup>, Matthew G Johnson<sup>5</sup>, Jane Grimwood<sup>2</sup>, Kerrie W. Barry<sup>6</sup>, Anna Lipzen<sup>7</sup>, Shengqiang Shu<sup>7</sup>, Dave Weston<sup>8</sup>, Jeremy Schmutz<sup>2</sup> and Jonathan Shaw<sup>9</sup>, (1)HudsonAlpha Institute For Biotechnology, Huntsville, AL, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (3)Duke University, (4)Oakridge National Laboratory, (5)Department of Biological Sciences, Texas Tech University, Lubbock, TX, (6)DOE Joint Genome Institute, Berkeley, CA, (7)DOE Joint Genome Institute, Walnut Creek, CA, (8)Oak Ridge National Laboratory, Oak Ridge, TN, (9)Duke University, Durham, NC

Sphagnum (peat moss) is an excellent model organism for niche ecosystem evolution and global carbon cycling. Sphagnum engineers its microenvironment by lowering soil pH and reducing carbon and nitrogen availability to generate acidic and anoxic conditions, allowing it to thrive while suppressing growth in other plants and microbes. In peatlands, Sphagnum often grows in sympatry, with species occupying space at varying heights relative to the water-table, growing in low valleys (hollows) or high mounds (hummocks). For a better understanding of the genetic traits that allow Sphagnum to engineer its ecosystem, we present the genomic resources of the Sphagnum Project, whose aim is to investigate the genetics that underlie ecologically relevant traits and extend those associations across the phylogeny of Sphagnum. For this purpose, the Sphagnum project has generated two high-quality reference genomes for *S. fallax* (395 Mb; Scaffold N50: 21 Mb) and *S. magellanicum* (439 Mb; Scaffold N50 17 Mb) representing the hollow and hummock ecosystem niches, a fully sequenced half-sibling pedigree of 184 offspring and a dense 5,395 cM genetic map in 19 linkage groups for marker trait association, and high-coverage Illumina re-sequencing of 35 individuals representing 15 taxa within the 5 major subgenera of Sphagnum. The combination of these resources position Sphagnum as an excellent model system for functional, ecological and evolutionary genomics.

## **W757: Non-Seed Plants**

### **The First High-Quality Reference Genome for Hornworts (*Anthoceros*)**

**Peter Szovenyi**, Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland, Fay-Wei Li, Cornell, BTI, New York, NY, Tomoaki Nishiyama, Kanazawa University, Kanazawa, Japan, Keiko Sakakibara, Rikkyo University, Japan, Juan-Carlos Villarreal, Laval University, Quebec City, Quebec, Canada, Eftychios Frangedakis, University of Cambridge, United Kingdom, Stefan A. Rensing, University of Marburg, Marburg, Germany and Waller Manuel, University of Zurich, Switzerland

The monophyletic group of hornworts is believed to represent one of the three lineages of bryophytes exhibiting many unique features within land plants, such as symbiosis with cyanobacteria, the presence of pyrenoids and a basal sporophytic meristem. Until now, gaining new knowledge on the unique biology of hornworts and its overall significance for land plant evolution was primarily hindered by the lack of genomic resources for a hornwort model species. Here we provide a high-quality genome draft of the model hornwort, *Anthoceros agrestis*, and some of its relatives. With the aid of Chicago and Hi-C libraries we assembled the *A. agrestis* genome into 5-6 chromosomes spanning a total length of ca. 120 Mb. The *A. agrestis* genome drastically differ from the published bryophyte genomes (moss and liverwort) and exhibit a number of unique features. The *A. agrestis* genome is small and strongly streamlined. In particular, we show that it has few recent paralogs, exhibits no signs of whole-genome duplication and has a moderate repeat content (ca. 30%). Despite of its small size, the genome is densely packed with genes (28 000 genes) of which 30% seems to be unique to hornworts showing no homology to genes of land plant and green algal genomes. We further show that the genome contains over 3000 PPR genes (pentatricopeptide repeat) known to be involved in organellar RNA editing that is pervasive in chloroplast-derived transcripts of *A. agrestis*. Finally, we present a preliminary gene expression atlas which shed light on the regulation of morphological and developmental traits that are either shared with other embryophytes or unique to hornworts. Overall, our analysis suggests that the hornwort genome is radically different from that of other land plants. Its unique and shared features

within land plants provides better understanding on the evolutionary genomic events accompanied the origin and evolution of land plants.

### **W758: Non-Seed Plants**

#### **Evolutionary Analyses of Machine-Learned Gene Regulatory Networks in Plants**

**Andrea Braeutigam**, Bielefeld University, Bielefeld, Germany

Transcriptional regulatory networks are at the core of plant responses to developmental, biotic, and abiotic cues. Machine learning algorithms offer a new avenue to predict the gene regulatory network (GRN) of a species based on large scale RNA-seq data collections, however, predictions suffer from large numbers of false positives. We developed methods to extract biological data from GRN prediction even in the presence of these positives. Phylogenetic reconstructions of transcription factor relationships from eight phototrophic species enabled the analyses of transcription factor functions in their phylogenetic context. Analyses of target genes for orthologous transcription factors revealed extensive conservation. In most plant species, a genome history with multiple polyploidizations results in complex patterns. The potential of gene regulatory networks of the liverwort *Marchantia polymorpha* with its frozen genome is discussed.

### **W759: Oats, Wild and Cultivated**

#### **Oat Genomes: Their Sequences, Diversity, and Domestication**

**Tim Langdon**, IBERS, Aberystwyth University, Aberystwyth, United Kingdom

### **W760: Oats, Wild and Cultivated**

#### **PanOat: Prospects of Pan-Genomics in Oat**

**Martin Mascher**, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Oat (*Avena sativa*) is an important cereal crop that is grown in temperate regions throughout the world. The large size and polyploid nature of its genome have long posed obstacles to assembling an oat reference genome sequence. Progress in genome sequencing and assembly methodology has now enabled the construction of chromosome-scale reference sequences for multiple hexaploid oat genotypes. Guided by the experiences in the wheat and barley pan-genome projects, I will outline a strategy for PanOat, a pan-genome project for oat. PanOat will select representative genotypes for major germplasm groups based on genotyping-by-sequencing data of a large diversity panel. Chromosome-scale sequence assembly will be obtained for these representatives using the TRITEX pipeline. Comparative analyses of 10+ oat genomes together with diploid and tetraploid ancestors will inform about the origin and extent of structural variants and their possible association with agronomic traits. PanOat will contribute towards translational goals such as the design of genotyping platforms and efficient haplotype imputation.

### **W761: Oats, Wild and Cultivated**

#### **Toward the World's Most Interesting Pan Genome**

**Eric N. Jellen**, Brigham Young University, Provo, UT

Domesticated hexaploid oat (*Avena sativa* and *A. byzantina*,  $2n=6x=42$ , AACDD) is the most nutritious of the cereal grains. The hexaploid oats evolved over approximately 7My through multiple rounds of polyploidization in the western and central Mediterranean basin. While Mediterranean-type environments are notorious for their wide variation for annual precipitation and, to a lesser extent, temperature, the West-Central Mediterranean home of *Avena* additionally experienced repeated cycles of desertification through glacial-interglacial cycles. One of these desertification cycles resulted in extreme sea-level decline in the Mediterranean Sea, creating lakes and basins experiencing extreme temperatures at

altitudes hundreds of meters below sea level, and was abruptly reversed through a tectonically mediated event, the Zanclean Flood, approximately 5Mya. Additionally, the centers of oat origin and domestication have been highly disturbed by concentrated human activity, leading to rapid dispersion of weedy and extreme habitat fragmentation of non-invasive forms, for the past 10Ky. During this latter Neolithic Period, hexaploid oat was probably domesticated at least twice in the ancient Near East and eventually overspread the globe in its domesticated and free-living (*A. fatua* and *A. sterilis*) forms. Recent whole-genome sequencing of  $A_sA_s$  and  $C_pC_p$  *Avena* diploids revealed that these sub genomes are highly divergent at the level of chromosome structure. Intergenomic CC-DD chromosome rearrangements are a well-known feature of the hexaploid oat genome, with several large translocations characterizing different gene pools. We know of only one large AA-CC translocation (involving chromosomes 17A and 7C) and whole-genome sequencing will be necessary to identify rearrangements between the highly similar AA and DD sub genomes. We expect that oat disease breeding using wild diploid and tetraploid species has introduced exotic chromosome segments that have served to further differentiate oat breeding pools - for example, for rust resistance from *A. magna* and *A. longiglumis*-derived 'Amagalon' in the USA and Canada. Consequently, "oat" might actually be a series of genetically distinct genepool-races or even subspecies which, when intercrossed, produce hybrids whose offspring display elevated levels of sterility.

## **W762: Oats, Wild and Cultivated**

### **Prospects for Improved Genomic Selection through Pan Genomics**

**Wubishet Bekele**<sup>1</sup>, Asuka Itaya<sup>1</sup>, Jennifer Mitchell Fetch<sup>2</sup>, Weikai Yan<sup>1</sup> and Nicholas A. Tinker<sup>1</sup>, (1)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)AAFC BRDC, Brandon, MB, Canada

This talk presents the development and application of genotyping-by-sequencing marker systems for genomic selection in the Ottawa oat breeding program and encouraging forward validation results. We will also share the strategy we have initiated to implement practical haplotype graph (PHG) in our genomics-assisted breeding scheme. The critical elements of our approach include whole-genome sequencing of founder lines, denovo assembly of important oat cultivars and pangenome analysis.

## **W763: Oats, Wild and Cultivated**

### **Towards an Understanding of Beta-Glucan Regulation in Oat**

**Annis Fatmawati**, Department of Plant Science, McGill University, Saint-Anne-de-Bellevue, QC, Canada, Thomas Donoso, Department of Plant Science, McGill University, Canada, Mohannad Mahmoud, Department of Plant Science, McGill University, Ste Anne de bellevue, QC, Canada, Rajvinder Kaur, McGill University, Ste Anne de bellevue, QC, Canada, Nicholas A. Tinker, Agriculture and Agri-Food Canada, Ottawa, ON, Canada and Jaswinder Singh, Department of Plant Science, McGill University, Ste Anne de Bellevue, QC, Canada

Oat is a major cereal crop that is grown worldwide for human food and animal feed. Its use in human diets is growing in popularity in part due to its ability to reduce serum cholesterol and glucose level. This has been attributed to its high  $\beta$ -glucan content.  $\beta$ -glucan is a major non-starch carbohydrate component, consisting of double  $\beta$ -1,3 and  $\beta$ -1,4 linkages. Information about  $\beta$ -glucan synthesis in oat is limited in part due to a lack of genomic resources. We are introducing the maize *Ac* and *Ds* transposable elements into the oat genome with the goal to create an experimental transposon-mediated functional genomic resource to identify genes encoding important traits such as  $\beta$ -glucan content. Recently, a Thaumatin Like Protein, *TLP8* has been identified in barley that interacts with  $\beta$ -glucan to regulate its content in the grain. Higher transcript abundance of *TLP8* in barley grains reflect lower amounts of  $\beta$ -glucan, and vice-versa. We hypothesize that the downregulation of *TLP8* could increase  $\beta$ -glucan content in oat. The *TLP8* homolog in oat was retrieved and an RNAi constructs created. Genetic transformation was then conducted via the bombardment gun method. Transformants were generated and selected using a phosphinothricin N-acetyltransferase (*PAT*) marker gene, yielding a 5-13% transformation efficiency. Histochemical assays confirmed the expression *PAT*, and transgenic plants

were resistant to herbicide LIBERTY (0.2%). Currently, we are characterizing transgenic lines at the molecular and biochemical levels in order to explain the association between TLP8 and  $\beta$ -glucan in oat.

#### **W764: Oats, Wild and Cultivated**

##### **Harvesting Oat CORE QTLs: Milling Quality, Seedling Growth, and Phenotypic Stability**

**Kathy L. Esvelt Klos**, USDA-ARS, Aberdeen, ID, Aaron D. Beattie, University of Saskatchewan, Saskatoon, SK, Canada and Catherine Howarth, IBERS, Aberystwyth University, Aberystwyth, United Kingdom

The Collaborative Oat Research Enterprise (CORE) panel of elite lines remains a useful resource for investigating the genetic architecture of complex oat traits and for identifying quantitative trait loci. In 2010, 2011, and 2017 lines were evaluated at multiple locations for the milling quality-related mature seed characteristics summarized by measures of test weight, thousand kernel weight, groat content, the percent of groats broken during de-hulling, and percent of plump and thin kernels. In 2018, many of these lines were evaluated at the National Plant Phenomics Centre, Aberystwyth, UK for seedling growth under soil moisture contents of 25%, 35%, 75%, and 105%. Together, these sets of phenotypic data allow us to examine the genetic architecture of quantitative traits from both ends of the oat plant's life, and to identify QTL influencing trait variation *per se* as well as QTL influencing trait stability across environments.

#### **W765: Organellar Genetics**

##### **Talen-Based Mitochondrial Genome Editing in Plants**

**Shin-ichi Arimura**, The University of Tokyo, Tokyo, Japan, Tomohiko Kazama, Kyushu University, Japan and Nobuya Koizuka, Tamagawa University, Japan

A lack of methods for transforming plant mitochondrial genomes has hampered our understanding of mitochondrial genes. Some of these genes are involved in the agronomically important trait of cytoplasmic male sterility (CMS). Here we attempted to edit CMS-related genes (*orf79* and *orf125*) in their mitochondrial genomes of male-sterile lines, BTA (Boro II type) of rice, *Oryza sativa*, and SW18 (kosena, a kind of Ogura-type) of rapeseed, *Brassica napus*, by using transcription activator-like effector nucleases (TALENs) with mitochondrial localization signals (mitoTALENs). Targeted disruption of these genes led to deletions ranging from ca 100 bp to 5 kb and restoration of fertility. The sequences adjacent to the deletions did not reconnect each other but connected to distant loci by illegitimate homologous recombination. The configurations of the mitochondrial genomes after treatment of mitoTALENs changed some or a lot but without lacking other (essential) genes. mitoTALENs appear to be an effective tool for analyzing and modifying plant mitochondrial genomes.

#### **W766: Organellar Genetics**

##### **Nucleus-to-Plastid Phytochrome Signaling in Controlling Plastid Transcription**

**Meng Chen**, University of California Riverside, Riverside, CA

Phytochromes are red and far-red photoreceptors that regulate every facet of plant development growth. When seedlings emerge from the soil and encounter light for the first time, phytochromes trigger a developmental transition from a dark-grown program called skotomorphogenesis to a light-dependent program called photomorphogenesis. The photomorphogenetic program enables the biogenesis of photosynthetically-active chloroplasts and thus transitions seedlings into a photoautotrophic lifestyle. Chloroplast biogenesis requires the activation of photosynthesis-associated genes encoded by both the nuclear and plastidial genomes. It is well understood that light triggers the translocation of phytochromes from the cytoplasm to the nucleus to activate photosynthesis-associated nuclear-encoded genes, but how phytochromes – which do not localize in plastids – control the expression of photosynthesis-associated plastid-encoded genes (*PhAPGs*) remains elusive. The plastidial genome is transcribed by two types of RNA polymerase: a phage-type nuclear-encoded RNA polymerase that transcribes

housekeeping genes and a bacterial-type plastid-encoded RNA polymerase (PEP) that transcribes *PhAPGs*. Our genetic studies on phytochrome signaling have serendipitously revealed that phytochrome signaling and the PEP are connected by a dual-targeted nuclear/plastidial protein named HEMERA (HMR). While nuclear HMR is a transcriptional activator required for phytochrome signaling, plastidial HMR is a PEP-associated protein essential for *PhAPG* expression. In my talk, I will discuss our latest work on the mechanistic link between phytochrome signaling and the regulation of plastidial transcription.

### **W767: Organellar Genetics**

#### **Structure, Biogenesis and Engineering of the Eukaryotic CO<sub>2</sub>-Concentrating Organelle, the Pyrenoid**

**Martin Jonikas**, Princeton University, Princeton, NJ

Approximately one-third of global carbon-fixation occurs in an overlooked algal organelle called the pyrenoid. The pyrenoid contains the CO<sub>2</sub>-fixing enzyme Rubisco, and enhances carbon-fixation by supplying Rubisco with a high concentration of CO<sub>2</sub>. The molecular structure and biogenesis of this ecologically fundamental organelle have remained enigmatic. By using high-throughput localization of proteins and identification of protein-protein interactions in the model alga *Chlamydomonas reinhardtii*, we increased the number of known pyrenoid components from 6 to over 80, and discovered the existence of three new protein layers in the pyrenoid: a plate-like layer, a mesh layer, and a punctate layer. We discovered that an abundant pyrenoid protein, Essential Pyrenoid Component 1 (EPYC1), works as a molecular glue that binds Rubisco holoenzymes together to form the matrix at the core of the pyrenoid. We then found that a simple mechanism involving a Rubisco-binding motif explains both targeting of proteins to the pyrenoid and the overall architecture of the three pyrenoid sub-compartments. Finally, contrary to longstanding belief that the pyrenoid matrix is a solid structure, we discovered that the matrix behaves as a liquid droplet, which mixes internally, divides by fission, and dissolves and condenses during the cell cycle. Our data provide insights into pyrenoid protein composition, structural organization and biogenesis. Working with our collaborators in the Combining Algal and Plant Photosynthesis project, we aim to transfer algal pyrenoid components into higher plants to enhance carbon fixation and yields in crops.

### **W768: Organellar Genetics**

#### **Engineering a Cyanobacterial Carbon-Concentrating Mechanism into Plants**

**Douglas Orr**<sup>1</sup>, Dawn Worrall<sup>1</sup>, Myat T Lin<sup>2</sup>, Vishal Chaudhari<sup>2</sup>, Elizabete Carmo-Silva<sup>1</sup>, Maureen R. Hanson<sup>2</sup> and Martin A.J. Parry<sup>1</sup>, (1)Lancaster University, Lancaster, United Kingdom, (2)Cornell University, Ithaca, NY

Using a CO<sub>2</sub> concentrating mechanism (CCM), cyanobacteria encapsulate a Rubisco with poor specificity but a relatively fast catalytic rate within a carboxysome micro-compartment. Coupled with the active transport of bicarbonate into the cell, and localization of carbonic anhydrase within the carboxysome shell with Rubisco, this creates a localised high CO<sub>2</sub> environment around Rubisco and overcomes many of its limitations, thereby improving photosynthetic efficiency. We are using a synthetic biology approach to engineer a  $\beta$ -cyanobacterial CCM into land plants. Our recent data shows that Rubisco from the  $\beta$ -cyanobacteria *Synechococcus elongatus* (Se) forms aggregated Rubisco complexes with the carboxysome linker protein CcmM35 within tobacco chloroplasts. This led us to a number of novel and interesting findings with transplastomic plants that were able to form a hybrid Rubisco enzyme utilizing tobacco small subunits and cyanobacterial large subunits, allowing carbon fixation and slow autotrophic growth in high CO<sub>2</sub>.

### **W769: Organellar Genetics**

#### **Carbonic Anhydrase Mutants for Engineering of Carbon-Concentrating Mechanisms into Chloroplasts**

Kevin M. Hines<sup>1</sup>, Kristen Edgeworth<sup>2</sup>, Thomas G Owens<sup>1</sup> and **Maureen R. Hanson**<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)Kenyon College, Gambier, OH

The enzyme carbonic anhydrase (CA) catalyzes the interconversion of bicarbonate (HCO<sub>3</sub><sup>-</sup>) with carbon dioxide (CO<sub>2</sub>) and water. In order for a carbon-concentrating mechanism (CCM) to operate within a chloroplast, CA must be confined to the microcompartment or other structure that encapsulates Rubisco, in order to prevent conversion of bicarbonate to CO<sub>2</sub> before it reaches the compartment. However, C3 plants such as tobacco contain two nuclear-encoded CAs in their chloroplast stroma, βCA1 and βCA5. In order to remove them from the stroma, we produced CRISPR/Cas9 mutants affecting both genes.

While the single knockout lines Δβca1 and Δβca5 had no striking phenotypic differences compared to WT plants, Δβca1ca5 leaves developed abnormally and exhibited large necrotic lesions, even when supplied with sucrose. Leaf development of Δβca1ca5 plants normalized at the high CO<sub>2</sub> concentration of 9000ppm. High CO<sub>2</sub>-grown Δβca1ca5 mutants had no measurable defect in photosystem II efficiency when measured at ambient CO<sub>2</sub>. Nevertheless, emerging Δβca1ca5 leaves show an upsurge in chloroplast reactive oxygen species (ROS). The chloroplasts within the leaves of the double mutant exhibit a higher pH than wild-type, according to assays with a pH-sensitive GFP. Δβca1ca5 seedling germination and development is negatively affected when seedling development occurs at ambient CO<sub>2</sub>.

A series of complementation experiments using altered forms of βCA1 were carried out in Cas9 -lacking Δβca1ca5 plants. Constructs expressing full-length βCA1 and βCA5 proteins complemented the Δβca1ca5 mutation, but inactivated (ΔZn-βCA1) and cytoplasm-localized (Δ62-βCA1) forms of βCA1 failed to reverse the mutant phenotype. When infected with tobacco mosaic virus (TMV) Δβca1 and Δβca1ca5 tobacco failed to show the hypersensitive response (HR), while expression of ΔZn-βCA1 restored the response.

Taken together, our data suggests that a deficiency of bicarbonate in the chloroplast stroma inhibits important biosynthetic reactions, leading to impaired plant development. If bicarbonate transporters are engineered onto the chloroplast envelope, adequate bicarbonate should be supplied for biosynthesis, likely normalizing the phenotype, which would make installation of a CCM feasible.

## **W770: Organellar Genetics**

### **Target Excision By Direct Export of a Site-Specific Recombinase from Agrobacterium to Tobacco Chloroplasts**

**Pal Maliga**, Rutgers University, Piscataway, NJ

T-DNA export from *Agrobacterium* to plant cells occurs by the type 4 protein secretion machinery. The T-complex is naturally targeted to the nucleus. We plan to retarget the T-complex to chloroplasts. As a proof of concept for retargeting we have shown target excision in the chloroplast genome by the Int recombinase exported from *Agrobacterium* to chloroplasts. The test system involves first introducing a silent spectinomycin-resistance (*aadA*) gene in the plastid genome. This gene is then activated by the excision of a blocking sequence by Int exported from *Agrobacterium*. The engineered chloroplast genomes were identified by spectinomycin resistance, conferred by the expression of the *aadA* marker gene. Because T-DNA is transferred from *Agrobacterium* to the plant cells by the Type 4 secretion machinery, successful re-targeting of proteins from *Agrobacterium* to plastids suggests that plastid transformation may be feasible by retargeting T-DNA delivery to plastids.

## **W771: Ornamentals**

### **Genetic Structure in *Hydrangea quercifolia* Bart. throughout its Natural Range**

**Andrew P. Sherwood**<sup>1</sup>, Stan C Hokanson<sup>1</sup>, Lisa Alexander<sup>2</sup>, Matthew D. Clark<sup>1</sup>, Yaniv Brandvain<sup>3</sup> and Steve McNamara<sup>1</sup>, (1)University of Minnesota, St. Paul, MN, (2)USDA-ARS U.S. National Arboretum, (3)University of Minnesota, Saint Paul, MN

*Hydrangea quercifolia* is a shrub native to the southeastern United States, which is also cultivated for its ornamental appeal. Little is known about the genetic diversity of *H. quercifolia*, or about any *Hydrangea* species in the wild. The range of *H. quercifolia* extends into six states and is therefore small enough to be amenable to comprehensive sampling. Tissue samples were collected from 74 geographically distinct populations from the extent of the native range. A subset of 188 samples were chosen to represent each population for SNP discovery using GBS. SNP loci were identified using Stacks de novo and 6,052 passed all filters. Loci contained on average 1.9 SNPs and were therefore considered to be microhaplotypes for subsequent analyses. Genetic diversity was analyzed using 1,701 of the SNP loci using Structure and Structure Harvester. The optimal K was determined to be either 3 or 5 clusters. The largest cluster, which was identified with both K=3 and K=5, includes the populations in Louisiana, Mississippi and the western half of Tennessee. The populations in Alabama, Georgia, Florida and the eastern half of Tennessee belong to the remaining 2 or 4 clusters. This follows the Tombigbee River discontinuity pattern that is found in several species in the southeastern US. No evidence of isolation by distance was found at the rangewide scale using a Mantel test. Genetic structure was also analyzed using k-means clustering, PCA of allele frequencies and a neighbor joining tree constructed using Nei's genetic distance. All these methods produced congruent results and provided support for either 3 or 5 hierarchical genetic clusters.

### **W772: Ornamentals**

#### **DArTseq LD Differentiates Species and Populations of Historic and Extant Wild *Chrysanthemum arcticum*, *C.a.* subsp. *arcticum*, and *C.a.* subsp. *polaré***

Yunjia Liu, University of Minnesota, Saint Paul, MN, Neil O. Anderson, University of Minnesota, Department of Horticultural Science, Saint Paul, MN and Andrzej Noyszewski, University of Minnesota, Department of Horticultural Science, St. Paul, MN

*Chrysanthemum arcticum* L. (= *Leucanthemum arcticum* L.) and its two subspecies, *C.a.* subsp. *arcticum* L. and *C.a.* subsp. *polaré* Hultén, are native to the New World, with the center of origin and diversity in the State of Alaska. The species is of interest since it is salt tolerant, growing only in maritime plant communities either on the beachfront or oceanic cliffs/cliff face edges. The purpose of this research was to examine the genetic structure of the *C. arcticum* species, subspecies, and populations. DArTseqLD is a proprietary genotyping method, based on genome complexity reduction; single nucleotide polymorphisms (SNPs) were generated from a pilot study of diverse genotypes. Historic (herbaria) and extant (living) specimens were collected from eleven N. America herbaria (using destructive sampling) and in forty-three wild populations throughout Alaska and Canada, respectively. Herbaria specimens reflect a wider range of species distribution and genetic variation compared with the extant specimens. Extant populations of *C. arcticum* (n=9 populations) and *C.a.* subsp. *arcticum* (n = 21 populations) separated, based on taxonomic classification; *C. arcticum* also had significant genetic differences among populations on mainland Alaska. Sympatric *C.a.* subsp. *arcticum* on Attu Island, Alaska, constituted two distinct genetic groups among the 21 collected populations. Herbaria specimens were predominantly similar to the corollary extant species' genetic structure, separating the taxonomic species delineations. *Chrysanthemum a.* subsp. *polaré* (n = 13 populations) was the most distant to both *C. arcticum* and *C.a.* subsp. *arcticum*, since it is geographically isolated. DArTseqLD also clearly distinguished *C. arcticum* and subspecies from other related Asiatic *Chrysanthemum* species. Provided high quality/quantity of DNA could be extracted, DArTseqLD is a powerful method of differentiating among and within species, regardless of whether the DNA was derived from herbaria or fresh samples.

### **W773: Ornamentals**

#### **Dissecting Genetic Diversity and Genomic Background of Petunia Cultivars with Contrasting Growth Habits using Transcriptome-Derived SNPs**

Yufang Guo and Ryan M. Warner, Michigan State University, East Lansing, MI

The modern cultivated garden petunia (*Petunia ×hybrida*) is one of the most economically important ornamental plant species. *Petunia axillaris* and *P. integrifolia* are the progenitor species of *P. ×hybrida*,

which dates back only to the 1830s. To investigate the genetic diversity and genomic background of current petunia cultivars with contrasting growth habits (upright vs. trailing), a panel of 13 petunia cultivars representing several breeding programs were genotyped with transcriptome-derived SNPs, and compared to the two progenitor species. A broad range of numbers of polymorphic SNPs were detected across the petunia cultivars and species. The relative genomic contributions of *P. axillaris* and *P. integrifolia* varied by cultivar. Cultivars with an upright growth habit had a higher proportion of homozygous alleles as well as a larger proportion of *P. axillaris*-derived alleles, compared to trailing cultivars. Additionally, trailing cultivars were enriched for “out group” SNPs (i.e. loci that were monomorphic between *P. axillaris* and *P. integrifolia*, but polymorphic between the cultivars and the progenitor species) that were clustered in regions across the seven chromosomes, suggesting these regions may have been introgressed from a third progenitor species and may be important regions for identifying genes conferring the trailing habit.

## **W774: Ornamentals**

### **Sterile Eucalyptus for Ornamental and Forestry Uses through CRISPR Disruption of *Leafy* Function**

Estefania Elorriaga<sup>1</sup>, Amy Leigh Klocko<sup>2</sup>, Cathleen Ma<sup>1</sup>, Alexander A. Myburg<sup>3</sup>, Xinmin An<sup>4</sup>, Marc du Plessis<sup>5</sup> and **Steven H. Strauss<sup>1</sup>**, (1)Oregon State University, Corvallis, OR, (2)University of Colorado Colorado Springs, Colorado Spring, CO, (3)Department of Genetics, University of Pretoria, Pretoria, South Africa, (4)Beijing Forestry University, Beijing, China, (5)University of Pretoria, Pretoria, South Africa

*Eucalyptus* is among the most widely planted and economically important taxa of ornamental and forest trees in the world. However, its spread as an exotic or genetically engineered form can create ecological and social problems. We sought to eliminate pollen and seed dispersal by mutation of the *Eucalyptus* ortholog of *LEAFY* (*LFY*). We transformed a wild type (WT) *Eucalyptus urophylla* x *grandis* hybrid and two *Flowering Locus T* (*FT*) overexpressing (i.e., early flowering) lines with CRISPR Cas9 constructs targeting *LFY*. We found highly efficient rates of *lfy* biallelic knock-outs using CRISPR Cas9 vectors, often approaching 100% of transgenic insertion events. Frameshift mutations in early-flowering, *AtFT*-overexpression backgrounds failed to produce any viable male or female gametes and did not differ statistically in growth rate from transgenic or wild type trees in a greenhouse trial. Expression of genes upstream or near to *LFY* in the floral development pathway were hyperexpressed, whereas floral organ identity genes downstream of *LFY* were severely depressed, showing an inability to progress towards floral organ differentiation. We conclude that disruption of *LFY* function induces indeterminacy in inflorescence development, and appears to be capable of efficient genetic containment and an absence of detectable effects on vegetative growth rate. The main constraints to the wide use of CRISPR and/or genetic containment methods in eucalypts are widespread market restrictions to recombinant DNA modified varieties, stringent and often costly regulatory impediments to field research and commercial authorizations, the recalcitrance of many eucalypt species and genotypes to transformation and regeneration, and the lack of an efficient somatic system for removal of CRISPR machinery when required for market/regulatory acceptance.

For financial support, we thank the industrial members of the Tree Biosafety and Genomics Research Cooperative (TBGRC) at OSU, the United States Department of Agriculture (USDA award 2011-68005-30407, the USDA System For Advanced Biofuels Production From Woody Biomass In The Pacific Northwest), the USDA Biotechnology Risk Assessment (grants 2011-68005-30407 and 2010-33522-21736), and the NSF I/UCRC Center for Advanced Forestry (# 0736283).

## **W775: Palm Genetics and Genomics**

### **Coconut Genetics and Genomics: Updates and Opportunities Towards a Vibrant Coconut Industry**

**Hayde C. Flandez-Galvez<sup>1</sup>**, Darlon V. Lantican<sup>1</sup>, Anand Noel C. Manohar<sup>1</sup>, Consorcia E. Reano<sup>1</sup>, Maria Luz J. Sison<sup>1</sup>, Susan M. Rivera<sup>2</sup>, Ramon L. Rivera<sup>2</sup>, Roanne R. Gardoce<sup>1</sup>, Ernesto E. Emmanuel<sup>2</sup>, Ma. Genaleen Q. Diaz<sup>1</sup>, Cristeta A. Cueto<sup>2</sup>, Barbara L. Caoili<sup>1</sup>, Melvin P. Dancel<sup>1</sup>, Don Emanuel M. Cardona<sup>1</sup>, Romnick A. Latina<sup>1</sup>, Edward Cedrick J. Fernandez<sup>1</sup>, Cris Q. Cortaga<sup>1</sup>, Jomari C. Domingo<sup>1</sup>, Frenzee Kroeizha L. Pammit<sup>1</sup>, Cecille Anne L. Osio<sup>1</sup>, Michelle S. Guerrero<sup>1</sup>, Jerice P. Monge<sup>1</sup>, Cris Ann M. Lim<sup>1</sup>, Ana Clarissa M. Ambagan<sup>1</sup>, Rei Jedial O. Guevarra<sup>1</sup>, Angelica Kate G. Gumpal<sup>1</sup>, Erica Joy Amarilla<sup>1</sup>, Dominic P. Agudera<sup>1</sup>, John Carlo O. Campomanes<sup>1</sup>, Maria Salve C. Vasquez<sup>1</sup>, Hidelisa P. Hernandez<sup>1</sup>, Susan R. Strickler<sup>3</sup> and Lukas A. Mueller<sup>3</sup>, (1)University of the Philippines Los Baños, Philippines, (2)Philippine Coconut Authority (PCA), Philippines, (3)Boyce Thompson Institute, Ithaca, NY

Philippines is the second world supplier of coconut by-products. The region has been threatened with devastating production constraints ranging from agro-climatic and weather calamities to widespread prevalence of disease/insect pests outbreaks, and increasing existence of non-bearing and senile palms in coconut plantations. To facilitate the development of resilient and outstanding varieties especially for added high-value traits, advancements in genomics and related technologies are harnessed towards their effective integration in a coconut breeding program. Coconut whole genome sequence reads were generated using 'Catigan Green Dwarf' (CATD) as the reference variety and combinations of advanced next generation sequencing (NGS) platforms. High quality genome assembly was generated and used to characterize adaptation and economically important genes i.e. candidate resistance genes, drought tolerance, productivity, and coconut oil related genes. Genome-wide and gene specific DNA markers are generated. A user-friendly database is being developed to house the coconut genome sequence data, gene/trait models and associated DNA markers.

Updates from the Philippines coconut genomics project will be presented. These include gene mining for host resistance against coconut scale insect (CSI) and screening for CSI least damaged coconut varieties, as well as characterization of coconut genes related to fruit flesh/endosperm mutations and coconut oil qualitative/quantitative traits. Significant result from initial molecular and biochemical studies that support nutritional and medicinal claims will also be presented. The unprecedented opportunities beyond basic science from these major S&T achievements in coconut and in integration with applicable new breeding technologies will be discussed.

## **W776: Palm Genetics and Genomics**

### **Inbreeding Management and Optimization of Genetic Gain with Phenotypic and Genomic Selection in Oil Palm (*Elaeis guineensis*)**

**David Cros**, CIRAD, UMR AGAP, Montpellier, France

Oil palm breeding relies on reciprocal recurrent selection between two heterotic groups complementary for bunch number and average bunch weight. Given the long generation interval and the limited selection intensity imposed by the progeny tests currently used in the program, genomic selection (GS) is a very promising solution for this species. However, GS also accelerates the annual increase in inbreeding in oil palm parental populations. This can generate inbreeding depression, which can be detrimental for seed production, and cause the loss of favourable alleles, which can reduce the long-term genetic progress. Here, we investigated the effect of three approaches of inbreeding management on parental inbreeding and genetic progress in hybrids. We simulated two widely used parental populations, La Mé and Deli, and four generations of selection. Inbreeding was measured in La Mé and genetic progress on hybrids bunch production. Inbreeding management in La Mé was made by: (i) mate selection, which uses the simulated annealing optimization algorithm, (ii) limiting deterministically the number of full-sibs selected and (iii) prohibiting selfings. The results showed that all methods slowed down the increase in parental inbreeding. Mate selection was also able to simultaneously increase the genetic progress. Stronger slowing-down in inbreeding were achieved with deterministic methods, in particular by selecting at best one individual per full-sib family and prohibiting selfings. However, this was associated with a decreased genetic progress. Finally, mate selection will allow oil palm breeders to control the rate of increase in inbreeding in the parental populations while maximizing the genetic gain.

### **W777: Palm Genetics and Genomics**

#### **Defense Strategies of Oil Palm Against *Phytophthora palmivora* Causal Agent of Bud Rot**

**Hernán Mauricio Romero**<sup>1</sup>, Kelly Avila<sup>2</sup>, Juanita Gil<sup>3</sup>, Leonardo Araque<sup>2</sup>, David Botero<sup>2</sup>, Greicy Sarria<sup>2</sup>, Jorge Duitama<sup>3</sup> and Silvia Restrepo<sup>3</sup>, (1)Universidad Nacional de Colombia / Cenipalma, Bogota, Colombia, (2)Colombian Oil Palm Research Center - Cenipalma, Colombia, (3)Universidad de los Andes, Bogota, Colombia

### **W778: Palm Genetics and Genomics**

#### **Coconut Genetics and Genomics: Updates and Opportunities Towards a Vibrant Coconut Industry**

**Hayde Flandez-Galvez**, University of the Philippines Los Banos, Los Banos, Laguna, Philippines

### **W779: Palm Genetics and Genomics**

#### **A Genetic Map Linked, Long-Read Reference Genome for Date Palm and Its Application in Genotype:Phenotype Mapping**

**Joel Malek**, Weill Cornell Medicine-Qatar, Doha, Qatar

The date palm tree produces fruit of high importance to the regions of north Africa, the middle east and parts of south asia. Until recently only a few genetic resources were available for date palm research. However, over the past 10 years significant genetic resources have been produced for the date palm. We produced the first draft genome sequence assembly and genetic map of the date palm genome Khalas cultivar using short read technology and others have improved on this genome. Multiple date palm genomes have been sequenced for comparative genomics. Here we discuss a new version of the date palm genome that is linked to the genetic map. It was sequenced using a combination of Pacific Bioscience long reads and 10X Genomics linking reads. Features include an N50 scaffold length of 1.4Mb a total span of 820Mb of which 700Mb (85%) is linked to the genetic map. We believe this represents a significant improvement in the resource available to date palm researchers. We discuss the use of the genome in mapping genetic control of sucrose content in date fruit among other phenotypes.

### **W780: Perennial Grasses**

#### **Prioritizing Core Microbiome Members on Switchgrass to Inform Assembly and Host-Microbe Interactions**

**Nejc Stopnisek**, Michigan State University, East Lansing, MI

Perennial grasses are promising feedstocks for biofuel production. Native plant microbiome possesses an unprecedented resource of beneficial functions which can be leveraged to increase the productivity and stress resilience of these important crops. We characterize the 16S rRNA gene diversity and seasonal assembly of bacteria and archaea of two perennial cellulosic feedstocks, switchgrass (*Panicum virgatum* L.) and miscanthus (*Miscanthus x giganteus*). We sampled leaves and soil every three weeks from pre-emergence through senescence for two consecutive switchgrass growing seasons and one miscanthus season and identify core leaf microbial taxa based on occupancy. Virtually all leaf taxa are also detected in soil suggesting that soil is an important reservoir of phyllosphere diversity. Core leaf taxa include early, mid, and late season groups that were consistent across years and crops. This consistency in leaf microbiome dynamics and core members is promising for microbiome manipulation or management to support crop production. We have now isolated a diverse collection of bacteria from both the phyllosphere and rhizosphere of field-grown switchgrass plants, which contains many of taxa seen in our cultivation-independent core microbiome. Ongoing whole genome sequencing of cultured isolates will provide an excellent framework to improve analysis and interpretation of field-based metatranscriptome and metagenome sequencing. Together, these multi-faceted approaches provide an

exciting framework to both understand and improve the resilience of the plant microbiome, and the host plant itself.

## **W781: Perennial Grasses**

### **Novel Genomic Resources for the Polyploid Forage Grass *Eragrostis curvula***

José Carballo<sup>1,2</sup>, Cristian Gallo<sup>2</sup>, Juan Pablo Selva<sup>2</sup>, Ingrid Garbus<sup>2</sup>, Diego Zappacosta<sup>1,2</sup>, Bruno A. Santos<sup>3</sup>, Emidio Albertini<sup>4</sup>, Vicent Were<sup>5</sup>, Nick Talbot<sup>5</sup>, Matthew James Moscou<sup>5</sup>, Mario Caccamo<sup>3</sup> and **Viviana Echenique**<sup>1,2</sup>, (1)Dpto. Agronomía, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina, (2)Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) Universidad Nacional del Sur-CONICET, Bahía Blanca, Argentina, (3)NIAB, Cambridge, United Kingdom, (4)University of Perugia, Perugia, Italy, (5)The Sainsbury Laboratory, Norwich, United Kingdom

*Eragrostis* is a polyphyletic genus with more than 400 species originated from Africa and now distributed in tropical and mid-warm season regions all over the world. Weeping love grass (*E. curvula*) is an apomictic C4 grass from this genus used as forage in marginal areas due to its resistance to biotic and abiotic stresses. The basic chromosome number of this grass is X=10 with ploidy levels ranging from 2x to 8x. *E. curvula* has been classified as an orphan, or underutilized crop and despite its importance, few investigations have been performed on this species.

We have generated sequencing data for genome, transcriptome and miRNA libraries obtained from different accessions. Three genomes have been sequenced, one diploid sexual and two apomictic tetraploids and the first high density linkage map for the species was obtained following a genotyping-by-sequencing (GbS). Functional analysis of the genetic and epigenetic mechanisms underlying apomixis were initiated through the sequencing of diverse miRNA libraries and transcriptomes. Finally, 17 accessions were sequenced through RNA-seq to discover NLR diversity and the identification of resistance genes involved in resistance to blast (*Magnaporthe oryzae*).

We have produced different genomic resources for this species which we have used to advance our understanding of the mechanisms involved in diplosporous apomixis as well as the identification of resistance genes for blast. Collectively, these resources contribute to the transition of *E. curvula* from orphan to well-characterized species.

## **W782: Perennial Grasses**

### **Controlling Transgene Gene Flow *via* Pollen Biocontainment and Herbicide Mitigation in *Panicum virgatum***

**Jonathan Willis**, Ray Collier, Bryan Hernandez and Roger Thilmony, USDA-Agriculture Research Service, Albany, CA

*Panicum virgatum* (switchgrass) is a C4 perennial grass native to North America that is being developed through breeding and biotechnology programs for forage and bioenergy. Switchgrass is an obligate out-crosser, thus if biotech switchgrass is grown within its native home range, that raises concerns of transgene escape. We used an 'Alamo' A4 clonal genotype of *P. virgatum* to test the ability of four different pollen-specific promoters (*PS1*, *PS2*, *PS3*, *OsGEX2*) to express the RNAse *Barnase* and ablate transgenic pollen. Multiple independent transgenic switchgrass events were generated for each construct, their transgene copy number was measured using droplet digital PCR and low-copy lines were used for further experiments. A modified Alexander's staining method was used to assess pollen viability in the transgenic plants and multiple *PS1* and *PS2* events with apparent transgenic pollen ablation were identified. These transgenic plants were used as parents in crosses with the 'Performer' P606 clonal switchgrass genotype and T<sub>1</sub> seeds are being examined for the pollen-mediated heritable transmission of the introduced transgenes.

In an effort to mitigate the effect of potential switchgrass transgenic escapes, we are also developing a construct that confers herbicide sensitivity to switchgrass. Rice and switchgrass are normally insensitive to the herbicide Bentazon. Others have shown that silencing the expression of a specific rice cytochrome P450 (*CYP81A6*) confers herbicide sensitivity. We are deploying a similar strategy in switchgrass by

designing and testing a silencing construct that targets the putative switchgrass *CYP81A6* orthologs. Transgenic P606 plants carrying this RNAi construct have lower levels of *PvCYP81A6* transcripts based on qRT-PCR. When these silenced transgenic plants are sprayed with 1500 mg/L Bentazon, they exhibit herbicide sensitivity (yellowing and necrosis) and/or stunted growth phenotypes, while the wildtype switchgrass controls are unaffected. We hope that this research will provide viable approaches for the effective biocontainment and safe commercial release of biotech switchgrass in the future.

### **W783: Perennial Grasses**

#### **Genetic Control of Flowering Time and Plant Biomass Traits in Switchgrass**

**Yiwei Jiang**, Department of Agronomy, Purdue University, West Lafayette, IN

Switchgrass (*Panicum virgatum*) is a native and valuable bio-energy grass. Identification of genetic mechanisms of flowering time and biomass traits will facilitate genetic manipulation of switchgrass for enhancing biomass yield. The objective of this research was to identify genetic signals/candidate genes for controlling heading and anthesis and plant biomass in segregating switchgrass populations. Four pseudo-F<sub>2</sub> populations (two pairs of reciprocal crosses) were developed from lowland (late flowering) and upland (early flowering) ecotypes, and heading and anthesis dates of these populations were collected in Lafayette, IN and DeKalb, IL in 2015 and 2016. A genome-wide association study (GWAS) identified five significant signals at three loci for heading and two loci for anthesis. A homolog of FT on chromosome (Chr) 5b and a homolog of PRR5 on Chr 8a were associated with heading date across locations and/or years. Transcriptomic profiling identified six upregulated (*TOC1*, *FKF1*, two *PHYA* and two *PFT1*) and five downregulated flowering genes (*FT*, *GASA*, *COP1*, two *CHS*) in the early-flowering parent, while one upregulated *FT* and one downregulated *GASA* were found in late-flowering parent. Moreover, *PFT1* and two *FT* were upregulated in the early-flowering F<sub>2</sub> genotype, and *PRR7*, *PRR5*, *PHYB*, *TOC1*, two *CKB4* were upregulated and *COP1* was downregulated in the late-flowering F<sub>2</sub> genotype. Through linkage mapping analysis, three QTLs on Chr 2b for plant height in 2015 or crown diameter (CD) in 2016, two QTLs on Chr 2a for CD and plant biomass in 2016, and one QTL on Chr 5b for CD in 2016 were detected. Among them, QTLs on Chr 2b and 2a also controlled flowering time. The identification of key genes and QTLs involved in the flowering pathways or plant biomass accumulation could assist in the development of desirable phenotypes of switchgrass with improved biomass yield.

### **W784: Perennial Grasses**

#### **Optimizing Selective Breeding in Switchgrass (*Panicum virgatum*) Combining Genetics and Genomics**

**Serge Edme**<sup>1</sup>, Gautam Sarath<sup>1</sup>, Nathan A. Palmer<sup>1</sup>, Rob Mitchell<sup>2</sup>, Christian Tobias<sup>3</sup> and PAG Authors' Group, (1)USDA-ARS Wheat, Sorghum, and Forage Research Unit, Lincoln, NE, (2)USDA-ARS Wheat, Sorghum, and Forage Research Unit, (3)USDA-ARS, Western Regional Research Center, Albany, CA

Breeding perennial grasses, like switchgrass (*Panicum virgatum* L.), can benefit from an integration of genomic selection (GS) into the decision making of advancing candidates for breeding and for cultivar deployment based on estimated breeding values. These grasses are being targeted for breeding to meet the demands for forage and bioenergy using high biomass and/or ethanol yield, digestibility, disease resistance, and low lignin content as objective traits. A recurrent selection strategy is used by the USDA-ARS breeding program in Lincoln, NE to increase the frequency of genes or alleles based on a two-stage strategy, whereby data are collected on plot samples the first year of selection and on individual plants from the best half-sib families the following year. GS, as a genome-wide selection approach, scans the genome to find polymorphic markers in linkage disequilibrium with putative QTL (quantitative trait loci) associated with the traits of interest. The implications of bridging GS and classical genetics (i.e. PS for BLUP-phenotypic selection) will be discussed using a three-generation pedigree population derived from crossing an upland ('Summer' as ♀) and a lowland ('Kanlow' as ♂) switchgrass ecotypes and genotyped with the DArT-seq platform to isolate single nucleotide polymorphisms. Gains from selection

by GS+PS will be predicted in the contexts of a two-stage selection index strategy, reduction of generation intervals, and optimum contribution of selection candidates for breeding and deployment.

### **W785: Perennial Grasses**

#### **Large-Effect QTLs for Rust Fungus Resistance in Switchgrass are Effective in Northern, but NOT Southern Locations**

**Acer VanWallendael**, Michigan State University, East Lansing, MI

Pathogens play an important role in the evolution of plant populations, but genetic mechanisms underlying disease resistance may differ greatly between geographic areas as well as over time. Local adaptation is a vital process in plant evolution, and may be impacted by differential pathogen pressures in concert with abiotic factors. This study uses two locally adapted ecotypes of the native perennial switchgrass (*Panicum virgatum*) to examine the temporal and spatial variation in the genetic architecture of resistance to a fungal pathogen, switchgrass leaf rust (*Puccinia spp.*). The northern upland ecotype is more susceptible to fungal pathogens than the southern lowland ecotype. To identify loci underlying variation in rust resistance in switchgrass, we scored rust severity across an outcrossed mapping population that combined the upland and lowland ecotypes at eight locations across the central United States from southern Texas to Michigan. We followed rust progression at these sites for three years and mapped quantitative trait loci (QTLs) using function-valued transformations of rust progression curves. Overall, we mapped 51 QTLs that varied in presence and strength over the three-year period, and corresponded to 18 unique non-overlapping regions of the genome. Two large-effect QTLs were consistently associated with variation in rust progression in multiple sites and years, and are therefore potentially the result of the same two underlying resistance genes. These two large-effect QTLs were almost exclusively detected in northern sites, whereas southern sites showed numerous small-effect QTLs, indicating a genotype-by-environment interaction in efficacy of rust resistance loci. The beneficial alleles at northern large-effect loci reduced rust severity by 34% and increased biomass by 44%, indicating a direct benefit by pleiotropy or indirect benefit through genetic linkage. Interestingly, these two large-effect loci exhibit a negative epistatic interaction, which also varied in its strength across space. Our results suggest an important role for fungal pathogens in the local adaptation of switchgrass and illustrate an influential geographic component of the genetic architecture of plant disease resistance.

### **W786: Plant Chromosome Biology**

#### **A Centromere's Perspective on Genetic Recombination**

**Gernot Presting**, University of Hawaii, Honolulu, HI

Centromeres of domesticated have passed through genetic bottlenecks, with three and four of the ten centromeres represented by a one or two pre-domestication haplotypes, respectively. This reduced genetic diversity correlates with deletion of large regions of the ancestral centromeres that are composed of tandem repeats, and a high frequency of neocentromere formation at a nearby location. We will present data on the role of genetic recombination and selection for key genes as drivers of centromere evolution.

### **W787: Plant Chromosome Biology**

#### **Homologous and Non-Homologous Somatic Recombination Leads to Genome Remodeling during Clonal Growth of Potato**

**Luca Comai**, Plant Biology and Genome Center, UC Davis, Davis, CA

Genome integrity is dependent on controlling and balancing DNA replication, recombination, cell division, mutation, and transposition. The orchestration of these processes can fail resulting in genome instability, an outcome well documented during stress, such as when cultured cells are induced to regenerate plants. In this setting, genome instability and "somaclonal variants" are noticeable because they appear very frequently. It is possible that similar processes occur during normal growth as well, albeit a lower frequencies. Plants' flexible ploidy, meristematic growth, and lack of predetermined germline might

enable retention of karyotypic novelty, perhaps explaining the formation of sports during clonal growth. We explored this hypothesis using potato, in which we document non-homologous and homologous mitotic recombination preferentially located on certain chromosomes. These events are consistent with fragile sites and can result, alternatively, in chromoanagenesis or copy-neutral loss of heterozygosity. Our observations provide a framework for understanding both genome instability and its consequences. In addition, they suggest strategies for engineering of plant genomes.

### **W788: Plant Chromosome Biology**

#### **Motor Proteins and Tandem Repeats in Maize Meiotic Drive**

**Jonathan I Gent**, University of Georgia, Athens, GA

The knobs of maize chromosome arms stand out cytologically as even more densely-staining than the retrotransposon-rich background of the rest of the genome. They are dominated by two types of tandem repeats, a major one called *knob180*, and a minor one called *TR-1*. Knobs with either type of tandem repeat can engage the meiotic spindle and pull chromosome arms to spindle poles faster than centromeres, which leads to preferential chromosome segregation known as meiotic drive. We discovered that *knob180* knobs utilize a specialized kinesin motor protein called Kinesin Driver (KINDR), which is encoded on a rare chromosome variant called Abnormal chromosome 10 (Ab10). KINDR is closely related to a maize kinesin that functions in normal spindle organization. Now we report a second kinesin encoded on Ab10 that associates with *TR-1* knobs: TR-1 Kinesin, or TRKIN. Unlike KINDR, TRKIN has no clear relationship to any known kinesin. We are investigating the relationship between these two kinesins as well as both of their associated tandem repeats, *knob180* and *TR-1* and their roles in meiotic drive. We have also produced a reference genome for the Ab10 chromosome, which is revealing clues about the origin of both kinesins.

### **W789: Plant Chromosome Biology**

#### **Effects of CENH3 Modifications on its Deposition in Maize**

**Jing Yuan**, State Key Laboratory of Plant Cell and Chromosome Engineering, Institution of Genetics and Developmental Biology - Chinese Academy of Sciences, Beijing, China

The centromere, as an essential element to mediate chromosome segregation, is epigenetically determined by CENH3-containing nucleosomes as a functional marker; therefore the accurate deposition of CENH3 is crucial to chromosome transmission. We characterized the deposition of CENH3 in maize by over-expression and mutational analysis. Our results revealed that over-expressing CENH3 in callus is lethal while over-expressing GFP-CENH3 and CENH3-YFP in callus and plants is not and can be partly deposited normally. Different mutations of GFP-CENH3 demonstrated that CENH3-Thr4 in the N terminus was needed for the deposition as a positive phosphorylation site and the last five amino acids in the C terminus are necessary for deposition. The C terminal tail of CENH3 is confirmed to be responsible for the interaction of CENH3 and histone H4, which indicates that CENH3 maintains deposition in centromeres via interacting with H4 to form stable nucleosomes. For GFP-CENH3 and CENH3-YFP, the fused tags at the termini probably affect the structure of CENH3 and reduce its interaction with other proteins, which in turn could decrease proper deposition. Taken together, multiple amino acids or motifs were shown to play essential roles in CENH3 deposition, which is suggested to be affected by numerous factors in maize.

### **W790: Plant Chromosome Biology**

#### **The Kinetochores Protein *Kn1* Affects Spindle Assembly Checkpoint Architecture in Maize**

Handong Su, **Yang Liu** and Fangpu Han, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

The Kn1-Mis12-Ndc80 (KMN) network is an essential component of the kinetochore-microtubule attachment interface, which is required for genomic stability in eukaryotes. However, little is known about plant Kn1 proteins due to their complex evolutionary history. Here, we identified the central kinetochore

Kn1 homolog in maize (*Zea mays*) via yeast two-hybrid screening. Two KI-like motifs in Kn1 mediate its binding to the TPR domains of the spindle assembly checkpoint (SAC) components Bub1 and BubR1, but fail to interact with Bub3. The KI-like motifs were detected in monocots but not dicots, suggesting that distinct kinetochore architectures are present within plants. ZmKn1 deficiency may disrupt cell division during early endosperm development, alters the transcription of stress responsive genes and the associated metabolic reactions, thereby impairing kernel development. These findings elucidate the conserved role of the KMN network in cell division and the varied architectures of kinetochore partners during the evolution of plants.

## **W791: Plant Chromosome Biology**

### **Functional Characterization of Cowpea Centromeres**

**Takayoshi Ishii**<sup>1</sup>, Martina Juranic<sup>2</sup>, Shamoni Maheshwari<sup>3</sup>, Fernanda D. O. Bustamante<sup>4</sup>, Maximilian M. Vogt<sup>5</sup>, Rigel Salinas-Gamboa<sup>6</sup>, Steven Dreissig<sup>4</sup>, Nial Gursansky<sup>2</sup>, Tracy How<sup>2</sup>, Joerg Fuchs<sup>4</sup>, Veit Schubert<sup>4</sup>, Andrew Spriggs<sup>2</sup>, Jean P. Vielle-Calzada<sup>6</sup>, Luca Comai<sup>3</sup>, Anna M. G. Koltunow<sup>7</sup> and Andreas Houben<sup>4</sup>, (1)Arid Land Research Center, Tottori university, Tottori, Japan, (2)Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, (3)Plant Biology Department and Genome Center, University of California, Davis, (4)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany, (5)Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Switzerland, (6)UGA Laboratorio Nacional de Genómica para la Biodiversidad CINVESTAV, Mexico, (7)Queensland Alliance for Agriculture and Food Innovation (QAAFI), University of Queensland, Australia

The legume cowpea (*Vigna unguiculata*,  $2n=2x=22$ ) has significant tolerance to drought and heat stress. Here we analysed and manipulated cowpea centromere-specific histone H3 (CENH3) genes, aiming to establish a centromere-based doubled-haploid method for use in genetic improvement of this dryland crop in future. Cowpea encodes two functional CENH3 variants (CENH3.1 and CENH3.2) and two CENH3 pseudogenes. Both functional cowpea *CENH3* variants are transcribed, and the corresponding proteins are intermingled in subdomains of different types of centromere sequences in a tissue-specific manner together with the outer kinetochore protein CENPC. CENH3.2 is removed from the generative cell of mature pollen, while CENH3.1 persists. The complete CRISPR/Cas9-based inactivation of *CENH3.1* resulted in delayed vegetative growth and sterility, indicating that this variant is needed for plant development and reproduction. By contrast, *CENH3.2* knockout individuals did not show obvious defects during vegetative and reproductive development, suggesting that the gene is an early stage of subfunctionalization or pseudogenization.

## **W792: Plant Cytogenetics**

### ***SPO11-1* Mutants Provide Insights into Recombination Progression and Crossover Formation in Hexaploid Bread Wheat**

**Kim Osman**, University of Birmingham, Birmingham, United Kingdom

In most eukaryotes the distribution of meiotic crossovers (COs) along chromosomes is non-random due to multiple levels of control. In wheat and other cereals the predominantly distal location of COs creates a problem of linkage-drag in the recombinationally 'cold' centromere/proximal and interstitial regions where agronomically important traits cannot be readily separated from undesirable ones. We are investigating the factors influencing CO formation in hexaploid wheat, which has 3 sub-genomes (A, B and D), to find ways to modulate the process and unlock genetic diversity for crop improvement. In one approach we used CRISPR-Cas9 editing to introduce targeted mutations in *SPO11-1*, which encodes a key component of the highly conserved DNA topoisomerase VI-like recombination initiation complex. Fortuitously, one round of editing generated different edits in 5 of the 6 gene copies. The single remaining D-genome wild-type (WT) copy was sufficient to retain fertility and enable propagation and crossing to generate a series of lines with different allele combinations. In all cases a single WT copy from any of the sub-genomes was sufficient to maintain fertility. Two triple mutant lines which differ only in their B-genome allele were completely sterile. Cytological analysis of Pollen Mother Cells (PMCs)

revealed a complete absence of chiasmata/COs in one of the mutants and an extremely low level of residual chiasmata in the other, suggesting that these lines might be useful hosts for re-targeting recombination in the wheat genome. Immunolocalisation revealed that although the chromosome axis appeared to form normally in both mutants, chromosomes failed to synapse. Interestingly, the weaker of the two mutants still appeared to produce DNA double-strand breaks (DSBs). The nature of these is still being investigated and current data will be presented. However, the observation that DSBs are repaired yet are unable to progress efficiently to COs has clear implications for CO re-targeting experiments and underlines the need for a greater understanding of the biology of recombination initiation in large-genome polyploid plants. Ongoing detailed analysis of the effects of various allele combinations in the wheat *SPO11-1* edited lines will contribute to this.

### **W793: Plant Cytogenetics**

#### **Investigating the Role of the Meiotic Chromosome Axes in Mediating Crossover Designation in Wheat**

**Daisy Ogle**<sup>1</sup>, Stuart Desjardins<sup>1</sup>, Cristobal Uauy<sup>2</sup>, James Simmonds<sup>2</sup>, Ed Byrne<sup>3</sup>, Sybille Mittmann<sup>3</sup> and James D. Higgins<sup>1</sup>, (1)University of Leicester, Leicester, United Kingdom, (2)John Innes Centre, Norwich, United Kingdom, (3)KWS, Royston, United Kingdom

In wheat, there is a skewed crossover bias to the distal ends of chromosomes. This limits allelic exchange and the natural variation available to plant breeders. *Pachytene checkpoint protein 2 (PCH2)* is a AAA+ ATPase, involved in regulating the timing of meiotic recombination in plants. Work in *A. thaliana* has shown that PCH2 is required for depleting ASY1 from the chromosome axes, promoting full synapsis and class I crossover formation. We have obtained homozygous *pch2* TILLING mutant lines in tetraploid wheat cv. Kronos and performed a cytological analysis using meiotic metaphase I spreads and immunolocalisation, with the aim to further understand the function and explore the potential for plant breeding.

### **W794: Plant Cytogenetics**

#### **Controlling Homoeologous Pairing in Bread Wheat: A Step Towards Identification of *Ph2***

Heïdi Serra<sup>1</sup>, Radim Svačina<sup>2</sup>, Ute Baumann<sup>3</sup>, Tim Sutton<sup>4</sup>, Ryan Whitford<sup>5</sup>, Miroslava Karafiátová<sup>6</sup>, Jan Bartoš<sup>6</sup> and **Pierre Sourdille**<sup>7</sup>, (1)INRA UMR1095 GDEC, Clermont-Ferrand, France, (2)Institute of Experimental Botany, Olomouc, Czech Republic, (3)The University of Adelaide, Urrbrae, South Australia, Australia, (4)Australian Center for Plant Functional Genomics, Adelaide, Australia, (5)University of Adelaide, Glen Osmond, SA, Australia, (6)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (7)INRA UMR 1095 GDEC, Clermont-Ferrand, France

Introgression in crops of alleles originating from relatives mainly relies on homoeologous recombination. Two main genes affect this process in bread wheat (*Triticum aestivum* L.;  $2n = 6x = 42$ ), *Ph1* and *Ph2*. *Ph1* mapped on chromosome-arm 5BL and was recently identified and characterized as TaZIP4-B2. On the contrary, *Ph2* which localized on chromosome-arm 3DS, still remains to be cloned. Only two mutants developed in the international reference cultivar Chinese Spring are known for *Ph2*. *ph2a* is defined as 80 Mb-distal deletion on chromosome arm 3DS and *ph2b* is an EMS mutant. Using a high-density array of SNPs and exome capture together with the new anchored and annotated sequence of the wheat genome, we showed that *ph2a* deletion covered in fact ~121 Mb containing 1577 genes. Development of a set of 27 3DS-deletion lines allowed the restriction of the location of *Ph2* in a region of 12.3 Mb containing only 88 genes. A detailed analysis of transcriptome profiling (RNASeq) of a sub-staged meiotic time series of wheat meiocytes together with exome capture in Chinese Spring and *ph2b* mutant, open the way to identify a candidate gene, which will be subsequently verified using additional mutants derived from EMS or CRISPR-Cas9 approaches.

### **W795: Plant Cytogenetics**

## **Centromere Behaviour during Synaptonemal Complex Formation in the Female and Male Meiosis of Wheat**

**Adel Sepsi**<sup>1,2</sup>, Diána Makai<sup>1</sup>, Andrea Lenyko-Thegze<sup>1</sup>, Edit Mihók<sup>1</sup>, Attila Fabian<sup>1</sup>, Katalin Jager<sup>1</sup> and Trude Schwarzacher<sup>3</sup>, (1)Centre for Agricultural Research, Agricultural Institute, Martonvásár, Hungary, (2)Budapest University of Technology and Economics, Faculty of Chemical Technology and Biotechnology, Budapest, Hungary, (3)University of Leicester, Leicester, United Kingdom

During prophase I. of meiosis homologous maternal and paternal chromosomes recognize each other, synapse and exchange genetic information via meiotic recombination. The mechanism ensuring the fidelity of homologous chromosome pairing is still enigmatic. Its complexity is however reflected by the extensive nuclear reorganisation that precedes synapsis emergence, involving rapidly moving chromosome loops, polarized by peripheral centromere and telomere associations. Recombination and synapsis initiates at the subtelomeres and result in a higher frequency of CO distribution towards the chromosome ends in higher plants. We explore centromere and telomere behaviour during specific stages of synaptonemal complex formation in male and female meiosis of hexaploid bread wheat, providing a detailed understanding of chromosome dynamics at early meiosis. We interrogate chromosome orientation and behaviour to understand the role of the 3D nuclear organization in facilitating the asymmetry of the early events of meiosis. We propose that associations of landmark chromosomal regions, such as telomeres and centromeres, control chromatin motility during the important period of homologous pairing thus have an impact on the fidelity and coordination of pairing.

## **W796: Plant Cytogenetics**

### **Fishing eccDNA Elements that Defy Chromosome Control of Mitosis and Meiosis and Drive Rapid Adaptive Evolution**

**Bikram S. Gill**, Mithila Jugulam, Bernd Friebe and Dal-Hoe Koo, Kansas State University, Manhattan, KS

Mitosis ensures accurate copying of identical genomic material to daughter soma cells during the growth of an organism. In germ cells, meiosis requires pre-alignment of homologous chromosomes. Any aberrant chromosome(s) that may have arisen during numerous mitotic divisions, will misalign and not be passed on to the progeny. Thus, the processes of mitosis and meiosis have evolved to ensure organismal genomic integrity. While this has evolutionary advantages, it is also a liability in cases where an organism is faced with adverse stress or a xenobiotic agent such as a drug or an herbicide? Apparently, organisms have renegade genetic elements in the form of extrachromosomal circular (ecc) DNAs that are ubiquitous and can defy controls of mitosis and meiosis. The eccDNAs may arise as structural mutations (via intrachromosomal recombination as an example) during cell division leading to soma cell heterogeneity. In response to the xenobiotic agent (e.g. herbicide), rare soma cells with eccDNAs harboring target gene, can increase in copy number, fight the stress, and acquired resistance is passed on to the progeny for rapid adaptive evolution. We will describe the FISHing and visualization of eccDNA molecules, show how they defy the controls of mitosis and meiosis and lead to acquired herbicide resistance in *Amaranthus palmeri* (Koo et al. PNAS 115:332-337).

## **W797: Plant Cytogenetics**

### **Oligopaints for the Maize Chromosome Set and Some Applications**

Patrice Albert<sup>1</sup>, Tao Zhang<sup>2</sup>, Kassandra Semrau<sup>3</sup>, Jean-Marie Rouillard<sup>3</sup>, Yu-Hsin Kao<sup>4</sup>, Chung-Ju Rachel Wang<sup>5</sup>, Tatiana V. Danilova<sup>6</sup>, Jiming Jiang<sup>7</sup> and **James A. Birchler**<sup>1</sup>, (1)University of Missouri, Columbia, MO, (2)Yangzhou University, Yangzhou, China, (3)Arbor Biosciences, Ann Arbor, MI, (4)Institute of Plant and Microbial Biology, Taipei, Taiwan, (5)IPMB/Academia Sinica, Taipei, Taiwan, (6)Kansas State University, Manhattan, KS, (7)Michigan State University, East Lansing, MI

Chromosome specific oligopaint libraries were developed for each of the 10 chromosomes of maize. The B73 reference genome was computationally analyzed via k-mer analysis for unique sequences at 45 nucleotide segments using a step size of 3 nucleotides. For each of the ten chromosomes,

oligonucleotides were synthesized ranging from ~46,000 to 91,000 per chromosome. The libraries can be perpetuated once synthesized. Each library was shown to be chromosome specific and in addition had no detectable hybridization to the supernumerary B chromosome. The libraries paint the respective chromosome regardless of the line used, being genotype independent, and can identify corresponding segments in the chromosomes of maize relative *Tripsacum dactyloides*. The paints have been used to trace paired homologues in pachynema of meiosis. Translocations, transpositions and B-A translocations can be analyzed. Interphase nuclei analysis indicates that homologues occupy generalized different domains. The complete collection of chromosome paints can be used to study the structure of chromosomal aberrations, domains within the nucleus, meiotic processes and chromosomal evolution.

## **W798: Plant Disease Resistance**

### **Co-Evolutionary Mechanisms of the *Oryza sativa* and *Magnaporthe oryzae* Pathosystem**

**Yulin Jia**, USDA-ARS, Stuttgart, AR, Xueyan Wang, University of Arkansas Rice Research and Extension Center, Stuttgart, AR, Melissa Jia, USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR, Yeshi Wamishe, University of Arkansas, 130, AR and Barbara Valent, Department of Plant Pathology, Kansas State University, Manhattan, KS

Blast disease of rice caused by the filamentous fungus *Magnaporthe oryzae* is a serious threat for rice production worldwide. Major resistance (*R*) genes are effective in preventing blast disease when both the rice variety and *M. oryzae* contain the matched pair of *R* gene and avirulence (*AVR*) gene. In the USA, 8 blast *R* genes *Pi-a*, *Pi-i*, *Pi-ks*, *Pi-km*, *Pi-z*, *Pi-d*, *Pi-ta/Pi-ta2(Ptr)*, and *Pi-b* have been deployed in different commercial varieties from 1964 to 2015. To determine if deployed *R* genes influence DNA sequence changes of *AVR* genes, we analyzed *AVR* genes, *AVR-Pii*, *AVR-Pia*, *AVR-Pib*, *ACE1*, *AVR-Pita1* and *AVR-Pik* in selected blast isolates from 1964-2015 and evaluated pathogenicity with rice monogenic lines carrying these major blast *R* genes in the japonica rice variety Lijianxintuanheigu (LTH), YT14 with *Pi-ta*, and Katy with *Pita/Ptr* and *Piks*. LTH and M202 without these *R* genes were used as controls. We found deletions at *AVR-Pii* and *AVR-Pia*, point mutations at *AVR-Pik* resulting in different *AVR-Pik* variants, and transposon insertion and deletion at *AVR-Pita1* in isolates of *M. oryzae* several years after their cognate *R* genes were deployed in rice varieties. In contrast, few or no changes at *AVR-Pi9*, *ACE1* (*AVR* of *Pi-33*), *AVR-Pib* were detected indicating the changes of these *AVR* genes have not been influenced by their cognate *R* genes. In fact, *Pi-9*, *Pi-33*, *Pi-b* have not been widely deployed from 1964 to 2015 in the USA which supports that there was no selection pressure from these *R* genes. Together these findings suggest that host *R* genes may have influenced genomic changes of *AVR* genes in *M. oryzae*. A model of co-evolution of host *R* genes and pathogen *AVR* genes at the Host-pathogen interface will be presented.

## **W799: Plant Disease Resistance**

### **Mechanistic Insights into the Broad-Spectrum Resistance of a Wheat Pore-Forming Toxin-like Gene**

**Nidhi Rawat**, University of Maryland, College Park, MD

Fusarium Head Blight is a major disease of large economic significance caused by necrotrophic/hemibiotrophic fungal pathogen *Fusarium graminearum*. A Pore-forming toxin-like (PFT) gene was found to be the major determinant of resistance against *F. graminearum* in wheat. The protein does not belong to the NB-LRR type of resistance protein and is novel in multiple ways. It is a chimeric lectin, having a carbohydrate binding domain and a bacterial pore-forming toxin like domain, and does not possess a secretory peptide. Using Arabidopsis, Nicotiana and wheat as hosts for *F. graminearum*, we are working on understanding the mechanism of PFT at cellular and molecular level. The results of these studies will be presented.

## **W800: Plant Disease Resistance**

### **Characterization and Mapping of Resistance to Target Spot (*Corynespora cassiicola*) in Tomato**

**Samuel F. Hutton**, University of Florida, GCREC, Wimauma, FL

Fresh market tomato (*Solanum lycopersium*) is a high value commodity in the southeastern US, especially in Florida. The warm and humid conditions in this region are ideal for development of several diseases, including target spot, caused by *Corynespora cassiicola*. Target spot affects all aerial portions of the plant and can result in severe defoliation and reduction of marketable yields when conditions are favorable. Although the disease is common to Florida, it has become a greater concern in recent years due to increasing incidence and severity. There are currently no resistant cultivars available, and disease management relies entirely on chemical control. However, recent work has demonstrated that many *C. cassiicola* isolates are insensitive to a number of commonly used fungicide groups. In an effort to identify sources of resistance, disease assays were used to screen a collection of approximately 100 wild tomato accessions for response to target spot. Reduced infection was identified in 24 accessions from *S. pimpinellifolium*, *S. cheesmaniae* and *S. galapagense*, and this subset was used for further assays to characterize resistance and for initiation of mapping efforts to identify resistance loci. Mapping of resistance from *S. pimpinellifolium* accession LA2093 was studied using an available F9 recombinant inbred line (RIL) population from the cross between LA2093 and the susceptible line, NC EBR-1, along with published molecular marker data for these RILs. Inclusive composite interval mapping analysis revealed candidate QTLs on chromosomes 1 and 12. To characterize resistance among the 24 selected accessions, separate experiments were used to challenge these accessions individually with four diverse *C. cassiicola* isolates. A mixed model analysis revealed different levels of resistance among accessions, but accessions generally performed consistently across isolates. The highest and broadest levels of resistance were observed in LA2093, in *S. galapagense* accessions LA0483 and LA0532, and in *S. cheesmaniae* accession LA0524. Validation of the results is underway, and QTL mapping for the best *S. galapagense* and *S. cheesmaniae* accessions has also been initiated.

#### **W801: Plant Disease Resistance**

##### ***Salix purpurea* eQTL for Response to Infection by Willow Leaf Rust, *Melampsora americana***

**Dustin G. Wilkerson**<sup>1</sup>, Chase R. Crowell<sup>2</sup>, Craig H. Carlson<sup>1</sup>, Christine D. Smart<sup>2</sup> and Lawrence B. Smart<sup>1</sup>, (1)Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (2)Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY

Willow leaf rust caused by *Melampsora* spp. represent the greatest pathogen threat to shrub willow produced commercially in short rotation coppice (SRC) for biomass, often resulting in complete defoliation and yield decreases near 50 percent. Efforts to control willow rust in SRC have focused on genetic resistance and polyculture but due to its sexual phase, leaf rust often overcomes resistance before biomass production becomes profitable. Most studies in this system have occurred on European species of willow and leaf rust, creating a need for the identification of novel sources of resistance to North American species of willow leaf rust. Recent publications have helped broaden our understanding of resistance sources in the *Salix purpurea* F<sub>2</sub> population but while identifying novel QTL, this population also provides an opportunity to learn more about the transcriptional relationship between resistant and susceptible genotypes when infected with *Melampsora* leaf rust. Formed through the mating of *S. purpurea*, '94006' and '94001', then selecting and crossing two of their progeny, 'Wolcott' and 'Fish Creek', produced a population with 485 individuals. Leaf rust severity (as percent leaf area) collected in this population in 2015 and 2017 from a replicated field trial in Geneva, NY identified 25 'resistant' and 25 'susceptible' genotypes that in addition to the population's parents and grandparents have been artificially inoculated in the greenhouse with *M. americana*. Leaf punches were collected prior to inoculation and at 42 and 66 hours post inoculation, time points determined by a small pilot study to capture periods with the greatest amount of differential expression between treatment groups, for RNA extraction and sequencing using 3' RNAseq.

#### **W802: Plant Disease Resistance**

##### **Cisgenesis As Tool to Improve Disease Resistance in Apple - from the Lab to the Field**

**Ina Schlathoelter**<sup>1,2</sup>, Giovanni AL Broggin<sup>2</sup>, Bruno Studer<sup>2</sup> and Andrea Patocchi<sup>1</sup>, (1)Breeding Research, Research Division Plant Breeding, Agroscope, Wädenswil, Switzerland, (2)Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

Cisgenesis has the potential to speed up the development of resistant cultivars. Without undergoing a breeding process, it can be used to deploy natural resistances to established cultivars. New resistant cultivars could greatly reduce pesticide usage and yield losses caused by pathogens, thereby contributing to sustainable crop production. Cisgenesis is especially beneficial for vegetatively propagated crops like apple as its breeding is constrained by a long juvenility phase and allogamy. Apple breeding always results in a novel cultivar that should be able to outcompete (susceptible) established cultivars with a superior combination of favorable traits. In contrast, cisgenic plants should not differ from their original variety with the exception of the newly added trait. As a supportive tool in breeding, cisgenesis should allow a rapid adoption of disease resistant cultivars.

Here we report the development of seven cisgenic lines, three carrying the scab resistance gene *Rvi15* and four the fire blight resistance gene *FB\_MR5*. The popular apple cultivar 'Gala Galaxy', which is highly susceptible to several diseases as scab and fire blight, was chosen to be modified by cisgenesis. We assessed disease resistance in the greenhouse confirming that the approach significantly improved disease resistance. The copy number of the cisgene was assessed by qPCR and the integration site was investigated using the targeted locus amplification (TLA) approach based on short-read sequencing.

To further evaluate tree and fruit-related traits like growth habit, flowering time or fruit quality, a field trial with a fire blight resistant cisgenic line (C44.4.146) has been initiated. The design of the experiment gives the unique opportunity to compare this cisgenic line to its wild type but also to (natural) sport mutants of the same cultivar. We report on investigations of tree- and fruit-related traits as well as first results from transcriptomics analysis of uninfected leaves from the field trial. Herewith unintended effects resulting from the cisgenic process will be uncovered. Overall, we show the development of new cisgenic apple lines from their generation, initial evaluations in the greenhouse and laboratory until their final practical evaluation in the field.

### **W803: Plant Disease Resistance**

#### **The Function of Extracellular Pyridine Nucleotides and their Receptors in Plant Immunity**

**Zhonglin Mou** and Xudong Zhang, University of Florida, Gainesville, FL

The pyridine nucleotide nicotinamide adenine dinucleotide (phosphate) [NAD(P)] is a universal electron carrier that functions in metabolic reactions and intracellular signaling processes. Accumulating evidence indicates that NAD(P) also acts in the extracellular space. It has been shown in animals that necrotic cell death leads to passive release of cellular NAD(P) into the extracellular space, where it is either processed by ectoenzymes or perceived by potential cell-surface receptors, triggering outside-in signaling. We found in the model plant *Arabidopsis thaliana* that, upon wounding and bacterial infection, intracellular NAD(P) is released into the apoplast at concentrations sufficient for immune activation. We have identified two cell-surface receptors that bind NAD(P) with high affinity. Mutations of the receptors not only reduce exogenous NAD(P)-induced immune responses, but also compromise plant immunity against bacterial pathogens. Our results demonstrate that extracellular NAD(P) is a primary damage-associated molecular pattern, which plays an important role in plant immune responses. These results also suggest that exogenous application of NAD(P) and overexpression of the NAD(P) receptors are potential strategies to improve crop disease resistance. In this talk, we will present our findings in the context of the plant immune system and will compare the mechanisms utilized by plants and animals to process and/or perceive extracellular NAD(P). The function of the NAD(P)-binding receptors in plant disease resistance will also be discussed.

### **W804: Plant Epigenetics and Epigenomics**

## **Biogenesis of Short-Interfering RNAs Guiding RNA-Directed DNA Methylation: Rules of Substrate Recognition by DICER-LIKE3**

Andrew John Loffer<sup>1</sup>, Jasleen Singh<sup>1</sup>, Vibhor Mishra<sup>1</sup> and **Craig Pikaard**<sup>2</sup>, (1)Indiana University, Bloomington, IN, (2)Howard Hughes Medical Institute, Indiana University, Bloomington, IN

In plants, RNA-dependent DNA methylation (RdDM) leads to transcriptional silencing of transposons, DNA viruses and a subset of genes. The process begins with transcription of DNA into double-stranded (ds) RNA by the coupled activities of DNA-dependent RNA polymerase IV (Pol IV) and RNA-dependent RNA polymerase 2 (RDR2). DICER-LIKE3 (DCL3) then cuts the dsRNAs into short interfering RNAs that are diverse in sequence and can be 23 or 24nt in length. Using purified, recombinant DCL3, we investigated the molecular basis for siRNA size and sequence heterogeneity. Our results indicate that the molecular features of Pol IV and RDR2 transcripts, including initiating nucleotide bias, 5' end phosphorylation status, and presence of 3' overhangs, due to terminal transferase activity or internal initiation of RNA second strands, have direct consequences on DCL3 activity, affecting the polarity and size of resulting siRNAs .

## **W805: Plant Epigenetics and Epigenomics**

### **Maize and Maizecode**

**Thomas Gingeras**, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

The overarching goal of the MaizeCODE project is to produce empirical data sets for the identification of biochemically active and functional elements encoded in the genome sequences of four Maize lines (B73, W22, NC350 and Tt11). A comprehensive catalog of these elements will be a critical component in strategies to link genotype with phenotype in these important plant systems. A comprehensive analysis begins with the genomes of these four Maize lines consisting of their sequence, assembly and characterization leading to the diploid genomes of lines. These data allowed for their comparative analyses, highlighting genome-features like single nucleotide polymorphism (SNPs), insertion-deletions (INDELS) and copy number variations (CNVs). Based on the genome-wide analyses of 3-5 tissues (Endosperm, Coleoptile Nodes, Ears, Pollen, Root Tips) for each line, reference datasets have been collected that include transcriptional landscapes (long [ $>200$ nt]), short [ $>200$ nt], RAMPAGE [TSSs for each isoform]) and three types of histone modifications. In addition, using tagged strains provided by NSF-funded Maize Cell Genomics project (<http://maize.jcvi.org/cellgenomics/index.php>), has allowed us to collect tagged transcription factor (TF) lines in the B73 background to study the binding sites for each of four factors. High resolution analyses of RNA profiles has been carried out using newly developed means of cell wall digestion to generated protoplasts, coupled with efficient cell capture using combinations of the fluorescent dyes to capture specific cell layers at different depths of the B73 root by ratiometric fluorescence cell sorting and using the fluorescent protein based cell marking. These data and the corresponding metadata and analyses pipelines are available for the scientific community at CyVerse and the Maizecode website (<http://www.maizecode.org/>). Finally our outreach effort for the project at the CSHL DNA Learning Center (DNALC) in collaboration with Gramene, has developed software tools and identified quality metrics from the annotation pipeline MAKER-P to analyze how well a gene model matches available evidence and can be used as an educational resource for the community annotation of the maize genomes. PUI faculty and researchers are trained in the use of those tools that are currently being tested through course-based undergraduate research experiences (CUREs) to improve the B73 *Zea mays* v4 gene models. These data will serve as an evolving basis for further additions and improvements to be used for genome annotations. Further curation of this repository will allow community-wide standardization for plant genomics data, ease their integration into analysis pipelines.

## **W806: Plant Epigenetics and Epigenomics**

### **The Centromeres of Arabidopsis - Genetic and Epigenetic Features**

**Ian R Henderson**, University of Cambridge, Cambridge, United Kingdom

Despite *Arabidopsis thaliana* being one of the best studied model plant species gaps remain in its genome assembly, notably including the centromeres. The centromeres are known to contain many copies of the CEN180 satellite repeat which bind the centromeric histone variant CENH3, which in turn supports assembly of the kinetochore and attachment to the spindle microtubules. To provide a better understanding of the centromeres we have applied nanopore sequencing to obtain long-reads and generate a draft assembly across the centromeres. I will present these assemblies and their genetic and epigenetic features.

## **W807: Plant Epigenetics and Epigenomics**

### **The RNA NAD<sup>+</sup> Cap**

**Xuemei Chen**, University of California, Riverside, CA

## **W808: Plant Interactions with Pests and Pathogens**

### **Chemical Signaling in Plant Defense and Nodulation**

**Aardra Kachroo**, University of Kentucky, Lexington, KY

Systemic acquired resistance (SAR) is a form of broad-spectrum immunity in plants, which is induced in response to local infections and protects uninfected parts against subsequent secondary infections. SAR involves the generation of mobile signals at the site of primary infection, which are transported to and arm distal portions of the plant. Research in the last decade has identified a number of diverse chemical signals that are important for SAR. These chemicals function in two parallel branches to activate SAR. One branch is regulated by salicylic acid (SA) and functions in parallel with the second branch comprising piperolic acid (Pip), azelaic acid (AzA) and glycerol-3-phosphate (G3P). AzA and G3P function downstream of the free radicals nitric oxide (NO) and reactive oxygen species (ROS). The plant galactolipids, monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG), are also required for SAR and of these DGDG contributes to both NO and SA biosynthesis. SAR signaling is highly conserved amongst diverse plants and interestingly, some of the SAR-inducing chemicals also regulate systemic signaling which is important for the genetic exclusion of poor nitrogen-fixing, root-nodulating bacteria in legumes. The interrelationships of the various SAR chemicals and their systemic transport routes which eventually regulate immune responses, will be discussed.

## **W809: Plant Interactions with Pests and Pathogens**

### **Molecular Characterization of Hessian Fly Resistance Genes in *Aegilops tauschii***

Yunming Long<sup>1</sup>, Wegi Wuddineh<sup>2</sup>, Kirk Anderson<sup>1</sup>, Guotai Yu<sup>3</sup>, Qijun Zhang<sup>1</sup>, Sanu Arora<sup>3</sup>, Burkhard Steuernagel<sup>4</sup>, Ming-Cheng Luo<sup>5</sup>, Xiwen Cai<sup>6</sup>, Yong Q. Gu<sup>7</sup>, Justin D. Faris<sup>2</sup>, Mingshun Chen<sup>8</sup>, Brande B.H. Wulff<sup>3</sup>, Evans Lagudah<sup>9</sup>, Marion Harris<sup>1</sup> and **Steven S. Xu<sup>2</sup>**, (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND, (3)John Innes Centre, Norwich, United Kingdom, (4)The Sainsbury Laboratory and John Innes Centre, Norwich, United Kingdom, (5)University of California, Davis, CA, (6)Department of Plant Science, North Dakota State University, Fargo, ND, (7)USDA ARS, Western Regional Research Center, Albany, CA, (8)Department of Entomology, Kansas State University, Manhattan, KS, (9)Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT, Australia

*Aegilops tauschii* Cosson, the D-genome donor of bread wheat (*Triticum aestivum* L.), provides a vast reservoir of genetic variation for wheat improvement in disease and insect resistance. For the resistance to Hessian fly (*Mayetiola destructor* Say), six resistance genes (*H13*, *H22*, *H23*, *H24*, *H26*, and *H32*) have been identified in *Ae. tauschii*. The objectives of this study were to identify and characterize Hessian fly resistance genes from a worldwide collection of *Ae. tauschii* accessions deposited in the USDA-ARS National Small Grains Collection and Wheat Genetics Resource Center. To identify the resistance, we evaluated approximately 600 *Ae. tauschii* accessions for reactions to the Hessian fly biotype Great Plains and identified 90 resistant accessions. To identify potentially novel genes, we

conducted fine mapping of the resistance genes in three *Ae. tauschii* accessions (Clae 25, RL 5271, and TA 2377) using the large F<sub>2</sub> populations derived from their respective crosses with a susceptible *Ae. tauschii* accession AL8/78. Two tightly-linked resistance genes were identified from Clae 25 and mapped to the *H26* region (3DL). The resistance genes in RL 5271 and TA 2377 were both mapped onto the genomic region harboring *H13* (6DS). By surveying the genomic sequences of approximately 260 *Ae. tauschii* accessions, we identified several candidate genes in RL 5271 and TA 2377. Sequence analysis revealed that one of the candidate genes is located in the same region as the *H13* gene in the wheat genotype Molly, with two single-nucleotide mutations in comparison with the *H13* candidate, indicating that the Hessian fly resistance gene in RL 5271 and TA 2377 is likely a new haplotype of *H13*. This study also demonstrated that several simple sequence repeat (SSR) and semi-thermal asymmetric reverse PCR markers that co-segregate with or are tightly linked to these targeted genes are useful for developing resistant cultivars and cloning resistance genes for Hessian fly.

## **W810: Plant Interactions with Pests and Pathogens**

### **The Molecular Battle between the Bacterial Wilt Pathogen and its Plant Hosts**

**Raka Mitra**, Carleton College, Northfield, MN

*Ralstonia solanacearum* causes one of the most devastating bacterial diseases of plants worldwide, affecting hundreds of plant species including many major crops such as tomato and potato and the model plant *Arabidopsis thaliana*. *Ralstonia* typically infects plants through the root systems and ultimately colonizes the plant vasculature, where it interferes with water transport, resulting in plant wilting and death. During plant invasion, *Ralstonia* employs a type III secretion system to deliver an array of effector proteins directly into the plant cell. The role of many of these effectors in bacterial wilt disease has not been explored. For leaf-invading pathogens, effectors typically target and interfere with host defense pathways. As a large number of *Ralstonia* effectors are not found in other pathogens, studies of these effectors may reveal novel plant pathways that are targeted during pathogen invasion, possibly illuminating novel aspects of plant root-based defenses.

We are focusing upon three approaches for studying the *Ralstonia*-plant interaction. First, we are characterizing the suite of conserved effector proteins employed by a variety of *Ralstonia* strains. Students in the Carleton College Cell Biology Lab class are investigating the localization of these GFP-tagged bacterial effector proteins *in planta*. We have identified *Ralstonia* effectors with cytoplasmic, nuclear and membrane localization, indicating a variety of host targets. Second, we have initiated yeast two-hybrid studies to study tomato proteins that interact with *Ralstonia* effectors. Third, we are performing additional studies on nuclear-localized effectors in order to better understand their functions. Through these studies, we hope to develop a better understanding of the interface between the bacterial pathogen and plant cell during bacterial wilt disease.

## **W811: Plant Interactions with Pests and Pathogens**

### **Molecular Interactions between Soybean Aphids and Aphid-Resistant Soybean**

**Andrew Michel**<sup>1</sup>, Ashley Yates-Stewart<sup>2</sup>, Josquin Daron<sup>3</sup>, Saima Shahid<sup>4</sup> and R. Keith Slotkin<sup>4</sup>, (1)Dept. of Entomology, OARDC - The Ohio State University, Wooster, OH, (2)Translational Plant Sciences, The Center for Applied Plant Sciences, The Ohio State University, Wooster, OH, (3)The Ohio State University, Columbus, OH, (4)Donald Danforth Plant Science Center, St. Louis, MO

The interaction between plants and insects involves critical adaptations and counter-adaptations as insects evolve to overcome plant defenses. Understanding these interactions at the molecular level can help identify factors that facilitate adaptation and improve the sustainability of insect resistant crops. The soybean aphid (*Aphis glycines*), is an invasive pest that significantly impacts soybean production in North America. While a few soybean varieties have natural resistance to the soybean aphid, certain soybean aphid populations have adapted to overcome this resistance (i.e. virulent biotypes). We compared differential gene expression among virulent and avirulent soybean aphid populations on aphid-

susceptible and aphid-resistant soybean. Overall, the aphid biotypes differentially express ~2,000 regardless of plant host. Virulent soybean aphids showed decreased expression of many putative effector proteins, which modulate plant defenses; in this case decreased effector expression may enable virulent biotypes to evade detection by plant defenses. Interestingly, we also observed higher expression of several transposable elements in the virulent biotype, which affected the expression of proximal genes. Bisulfite sequencing of soybean aphid biotype genomes indicated that some of this differential gene expression is related to unique patterns of DNA methylation among soybean aphid biotypes. Our presentation will discuss mechanisms of virulence adaptation in the soybean aphid, including possible epigenetic causes, and how these can be used to improve the durability and sustainability of aphid-resistant soybean.

## **W812: Plant Interactions with Pests and Pathogens**

### **Signal Transduction in Resistance to Bacterial Speck Disease in Tomato**

**Fangming Xiao**, University of Idaho, Moscow, ID

In tomato, resistance to bacterial speck disease caused by *Pseudomonas syringae* pv. *tomato* (*Pst*) is determined by the intracellular immune receptor Prf. The activation of Prf is mediated by perception of *Pst*-derived effector AvrPto or AvrPtoB by the Pto kinase, leaving it largely unknown the transmission of a defense signal from the activated Prf to downstream defense events. We have found that the activated Prf interacts with and stabilizes the defense-related SINAC1 transcription factor, which otherwise is highly unstable due to the ubiquitin ligase SISINA3-mediated ubiquitination and degradation. Significantly, SINAC1 is a positive regulator of the Prf-mediated defense signaling and transgenic tomato overexpressing SINAC1 exhibit enhanced resistance to *Pst*. Our findings support a hypothesis that the signaling-competent Prf interferes with SINAC1 ubiquitination to prevent its degradation by sequestering SINAC1 away from the ubiquitin ligase SISINA3, thereby enhancing the transcriptional potential of SINAC1 and enabling robust transcriptional reprogramming.

## **W813: Plant Interactions with Pests and Pathogens**

### **Insect Control using Proteins from Non-Bt Sources and Expression through Novel Regulatory Elements**

**Scott Diehn**, Corteva Agriscience, Johnston, IA

Western corn rootworm (WCR) is a major pest of corn in North America and Europe. Damage caused by WCR has the potential to cost U.S. growers billions (US\$) in yield loss annually if not controlled. Current commercial traits based on proteins identified from *Bacillus thuringiensis* (Bt) bacteria have been an important tool to reduce yield loss for farmers. Reports of field resistance to current commercial traits highlight the need for new traits based on new modes of action. Corteva Agriscience<sup>™</sup> has pursued multiple strategies including screening of non-Bt sources for proteins that can be developed into insect resistance traits. These efforts identified several proteins that provide root protection from WCR feeding when expressed in transgenic plants. Non-Bt sources offer a rich avenue of actives for the development of future insect resistance traits.

## **W814: Plant long non-coding RNAs**

### **Characterisation of Plant Telomerase RNAs**

Petr Fajkus<sup>1</sup>, Vratislav Peška<sup>1</sup>, Michal Závodník<sup>2</sup>, Miloslava Fojtova<sup>2</sup>, Jana Fulneckova<sup>1</sup>, Agata Magdalena Kilar<sup>2</sup>, Jason Alessio Sims<sup>3</sup>, Eva Sýkorová<sup>1</sup> and **Jiri Fajkus**<sup>1</sup>, (1)Institute of Biophysics, CAS, v.v.i., Brno, Czech Republic, (2)Masaryk University, CEITEC and Faculty of Science, Brno, Czech Republic, (3)Max Perutz Labs, University of Vienna, Vienna, Austria

Phylogenetic divergence in Asparagales plants is associated with changes in otherwise widely conserved telomere DNA sequences. To elucidate these evolutionary telomere changes, we aimed to characterise telomerase RNA subunits (TRs) in these plants. This was a challenging task since the only

predictable region in these TRs is a short region that serves as a template for telomere synthesis by the telomerase reverse transcriptase. Therefore, we took advantage of the unusually long telomere repeat unit in *Allium* plants (12 nt) which allowed us to identify TRs in transcriptomic data of representative *Allium* species. Orthologous TRs were then identified in Asparagales plants harbouring TTAGGG (human-type) or TTTAGGG (*Arabidopsis*-type) telomeres. Further, we identified TRs across the land plant phylogeny, including model plants, crop plants, and plants with unusual telomeres. We demonstrate the templating function of these TRs and disprove a functionality of the only previously reported plant telomerase RNA in *Arabidopsis*. Our results change the existing paradigm in plant telomere biology which has been based on a conserved telomerase reverse transcriptase (TERT) associating with highly divergent TRs even between closely related taxa. The finding of a monophyletic origin of plant TRs opens the possibility to identify TRs directly in transcriptomic or genomic data and/or predict telomere sequences synthesized according to the respective TR template region. This knowledge expands existing telomerase research to include plant systems which (i) differ in lifespan and developmental strategies; (ii) must efficiently cope with environmental changes and (iii) could be easily regenerated from their totipotent cells.

*Funding:* Czech Science Foundation [17-09644S], ERDF [project SYMBIT, reg. no. CZ.02.1.01/0.0/0.0/15 003/0000477] and Ministry of Education, Youth and Sports of the Czech Republic [project CEITEC 2020 (LQ1601)].

## **W815: Plant long non-coding RNAs**

### **Elucidating the Function of a Novel Long Non-Coding RNA during *Arabidopsis* development**

**Marianne Kramer**<sup>1</sup>, Kyle R Palos<sup>2</sup>, Andrew Nelson<sup>3</sup>, Eric Lyons<sup>4</sup>, Mark A Beilstein<sup>2</sup> and Brian D. Gregory<sup>1</sup>, (1)University of Pennsylvania, Philadelphia, PA, (2)University of Arizona, Tucson, AZ, (3)Boyce Thompson Institute, Ithaca, NY, (4)The University of Arizona, Tucson, AZ

We previously identified a cohort of nuclear lincRNAs that were protein bound. These protein-bound lincRNAs were significantly more conserved than those without protein binding sites, indicating that these lincRNAs may have important functions in *Arabidopsis*. We have termed these lincRNAs *CONSERVED IN BRASSICA RAPA 1-14 (CONBR1-14)*.

After screening several *CONBR* mutants, we began to look closer at *CONBR1* since it contained two small nucleolar RNAs (snoRNAs). In mammals, two snoRNA containing lincRNAs have been studied and shown to be important in specific diseases, thus we hypothesized that since *CONBR1* contains two snoRNAs it may have an interesting function.

To determine if *CONBR1* had an important function in *Arabidopsis*, we obtained an insertion mutant within the lincRNA (*conbr1-1*) as well as generated a CRISPR mutant with a large deletion in the 5' end of the transcript (*conbr1-2*). Interestingly, both *conbr1-1* and *conbr1-2* mutant plants are significantly smaller than WT plants. To test if the plants were developmentally delayed, we examined germination efficiency of mutants compared to wild-type (WT) plants. After 48hrs, significantly fewer mutant plants have germinated as measured by emergence of their cotyledons from the seed coat.

We then introduced *CONBR1* back into *conbr1-1* plants (*conbr1-1/CONBR1pro::CONBR1*) under its endogenous promoter to examine if we can rescue the mutant phenotype. *conbr1-1/CONBR1pro::CONBR1* plants had WT levels of *CONBR1* expression and were similar in size and germination to WT plants, indicating we can rescue the developmental phenotype by reintroducing *CONBR1* expression into the mutants. Further, we introduced *CONBR1* under its endogenous promoter into WT background to examine plants overexpressing *CONBR1* (*Col-0/CONBR1pro::CONBR1*). Among several alleles, we see anywhere from a 5-20-fold increase in *CONBR1* levels. We find that *CONBR1* overexpressing plants appear to be larger than WT plants and appear to have a slight increase in germination efficiency, indicating that *CONBR1* may in fact be an important regulator of early *Arabidopsis* development.

To examine the molecular function of *CONBR1*, we first confirmed that *CONBR1* was nuclear localized using qPCR on nuclear and cytoplasmic RNA fractions from WT plants. Further, given the developmental delay phenotype, we tested when during development *CONBR1* is expressed and found that *CONBR1* is highly expressed early during early development, peaking in expression in 2-day-old seedlings, which agrees with previously published RNA-seq data.

Nuclear localized lincRNAs have been previously demonstrated to bind to a variety of proteins such as transcription factors and histone modifying enzymes to regulate gene expression. To determine if *CONBR1* functions in a similar manner, we are performing chromatin isolation by RNA purification followed by DNA sequencing (ChIRP-seq) to identify regions throughout the genome that *CONBR1* binds as well as ChIRP followed by mass spectrometry to identify the proteins bound to *CONBR1*. ChIRP uses biotinylated DNA probes antisense to *CONBR1* to pull down *CONBR1* as well as any regions of DNA and proteins bound. Further, we are performing mRNA-sequencing in WT and *conbr1-2* to determine what global changes in gene expression occur when *CONBR1* is knocked down.

### **W816: Plant long non-coding RNAs**

#### **MINE the Gap: LncRNA Identification as a Bridge between Computational and Bench Approaches in the Undergraduate Classroom**

**Rebecca Murphy**, Centenary College of Louisiana, Shreveport, LA and Andrew Nelson, Boyce Thompson Institute, Ithaca, NY

As emerging technological advances make computationally-driven, high-throughput assays increasingly accessible, acquiring the skills necessary to navigate the creation and analysis of the resulting large data sets becomes increasingly important for undergraduates in an ever-expanding range of STEM fields. However, gaining initial exposure to complex data analysis can be daunting for both students and instructors, particularly at primarily undergraduate institutions (PUI).

To introduce undergraduate students to this type of analysis, we have developed a teaching module for undergraduates that gives them exposure to mining publicly available transcriptome data from NCBI, mapping sequencing reads to reference genomes, and identifying novel long non-coding RNAs from these data using the RMTA and EvolIncl apps from CyVerse. Using this type of *in silico* approach combines multiple skill sets, encourages critical thinking, and translates directly to the bench through inquiry guided lincRNA validation experiments using PCR.

Furthermore, incorporating computational approaches into traditionally molecular-based research programs can be transformative for faculty at PUIs, where performing research presents a unique set of challenges. By applying the lincRNA identification pipeline both to publicly available data and RNAseq data generated through our own experiments, we have identified putative lincRNAs in several important crop species, and greatly expanded our research capabilities at Centenary College of Louisiana.

### **W817: Plant long non-coding RNAs**

#### **Comparison of lincRNAs identified through Nanopore and short read RNA-sequencing strategies**

**Sateesh Peri**, Genetics Graduate Interdisciplinary Group, University of Arizona, Tucson, AZ

Long non-coding RNAs (lincRNAs) are transcripts greater than 200 nt that are functionally diverse and yet poorly defined. Thanks to high throughput next-generation sequencing technologies, we are just beginning to unravel the ubiquity and functions of these enigmatic transcripts. To date, we have processed > 30,000 publicly available, short read Illumina RNA-seq experiments across the plant kingdom, creating the most exhaustive lincRNA curation and detailing when and where they are expressed. While these data promise to be incredibly useful, due to limitations in short-read sequencing, as well as transcript assembly algorithms, both false positives and negatives are abundant. Due to read-size restrictions in short-read sequencing technologies, there has always been a difficulty in generating full-length transcripts. This is compounded by the on-average low expression of lincRNAs, which leads to

errors in transcript assembly. To address these issues we have applied Oxford Nanopore Technology (ONT) long read sequencing technology to sequence full-length transcripts from several species within the Brassicaceae. Here in this study we present a comparison of short-read and long read identified *Arabidopsis thaliana* lncRNAs. In general, long read sequencing more accurately distinguishes between non-coding and coding due to intact and fully assembled ORFs. We also observe a much higher exon count, longer average length, and deeper evolutionary conservation for lncRNAs derived from ONT data. These data demonstrate the necessity of long read data, but also the usefulness of marrying these two technologies when addressing questions of differential expression, conservation, and ultimately functional characterization of lncRNAs.

### **W818: Plant long non-coding RNAs**

#### **The Landscape of Coding and Non-Coding RNAs Defined by Enhanced Super-Resolution Ribosome Profiling**

**Larry Wu**, Michigan State University, East Lansing, MI and Polly Hsu, Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI

In many organisms, genes that contain open reading frames with more than 100 amino acids are annotated as coding, and the rest are presumed to be non-coding. However, recent studies have shown that some of the small ORFs in these “non-coding” RNAs (ncRNAs) could be translated and have important biological functions. To systematically define small translated ORFs in annotated ncRNAs, we exploit ribosome profiling, the deep sequencing of ribosome footprints, to investigate genome-wide mRNA translation in *Arabidopsis*. With the improved read coverage of our super-resolution ribosome profiling method, we found hundreds of new small ORFs and unexpected regulatory features within annotated ncRNAs. In contrast, under our experimental conditions, some genes that are annotated as coding appear to be non-coding. In summary, our ribosome profiling method efficiently identifies small ORFs on ncRNAs, verifies the translation of annotated coding genes and reveals the important roles of translation in the biogenesis of ncRNAs in plants.

### **W819: Plant long non-coding RNAs**

#### **Targets of Opportunity: The Birth of Orphan Genes, Exemplified in Yeast, Arabidopsis, Maize and Humans**

**Eve Syrkin Wurtele**<sup>1</sup>, Urminder Singh<sup>2</sup>, Priyanka Bhandary<sup>1</sup>, Jing Li<sup>1</sup> and Arun S. Seetharam<sup>1</sup>, (1)Iowa State University, Ames, IA, (2)Iowa State University

Each organism contains genes that encode proteins with no homolog in other species (“orphan genes”). Some orphan genes have arisen *de novo* from non-genic material, others from within lncRNAs, others from novel reading frames within protein-coding genes, while others result from ultra-rapid mutation of existing genes. The challenge of distinguishing protein-coding orphan genes in genomes and predicting their functions is immense, resulting in under-appreciation of their importance.

Many transcripts containing open reading frames (ORFs) that bear no homology to other proteins are expressed and translated, but are annotated as lncRNAs, or not annotated as genes at all. Under the premise that some of these are protein-coding orphan genes, we created an aggregated dataset from RNA-seq raw reads in NCBI-Sequence Read Archive from four diverse eukaryotes: *Saccharomyces cerevisiae*, *Homo sapiens*, *Arabidopsis thaliana* and *Zea mays*. These datasets, comprising between 3000 and 15,000 samples for each species, were realigned to the respective genomes, and ORFs within these transcripts were subjected to phylostratigraphic analysis. In yeast alone, 15,806 transcripts contain ORFs inferred to be orphans (“orphan-ORFs”), about 40% of which are ribosome-bound. Taken together, these data help distinguish those transcripts that may be protein-coding proto-genes or genes, from lncRNAs, and from the random transcription that may provide fodder for new genes. We provide aggregated RNA-Seq datasets with sample metadata in MetaOmGraph (MOG), a tool enabling powerful, interactive, statistical analysis and visualization of specific transcripts under user-selected conditions.

This approach enables reuse of these data for exploratory discovery and provides a rich context for experimentalists to identify and make novel, experimentally-testable hypotheses about candidate genes.

### **W820: Plant Molecular Breeding**

#### **Exploiting Genetic Variation of *Gossypium* Gene Pools for Cotton Improvement**

**John Z. Yu**, James E. Frelichowski, Lori L. Hinze and Joshua A. Udall, USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX

The genus of *Gossypium* consists of more than 50 species that provide primary and secondary gene pools essential for cotton improvement. Yet the genetic potential of these gene pools is barely exploited due to various factors including cotton industry's transgenic breeding that relies largely on very few transformable genotypes. Characterization of significant variants inherited through the cotton breeding process has been challenging. With the recent sequencing of several diploid and tetraploid cotton genomes, abundant DNA variation is uncovered within and among the *Gossypium* species and accessions. In concert with the phenotypic observation, such DNA variants are being associated to priority traits including fiber and seed quality properties as well as biotic and abiotic stress tolerance. Now it is possible to exploit the beneficial genetic variation that may facilitate and expedite molecular breeding of cultivated cottons. Desirable genotypes can be discovered for specific cotton breeding programs. The current status and future perspective will be discussed.

### **W821: Plant Molecular Breeding**

#### **Phenotypic Implications of Structural Genome Variation in Elite Winter Canola (*Brassica napus*)**

**Paul Vollrath**, Justus Liebig University, Giessen, Germany

In three-year field trials at eight locations we phenotyped an interconnected multiparental population comprising of 354 accessions for a number of traits. The population is derived from crosses of six elite canola lines with a common parental line. Each subpopulation consists of 60 individual DH lines. The whole population was genotyped by the 60k SNP chip and GWAS was conducted. Whole genome sequencing using Oxford Nanopore Technology was performed with a coverage of 20x for all of the seven parental lines. The aim was to identify segregating long- and short-range deletions and insertions associated with traits of interest. Using the MinION device we reached an average N50 of ~25 kb and a mean read length of ~10 kb. For detecting structural genome variation, the alignment based approach and the reference genome Darmor-*bzh* v4.1 were used. By the use of the *NGMLR* aligner and the variant caller *Sniffles* we detected 5818 deletions and 3635 insertion within these lines. Developmental traits and disease resistance were found to be associated with structural genome variation.

### **W822: Plant Molecular Breeding**

#### **QTG-seq Accelerates QTL Fine-Mapping in the Era of “Big Data”**

**Lin Li**, Huazhong Agricultural University, Wuhan, China

Deciphering the genetic mechanisms underlying agronomic traits is of great importance for crop improvement. Most of these traits are controlled by multiple quantitative trait loci (QTL), and identifying the underlying genes by conventional QTL fine-mapping is time-consuming and labor-intensive. Here, we devised a new method we named quantitative trait gene sequencing (QTG-seq) to accelerate QTL fine mapping. QTG-seq combines QTL partitioning to convert a quantitative trait into a near-qualitative trait, bulked segregant sequencing on a large segregating population, and a robust new algorithm for identifying candidate genes. Using QTG-seq, we fine-mapped a plant height QTL in maize (*Zea mays* L.), *qPH7*, to a 300-kb genomic interval and verified that a gene in that region encoding an NF-YC transcription factor was the functional gene. Molecular evidence suggested that *qPH7* might influence plant height by interacting with proteins encoded by a CO-like gene and an AP2 domain containing gene. Selection analysis indicated that *qPH7* was subject to strong selection during maize improvement. In summary, QTG-seq provides an efficient method for QTL fine-mapping in the era of “big data”.

### **W823: Plant Molecular Breeding**

#### **Understanding and Addressing the Complexity of Integrating Disease Resistance from Unadapted Germplasm in Pepper**

**Allen Van Deynze**<sup>1</sup>, Jareerat Chunthawodtiporn<sup>2</sup>, William Rehrig<sup>3</sup>, Kevin Stoffel<sup>1</sup>, Armando Garcia-Llanos<sup>1</sup>, Shiyu Chen<sup>4</sup>, Randi C. Jimenez<sup>4</sup> and Theresa Hill<sup>5</sup>, (1)University of California, Davis, CA, (2)Kasesart University, Thailand, (3)Enza Zaden, (4)University of California, Davis, Davis, CA, (5)Department of Plant Sciences, Seed Biotechnology Center, University of California, Davis, California, USA, Davis, CA

Pepper is a fruit that has large variation in fruit shape, fruit size and horticultural traits. Pepper varieties are also routinely attacked by pathogens. Using unadapted sources of germplasm to bring in novel alleles, for disease resistance while maintaining favorable quality and horticultural traits is challenging for breeding in pepper. Fortunately, genetic diversity and resistance exists in crossable landraces of *Capsicum annuum*, the primary cultivated species. In an effort to understand complexity of disease resistance from multiple sources, we characterized the inheritance and developed molecular tools, strategies and germplasm to introgress quantitative disease resistances while maintaining fruit quality and horticultural traits for elite germplasm. We will present our understanding of disease resistances, plant architecture, as well as complexity of fruit quality including fruit size, shape and harvestability at the phenotypic and genomic levels. The incorporation of genomic tools associated with these traits has allowed us to develop integrated approaches to combine these complex traits into cultivated fresh market and processing pepper types.

### **W824: Plant Molecular Breeding**

#### **Variance Heterogeneity Genome-Wide Mapping for Cadmium in Bread Wheat Reveals Novel Genomic Loci and Epistatic Interactions**

**Gota Morota**<sup>1</sup>, Waseem Hussain<sup>2</sup>, Malachy Campbell<sup>3</sup>, Diego Jarquin<sup>4</sup> and Harkamal Walia<sup>3</sup>, (1)Virginia Polytechnic Institute and State University, Blacksburg, VA, (2)International Rice Research Institute, Los Banos, Philippines, (3)University of Nebraska Lincoln, Lincoln, NE, (4)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE

Genome-wide association mapping identifies quantitative trait loci (QTL) that influence the mean differences between the marker genotypes for a given trait. While most loci influence the mean value of a trait, certain loci, known as variance heterogeneity QTL (vQTL) determine the variability of the trait instead of the mean trait value (mQTL). In the present study, we performed a variance heterogeneity genome-wide association study (vGWAS) for grain cadmium (Cd) concentration in bread wheat. We used double generalized linear model and hierarchical generalized linear model to identify vQTL associated with grain Cd. We identified novel vQTL regions on chromosomes 2A and 2B that contribute to the Cd variation and loci that affect both mean and variance heterogeneity (mvQTL) on chromosome 5A. In addition, our results demonstrated the presence of epistatic interactions between vQTL and mvQTL, which could explain variance heterogeneity. Overall, we provide novel insights into the genetic architecture of grain Cd concentration and report the first application of vGWAS in wheat. Moreover, our findings indicated that epistasis is an important mechanism underlying natural variation for grain Cd concentration.

### **W825: Plant Molecular Breeding**

#### **Accelerating Molecular Breeding of Octoploid Strawberry through Development of Allo-Polyploid Reference Genomes and Subgenome-Specific Marker Platforms**

**Michael A Hardigan**<sup>1</sup>, Mitchell J. Feldmann<sup>1</sup>, Dominique D.A. Pincot<sup>1</sup>, Nicolas Cobo<sup>1</sup>, Anne Lorant<sup>1</sup>, Mirko Ledda<sup>1</sup>, Randi Famula<sup>1</sup>, Patrick Edger<sup>2</sup> and Steven J. Knapp<sup>1</sup>, (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Department of Horticulture, Michigan State University, East Lansing, MI

The garden strawberry (*Fragaria × ananassa* Duchesne ex Rozier) poses a unique challenge for plant breeders due its octoploid genome ( $2n=8x=56$ ) and interspecific origin. Today's cultivars arose through a combination of ancient polyploid hybridization, followed by repeated homoploid hybridization in the modern era. Strawberry lagged behind diploid crops in the development of molecular breeding tools due to the associated genomic complexity. Using the recently published Camarosa v1.0 octoploid genome, which discriminates ancestral subgenome homology at the assembly level, we identified an unbroken distribution of 41.8M subgenome-specific, diploid-behaving DNA variants spanning the octoploid genome of *F. × ananassa* and its progenitors. We utilized this diversity to construct and validate 850K and 50K single nucleotide polymorphism genotyping arrays, and begin molecular breeding for flowering time, disease resistance, and fruit quality within the University of California strawberry breeding program. With the challenge of resolving and assaying subgenomic variation addressed by up-front bioinformatics, we successfully performed mapping, genome-wide association, and genomic prediction studies for multiple traits in an octoploid plant breeding program using straightforward diploid approaches, including the economically valuable day-neutral flowering gene. We are working to further expand the molecular breeding toolbox for octoploid strawberry by developing phased long-read assemblies for *F. × ananassa* and beach strawberry (*F. chiloensis*). To support this work, we used recombination breakpoint prediction to map 1.6M Camarosa variants in 3,179 haploblocks spanning 2,017 cM and 1.9M beach strawberry variants in 5,521 haploblocks spanning 3,394 cM. These phased haplotype maps provide a framework for separating homeologous and homologous sequences to maximize octoploid genome quality. We have now successfully assembled parent-specific chromosome haplotypes of a heterozygous octoploid hybrid individual.

## **W826: Plant Phenotypes**

### **Laying the Foundations: Why Are Semantics Difficult in Agriculture?**

**Robert P. Davey**, Earlham Institute, Norwich, United Kingdom

Integrating plant datasets is typically labour intensive. Characterization of traits and measurements of treatments are vital parts in the process of crop experimentation, but datasets produced from this process are often unstructured. Furthermore, FAIR data is becoming a practice that crop scientists are urged to follow, but whilst there is much information about what FAIR is, there is far less information on how to implement FAIR practices. How can we use technology and societal processes to help capture the required context to power future methods of analysis?

## **W827: Plant Phenotypes**

### **Communication across Multidisciplinary Lines for Plant Phenotyping Research: An Engineer's Perspective**

**Amy Tabb**, USDA-ARS, Kearneysville, WV

Plant phenotyping projects with multidisciplinary teams have the potential to address interesting research questions, but with this potential comes added challenges. These challenges stem from differences in academic discipline, and are related to topics such as incentive structure, publishing venues, definitions of project success, terminology, and others. In this talk, I will outline some of these differences from my perspective as a computer engineer who has collaborated with plant biologists in plant phenotyping projects. I will also present some communication strategies for desirable outcomes in these types of projects, supported by examples from my own research program.

## **W828: Plant Phenotypes**

### **UAV Imaging and Analysis for Field Phenotyping**

**Ian Stavness**, University of Saskatchewan, Saskatoon, SK, Canada

Low-cost aerial imaging platforms, such as consumer drones, have the potential to democratize image-based field phenotyping by making image acquisition systems cheaper and easier to replicate. However, a fast and efficient image processing pipeline to go from raw drone images to per-plot phenotypic data

remains a bottleneck for the utilization of image-based phenotyping in large-scale applications. In this talk, I will present recent work on localizing micro-plots and rows within field trials from aerial drones. I will introduce our PlotVision platform, which speeds up the process of aerial image analysis for plant breeding experiments. I will also discuss recent image analysis and machine learning approaches for estimating phenotypes from aerial images.

## **W829: Plant Phenotypes**

### **Combining Phenomics, Modeling and Genomic Prediction for Identifying Ideotypes Adapted to a Changing Climate**

**François Tardieu**<sup>1</sup>, Boris Parent<sup>2</sup>, Llorenc Cabrera Bosquet<sup>2</sup> and Claude Welcker<sup>3</sup>, (1)INRA, Montpellier, France, (2)INRA, France, (3)INRA, montpellier, France

Seeking alleles for high yield in a variable environment poses a problem of time scales. Yield involves feedbacks operating over months, whereas physiological mechanisms of acclimation operate over minutes, following the variability of environmental conditions. Phenomics needs to analyze the genetic variability at these different scales under a range of environmental conditions, with a suite of installations with different temporal definitions and measured traits. Multi-site field experiments show that a given allele, resulting in a given value for trait measurement in a given environment, can result in positive, negative or null effects on yield depending on environmental scenarios. Breeders solved this difficulty by directly selecting for high and stable yield across environments. Retrospective analyses of the genetic progress suggest that, doing so, they selected for constitutive traits. The allelic diversity that governs adaptive traits was probably left largely unexploited because it results in positive or negative effects on yield depending on environmental scenarios. We propose a probabilistic approach that estimates the benefits and risks of vectors of alleles/traits in the most likely environmental scenarios in each region, with current or future climates and different management practices. A combination of phenomics, modeling and genomic prediction allows one to identify which vectors of traits/alleles are most likely to be positive for yield in each region and each field over e.g. 30 years. For that, we use phenotyping platforms in controlled conditions and multi-site field experiments to identify genotype-dependent parameters of models, in particular those driving the responses of grain number to temperature, intercepted light and soil water potential. Those parameters can in turn be predicted based on allelic values for new genotypes. This 'big data' approach potentially allows one to model yield for hundreds of genotypes in hundreds of fields in current and future environmental conditions. It may allow exploiting new sources of allelic diversity for yield in drought-prone regions, by explicitly taking into account the alleles that optimize adaptive responses in expected environmental scenarios for each region.

## **W830: Plant Phenotypes**

### **Integrated Phenomics for Connecting Plant Phenotypes to Genotypes with Real-Time Crop Management**

**Nadia Shako**r, Donald Danforth Plant Science Center, Saint Louis, MO

One of the most exciting scientific opportunities in digital agriculture to be realized is using HTP and sensor technologies to gather and analyze real time plant and environmental data from agronomic settings to identify dynamic QTL that are critical to the crop's growth phenology. A controlled-environment study at the Donald Danforth Plant Science Center identified candidate genomic intervals and genes controlling response to early cold stress. GWAS identified transient quantitative trait loci (QTL) strongly associated with each growth-related trait, permitting an investigation into the genetic basis of cold stress response at different stages of development. The analysis identified a priori and novel candidate genes associated with growth-related traits and the temporal response to cold stress. The critical piece that would unequivocally make these associations relevant to production agriculture is the identification of significant transient QTL in replicated field environments. I will discuss the development of an integrated phenomics platform for agronomic optimization and crop breeding that will aim to 1) leverage existing big agricultural data sets to inform scalable sensors and infield computing resources, benchmarking a comprehensive real-time in-field analytics platform; 2) combine high resolution spatial and temporal environmental data with genomic data, allowing for mapping of phenotype to genotype

correlations with key developmental timepoints; and 3) integrate real-time sensing information into ML and crop models to provide direct decision-making support for plant breeders and farmers.

### **W831: Plant Phenotypes**

#### **Developing New Methods to Measure Traits Impacting Soybean Shoot Architecture**

**Suma Sreekanta**<sup>1</sup>, Riley C W O'Neill<sup>2</sup>, Austin Dobbels<sup>3</sup>, Jacob Meierhofer<sup>4</sup>, Fumiaki Katagiri<sup>5</sup>, Jeff Calder<sup>6</sup>, Peter J Oliver<sup>6</sup>, Chehrzad Shakiban<sup>7</sup>, Aaron J Lorenz<sup>3</sup>, Robert M Stupar<sup>3</sup> and Gary J Muehlbauer<sup>3</sup>, (1)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (2)Department of Mathematics, St. Thomas University, (3)Department of Agronomy & Plant Genetics, University of Minnesota, St. Paul, MN, (4)College of Biological Sciences, University of Minnesota, St. Paul, MN, (5)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (6)School of Mathematics, University of Minnesota, Minneapolis, MN, (7)Department of Mathematics, St Thomas, St Paul, MN

Shoot Architecture (SA) is a result of complex interplay between many traits. In crops such as maize and wheat, altering SA has resulted in enhanced yield. However, study of SA has been limited to a few traits because measuring SA traits has traditionally been a slow, low throughput process. Many previous studies have relied on manual measurements of a few traits that can be error prone as well as susceptible to measurement biases. We used a combination of high-throughput technologies including an unmanned aircraft system as well as inexpensive smartphone images to parameterize SA in terms of multiple individual leaf, branch and whole plant traits of field grown plants. A panel of 40 genotypes with strong visual variation in SA were used to evaluate our methods. Canopy coverage was measured each week using a drone based approach. SA traits were captured on images by both destructive and non-destructive sampling of plants. We have developed protocols on a MATLAB platform to automate image analysis. We automated processes to recognize and measure angle of branching, petiole angle, leaf shape characteristics, petiole length as well as overall plant length from images. We standardized tools and techniques to capture images that can be accessed by our image processing pipeline. Our automated image processing shows over 90% co-relation to manual measurements while reducing overall processing time significantly. Our methods for both image capture and image analysis rely on easily accessible equipment and are readily deployed by any lab interested in SA traits.

### **W832: Plant Reproductive Genomics**

#### **Evolution of Parental Genome Dosage Sensitivity during Seed Development – a New Perspective from Water Lilies**

**Rebecca Povilus**, Whitehead Institute, Cambridge, MA, Pamela K. Diggle, University of Connecticut and William E. Friedman, Harvard University

Theoretical and empirical studies have long connected the evolutionary innovation of endosperm, a genetically biparental product of a double fertilization process unique to flowering plants, to conflicting parental interests over offspring provisioning. Yet none of these studies examined interparental conflict in representatives of any of the “early diverging” lineages (such as Amborella, Nymphaeales, Austrobaileyales). We therefore performed reciprocal interploidy crosses in the water lily *Nymphaea thermarum* (Nymphaeales). We find that an excess of paternal genomes is associated with an increase in endosperm growth. By contrast, maternal ploidy negatively influences development or growth of all seed components, regardless of paternal genome dosage. Most relevant to the conflict over distribution of maternal resources, however, is that growth of the perisperm (seed storage tissue derived from the maternal sporophyte, found in all Nymphaeales) is unaffected by paternal genome dosage—ensuring maternal control of maternal resources. We conclude that the evolutionary transfer of embryo-nourishing function from a genetically biparental endosperm to a genetically maternal perisperm can be viewed as an effective maternal strategy to recapture control of resource distribution among progeny. Our results are consistent with the hypothesis that interparental conflict influenced seed development even at the earliest stages of flowering plant evolution.

## **W833: Plant Reproductive Genomics**

### **The Genetic Architecture of Tristyly**

**Haoran Xue**<sup>1</sup>, Stephen I. Wright<sup>1</sup> and Spencer C. H. Barrett<sup>2</sup>, (1)University of Toronto, Toronto, ON, Canada, (2)University of Toronto, Canada

In flowering plants, diverse mechanisms have evolved to limit self-fertilization and/or promote cross-fertilization, one of which is the genetic polymorphism tristily. In tristylous species, populations are comprised of up to three floral morphs with a reciprocal arrangement of stigma and anther heights. This genetic polymorphism is governed by two diallelic, epistatically interacting loci (*S* and *M*). Studies of similar genetic polymorphisms suggest that these loci could be supergenes – clusters of tightly linked genes that facilitate the co-segregation of cooperating alleles. We are investigating the molecular genetic architecture of tristily in *Eichhornia paniculata* using genome assembly, gene mapping, and transcriptomic analysis.

We obtained PacBio and 10X Chromium sequences of *E. paniculata* and assembled a draft genome. Then Hi-C sequences were used to improve the draft assembly, which resulted in a chromosome-level assembly that contains eight large scaffolds, corresponding to the eight pairs of chromosomes in this species. We are verifying the assembly with a linkage map. To identify the genes governing tristily and how they are arranged in the genome, we conducted a genome-wide association study (GWAS) with Illumina sequences of 60 plants (20 of each floral morph) from a single population. We identified several regions that show a complete association with the floral morphs. To narrow down the candidate genes and characterize what components of the tristylous syndrome they control, we are doing an organ-specific transcriptomic analysis with transcriptomes of styles and stamens of each morph at different developmental stages. Collectively, these projects should provide answers to long-standing questions concerning the genetic architecture of tristily and how it evolves.

## **W834: Plant Reproductive Genomics**

### ***Dicer-like 5* Deficiency Confers Temperature-Sensitive Male Sterility in Maize**

**Chong Teng**, Donald Danforth Plant Science Center, St. Louis, MO

Small RNAs play important roles during plant development by regulating transcript levels of target mRNAs, maintaining genome integrity, and reinforcing DNA methylation. *Dicer-like 5* (*Dcl5*) is proposed to be responsible for precise slicing in many monocots to generate diverse 24-nt phased, secondary small interfering RNAs (phasiRNAs), which are exceptionally abundant in meiotic anthers of diverse flowering plants. The importance and functions of these phasiRNAs remain unclear. Here, we characterized several mutants of *dcl5*, including alleles generated by the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) *Cas9* system and a transposon-disrupted allele. We report that *dcl5* mutants have few or no 24-nt phasiRNAs, develop short anthers and defective tapetal cells, and exhibit temperature-sensitive male fertility. We propose that DCL5 and 24-nt phasiRNAs are critical for fertility under growth regimes for optimal yield.

## **W835: Plant Reproductive Genomics**

### **CLE-Family Peptide Signaling Regulates Diverse Floral Morphologies across Plant Taxa**

**Daniel Jones**, Amala John and Zachary L. Nimchuk, University of North Carolina, Chapel Hill, NC

Understanding how conserved signaling pathways can coordinate diverse developmental forms is a central question in biology. Signaling via CLE (CLV3/EMBRYO-SURROUNDING REGION-related) family peptides is a conserved cell-cell communication mechanism regulating stem cell identity, division-plane orientation and organogenesis across highly divergent plant taxa. In this study, we have identified and compared CLE-signaling components that regulate inflorescence development in two model systems that have very distinct inflorescence morphologies, *Arabidopsis thaliana* (*Arabidopsis*) and *Helianthus annuus* (sunflower). Species in the sunflower family (Asteraceae) all have a compact inflorescence known as a capitulum, an evolutionary innovation thought to be crucial for the Asteraceae's

large expansion (~23,000 species) and success (global distribution). It has been hypothesized that the capitulum evolved from a diffuse inflorescence type (a more common morphology among flowering plants), in a process that would require widening of the shoot meristem via stem cell proliferation as well as the suppression of internode elongation between flowers. Through the use of genetics, comparative genomics and transcriptomics; we have identified key differences in CLE signaling components that may account for the development of a capitulum instead of a more diffuse raceme (like in *Arabidopsis*). These results may provide evidence as to how the capitulum first evolved and can lend insight into understanding how CLE signaling affects overall plant morphology.

### **W836: Plant Reproductive Genomics**

#### **Genomic Evidence for Sexual Selection in a Moss**

**Stuart McDaniel**, University of Florida, Gainesville, FL

Sexual selection is widely recognized as a potent force in shaping genetic variation within and promoting divergence among species. In principle, sexual selection should result in different patterns of genetic variation between males and females. For example, loci contributing to sex-biased survival are expected to show elevated  $F_{ST}$  between mature males and females. Integrating sex-biased evolutionary processes across generations, however, requires comparing variation in male-transmitted to female-transmitted genome regions. Here we took advantage of the haploid UV sex chromosome system to test for genomic evidence of long-term sexual selection in the dioecious moss, *Ceratodon purpureus*. The UV sex chromosomes are both sex-specific where the U and homologous V are inherited only through the female and male lines, respectively. Both the U and V sex chromosomes are gene rich, and contain ~500 orthologous gene pairs. Using a global sample of genome re-sequencing data from 12 male-female sibling pairs, we found V-linked loci had lower effective population sizes than U-linked loci, regardless of what mapper we used. To test whether this result could be from a lower mutation rate in males, we compared the rates of protein evolution of U and V-linked genes. We found that although dN/dS ratios were equivalent between males and females, the male dN and dS values were higher than the corresponding female values, indicating that, in contrast, males may have a higher mutation rate than females. These results demonstrate the potential for anisogamy alone to generate sex-biased mutation and selection, thereby shaping variation within and among species.

### **W837: Plant Transgene Genetics**

#### **Simultaneous Gene Editing and Haploid Induction for Direct Editing in Broad Elite Crop Germplasm**

Tim Kelliher, Guozhu Tang, Xiujuan Su, Erin Burch, Weining Gu, Zhongying Chen, Shujie Dong, Yuejin Sun and **Qiudeng Que**, Syngenta, Durham, NC

Genome editing using CRISPR-Cas system works efficiently in plant cells, but delivery of genome-editing machinery into broad elite crop varieties is not feasible or efficient using established in vitro tissue culture-based methods. We co-opted the aberrant reproductive process of haploid induction (HI) to deliver CRISPR-Cas editing machinery for inducing edits in nascent embryos of monocot and dicot species. This new method, named HI-Edit, enables direct genomic modification of elite crop varieties. HI-Edit was tested in field and sweet corn lines using a native haploid-inducer line and extended to dicots using an engineered *CENH3* HI system. We also recovered edited wheat embryos using editing machinery delivered through wide crosses using maize pollen. Our data support the hypothesis that double fertilization precedes uniparental chromosome elimination in HI. Edited haploid plants lack both the haploid-inducer parental DNA and the editing machinery. Therefore, edited plants could be used directly in trait testing and integrated into commercial variety development program.

### **W838: Plant Transgene Genetics**

#### **On Improving Gene Targeting in Plants**

**Yunde Zhao**, Section of Cell and Developmental Biology, University of California, San Diego, San Diego, CA

Precise gene targeting in plants enables allelic replacements, tagging genes in the right genomic context, and many other applications. However, homologous recombination-based gene targeting/replacement (HDR) has been very challenging in plants because of two main obstacles: 1) plant cell wall makes it difficult to deliver repair template DNA into cells; 2) the non-homologous end joining (NHEJ) pathway is the predominant pathway for repairing double-strand DNA breaks (DSB). Intuitively, any increases in efficiency of repair template delivery would probably help HDR. Inhibition or elimination of NHEJ would also increase HDR efficiency by eliminating a major competition. In this presentation, I will present various strategies in facilitating efficient repair delivery into cells. I will also demonstrate that timing the generation of DSB can have a major impact on HDR efficiency. I will present our progress in HDR in both rice and Arabidopsis.

### **W839: Plant Transgene Genetics**

#### **Carbon Nanomaterials Enable Plant Genome Engineering without Transgene Integration**

**Gözde S. Demirer**, Department of Chemical and Biomolecular Engineering, Berkeley, CA

Food security is threatened by decreasing crop yields and increased consumption in the light of climate change, population growth, and a shortage of arable land. To mitigate these factors, genetic engineering of plants can be employed to create crops that have higher yields and nutritional value, and are resistant to biotic and abiotic factors. Despite the recent significant advancements in genome editing, most plant species still remain difficult to genetically engineer. The two bottlenecks of generating transgenic plants are (i) efficient biomolecule delivery into plant cells through the rigid cell wall and (ii) the regeneration of transformed tissues. The workhorse method of plant DNA delivery, *Agrobacterium*, limits the range of plant species that can be transformed and results in uncontrolled transgene integration, hence eliciting a genetically modified organism (GMO) regulatory purview of edited plants. To-date, there has yet to be a plant transformation method that enables high-efficiency plasmid DNA delivery, without transgene integration, in a plant species-independent manner for intact plants.

Here, we describe the generation, validation, and optimization of a nanomaterial-based gene delivery platform that can efficiently deliver genes into both model and agriculturally relevant crop plants, without mechanical aid, in a non-toxic and non-integrating manner; a combination of features that is not attainable with existing plant transformation approaches. More specifically, we chemically modified the surface of single-walled carbon nanotubes (SWNTs) with a cationic polymer (polyethylenimine, PEI) to adsorb the negatively charged genetic cargoes via electrostatic attractions. Delivery of plasmid DNA to tobacco, arugula, cotton, and wheat leaves with these modified SWNTs results in strong transient expression of reporter and functional proteins. We verified that the transgene does not integrate into plant nuclear genome with a highly sensitive digital droplet PCR analysis. Next, we show chemically modified SWNTs can deliver CRISPR plasmids encoding the nuclease protein Cas9 and guide RNAs targeting genes of interest. Through transient expression of Cas9 and guide RNA in plant cells, we obtain stable editing of endogenous plant genes with efficiencies comparable to *Agrobacterium*-mediated delivery in tobacco leaves. Gene editing rates are quantified using a combination of techniques, such as the restriction site loss assay, Sanger sequencing and TIDE/ICE analysis, and deep amplicon sequencing. We also demonstrate the value of our nano-scale gene delivery platform by editing genes in plant seeds with CRISPR-SWNT delivery, which could eliminate the need for laborious tissue culture protocols to regenerate edited plants.

Plant genome editing using carbon nanotubes loaded with CRISPR vectors is a breakthrough advancement addressing two crucial bottlenecks in the generation of transgenic plants: (i) gene delivery to enable transient expression without gene integration and (ii) production of stable gene editing without tissue regeneration. Additionally, non-integrating DNA delivery nanotechnologies could enable genetic engineering of plants in a manner that circumvents GMO labeling in the U.S. to facilitate introduction of engineered crops to the market in a time- and cost-effective manner.

## **W840: Plant Transgene Genetics**

### **A Simple Method for Spray-on Gene Editing *in planta*.**

**Katie Higginbottom**, University of Bristol, Bristol, United Kingdom

Potential innovation in Plant research using gene-edited and genetically modified plants is currently being hindered by inefficient and costly methods of plant transformation. Additionally, many species are very recalcitrant to transformation by the two current methods used, *Agrobacterium tumefaciens* mediated and particle bombardment. Consequently, they remain undeveloped, with their potential untapped.

Carbon dots (quasi-spherical, <10nm nanoparticles) formed from natural materials can act as a fast vehicle for carrying whole DNA plasmids into mature plant cells, resulting in transient plant transformation with no known negative impact on either developmental or photosynthetic processes. Carbon dots also fluoresce due to their surface functionalisation, so can be tracked without the need for a reporter gene, and additionally do not show cytotoxicity in mammalian cells.

The carbon dot/plasmid DNA nanocomplex can currently be applied to plants via either of two methods, plant tissue dipping, or via foliar (spray on) application, and is markedly simple to perform. This method has been successful in a number of species, from important crops such as wheat, barley and maize, to traditionally transformation recalcitrant crops such as sorghum, and even pitcher plants (*Nepenthes*). In all species we have successfully shown GFP or YFP expression, and in wheat we obtained Cas9 gene editing of the meiotic *SPO11* genes, therefore developing the first spray on Cas9 gene editing system for plants.

This unique method of applying carbon dots conjugated with plasmid DNA via foliar spray opens up this technique to readily be developed into very high throughput systems through the use of sprayers or misters, with great promise for use in both research and industry. This protocol for spray-on plant gene editing that is: simple, inexpensive, fast, non-cytotoxic, innately trackable, transforms *in planta* in mature and young plants, and is applicable to multiple species, creates many opportunities for the future of plant transformation in industry and allowing plant transformation to reach more research labs than ever before.

## **W841: Plant Transgene Genetics**

### **Modification of *Agrobacterium Tumefaciens* for Improving the Transformation Frequency in Crops**

**Satoko Nonaka**, Tsukuba Plant Innovation Research Center, Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan and Hiroshi Ezura, Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

*Agrobacterium tumefaciens* has a unique ability to mediate inter-kingdom DNA transfer. This ability of *A. tumefaciens* has been exploited in fields such as functional analysis of genes and breeding of Generic Modified (GM) crops. However, in some plants, transformation frequency is still low, which creates a technical problem. To overcome the problem, we focused on the transformation process comprising the following three steps: the first step is T-DNA transfer, second is transgenic cell selection, and third is transgenic cell regeneration.

To improve T-DNA transfer frequency, we tried to modify *A. tumefaciens*, focusing on the T-DNA transfer, which is the first step of transformation process. Several studies showed that ethylene, phytohormone, and gamma aminobutyric acid (GABA), a non-proteomic amino acid, are factors negatively affecting plant-*Agrobacterium* interactions. Thus, we hypothesized that the removal of these factors would be effective in improving the T-DNA transfer frequency. Based on this hypothesis, four types of Super-*Agrobacterium* were created and analyzed.

Super-*Agrobacterium* ver. 1: the target was ethylene. We tried to introduce the 1-aminocyclopropane-1-carboxylic acid (ACC) deaminase (*acdS*) gene from *Pseudomonas* into *Agrobacterium*. ACC deaminase cleaves ACC, the ethylene precursor, into ammonia and alpha-keto-butyrate. *Agrobacterium tumefaciens* with the AcdS activity was efficacious in suppressing ethylene evolution from plant tissues during co-cultivation and increasing T-DNA transfer.

Super-*Agrobacterium* ver. 2: a stronger promoter was used for driving the *acdS* gene. In Super-*Agrobacterium* ver. 1, the expression of the *acdS* gene was under the control of the lac promoter, which showed constitutive expression in *A. tumefaciens*. In Super-*Agrobacterium* ver. 2, instead of the lac promoter, the *virD* gene promoter was cloned from *A. tumefaciens* and used.

Super-*Agrobacterium* ver. 3: the target was GABA. GABA is degraded by GABA transaminase (GabT). Some bacteria utilize GabT in plant-bacterial interactions. Indeed, a *Pseudomonas* mutant line with inhibited GabT activity decreased infectivity. This result suggests that the GabT activity enables to accelerate plant-bacterial interaction. Based on this result, we hypothesized that the GabT activity would also accelerate plant-*Agrobacterium* interaction. Thus, we introduced the gene encoding GabT into *A. tumefaciens* (Super-*Agrobacterium* ver. 3).

Super-*Agrobacterium* ver. 4: the targets were ethylene and GABA. The AcdS and GabT activities were introduced simultaneously.

These four Super-*Agrobacterium* increased T-DNA transformation in melon, tomato, wild watermelon, *Withania somnifera*, *Erianthus ravennae*, *Colocynthis citrullus*, and *Solanum torvum*. Super-*Agrobacterium* increased the stable transformation frequency in tomato and grasses. Super-*Agrobacterium* vers. 1, 3, and 4 increased the transformation frequency by approximately 2.5, 2.6, and 3.6-fold higher than that of the original strain, respectively. Super-*Agrobacterium* ver. 4 enabled reduction in the time and labor required for transformation by approximately 72% and is therefore the most effective and powerful tool for plant genetic engineering and functional analysis.

## **W842: Plant Transgene Genetics**

### **Transgene Stacking in Potato using the GAENTRY System**

**James Thomson**<sup>1</sup>, Kent F. McCue<sup>2</sup>, Ethan Gardner<sup>3</sup>, Ron Chan<sup>3</sup> and Roger Thilmony<sup>1</sup>, (1)USDA-Agriculture Research Service, Albany, CA, (2)USDA-Agricultural Research Service, Albany, CA, (3)USDA-Agriculture Research Service

This research presents a novel plant biotechnology for the rapid genetic improvement of crops. Although single genes have been important in engineering herbicide and pest tolerance traits, future improvements of complex traits like yield and nutritional quality will likely require the introduction of multiple genes. The GAENTRY (Gene Assembly in Agrobacterium by Nucleic acid Transfer using Recombinase technology) system as a flexible and effective system for stably stacking multiple genes within an *Agrobacterium* virulence plasmid Transfer-DNA (T-DNA). The system utilizes unidirectional site-specific recombinases *in vivo* and an alternating selection scheme to sequentially assemble multiple genes into a single transformation construct. To demonstrate GAENTRY's capabilities, 10 cargo sequences were sequentially stacked together to produce a 28.5 kilobase pair T-DNA, which was used to generate transgenic potato. Approximately 89% of the events identified using the dual antibiotic selection screen exhibited all of the introduced traits. A total of 57% of the tested lines carried a single copy of the selection marker transgene located near the T-DNA left border and none of the plant tested contained sequence from outside the T-DNA. These results demonstrate that the GAENTRY is a powerful, yet simple to use, new tool for transgene stacking, plant synthetic biology and the generation of high quality genetically engineered plants.

## **W843: Polyploidy**

## **Novel Approaches in Phylogenomics Reveal Extensive Interspecific Hybridization and Polyploidization in the Genus *Phytophthora***

**Tom Ruttink**, ILVO Plant Sciences, Melle, Belgium

The genus *Phytophthora* comprises many economically and ecologically important plant pathogens. Interspecific hybrid species have previously been identified in six of the 12 known phylogenetic clades. Hybrid species can potentially infect a wider host range and display enhanced vigour compared to their progenitors. *Phytophthora* hybrids therefore pose a serious threat to agriculture and natural ecosystems and early and correct identification of hybrids is essential for adequate plant protection. Here, we report a novel phylogenomic approach combining Genotyping-by-Sequencing (GBS) and genome size estimation. The method delineates species, identifies interspecific hybrids based on deconvolution of genomic fingerprints, discriminates between 'pure' species, diploid hybrids, allo-polyploid and auto-polyploid chromosomal constitutions, and identifies the most likely progenitor species of various types of hybrids. Screening of a genus-wide collection of 614 *Phytophthora* isolates confirmed and characterized 27 known hybrid species and discovered 16 new hybrid species across the genus. We used both a concatenation- and a coalescent-based phylogenomic method to construct a reliable phylogeny using the GBS data of 140 non-hybrid *Phytophthora* isolates. Hybrid polyploid species were subsequently linked to their progenitors in this phylogenetic tree via bi- or trifurcate links. Thus, our phylogenomic analysis revealed that Clade I contains only diploid hybrids, Clade 6 contains both diploid and polyploid hybrids, and Clade 7 contains hybrids that are almost exclusively derived from at least three progenitor species. Our study paves the way for relatively low cost but high resolution identification of polyploid hybrids with complex chromosomal composition and their phylogenetic relations in many other phyla.

### **W844: Polyploidy**

#### **Genome Stabilization after Allopolyploidization in Nascent *Brassica napus***

**Julie Ferreira-De-Carvalho**, Solenn Stoeckel, Frédérique Eber, Maryse Taburel-Lodé, Marie-Madeleine Gilet, Gwenn Trotoux, Jérôme Morice, Cyril Falentin, Anne-Marie Chèvre and Mathieu Rousseau-Gueutin, INRA IGEPP, Le Rheu, France

Allopolyploidy, which results from the merger and duplication of two divergent genomes, has played a major role in the evolution of plant genomes and is thought to be at the origin of the astonishing plant biodiversity. Although established polyploids have been shown to have higher fitness, meiotic instability and lower fertility have been observed in some resynthesized polyploids. Thus, complex genome stabilization processes must occur during the formation of an allopolyploid species. Due to the similarity of homoeolog chromosomes in allopolyploids, multivalent and univalent chromosomes can be observed at meiosis leading to unbalanced gametes. To avoid this fate and produce viable offspring, polyploid species may either generate divergence between constitutive homoeologous genomes or carry a genetic control preventing non-homoeologous pairing. In that regard, impacts of non-reciprocal translocations via homoeologous rearrangements on genome stability and fertility have still to be established.

Using the *Brassica* model, we investigated the role of these rearrangements on meiotic behavior and fertility in the first generations following independent allopolyploidization events with different resynthesized lineages of *B. napus* and explored the role of several loci associated with bivalent formation.

We showed that the genetic background of the resynthesized polyploids explains part of the variation observed in seed number, presence of univalent during meiosis and the capacity of producing offspring. We also identified that the genetic background of the resynthesized lineages has an impact on the position, size and number of non-reciprocal translocations observed in resynthesized polyploids. Finally, we developed a statistical approach to identify absent SNPs significantly associated with bivalent formation and explore associated translocations. This fine genetic analysis of the underlying molecular processes provides important results on the mechanisms involved in genome stability and increased fertility in a nascent allopolyploid.

## **W845: Polyploidy**

### **Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms**

**Yuannian Jiao**, Institute of Botany, The Chinese Academy of Sciences, Beijing, GA, China

Ancient whole-genome duplications (WGD or polyploidy) are prevalent in plants, and some WGDs occurred during the timing of severe global environmental changes. It has been suggested that WGDs may have contributed to plant adaptation. However, it still lacks of empirical evidence from genetic level to support the hypothesis. Here, we investigated the survivors of gene duplicates from multiple ancient WGD events on the major branches of angiosperm phylogeny, and aimed to explore genetic evidence supporting the significance of polyploidy. Duplicated genes co-retained from three waves of independent WGDs (~120 million years ago (Ma), ~66 Ma and <20 Ma) were investigated in 25 selected species. Gene families functioning in low temperature and darkness were commonly retained gene duplicates after the eight independently occurred WGDs in many lineages around the Cretaceous–Paleocene (K-Pg) boundary, when the global cooling and darkness were the two main stresses. Moreover, the commonly retained duplicates could be key factors which may have contributed to the robustness of the critical stress related pathways. In addition, genome-wide transcription factors (TFs) functioning in stresses tend to retain duplicates after waves of WGDs, and the co-selected gene duplicates in many lineages may play critical roles during severe environmental stresses. Finally, our results shed new light on the significant contribution of paleopolyploidy to plant adaptation during global environmental changes in the evolutionary history of angiosperms.

## **W846: Polyploidy**

### **Ecological Epigenetics in Polyploid *Spartina***

**Christina L. Richards**, University of South Florida, Institute of Evolution and Ecology, FL, Germany  
ecological epigenetics in various models including polyploid *Spartina*.

## **W847: Polyploidy**

### **Origin and Evolution of the Octoploid Strawberry Genome**

**Patrick Edger**, Department of Horticulture, Michigan State University, East Lansing, MI

Cultivated strawberry emerged from the hybridization of two wild octoploid species, both descendants from the merger of four diploid progenitor species into a single nucleus more than 1 million years ago. Here we report a near-complete chromosome-scale assembly for cultivated octoploid strawberry (*Fragaria × ananassa*) and uncovered the origin and evolutionary processes that shaped this complex allopolyploid. We identified the extant relatives of each diploid progenitor species and provide support for the North American origin of octoploid strawberry. We examined the dynamics among the four subgenomes in octoploid strawberry and uncovered the presence of a single dominant subgenome with significantly greater gene content, gene expression abundance, and biased exchanges between homoeologous chromosomes, as compared with the other subgenomes. Pathway analysis showed that certain metabolomic and disease-resistance traits are largely controlled by the dominant subgenome. These findings and the reference genome should serve as a powerful platform for future evolutionary studies and enable molecular breeding in strawberry.

## **W848: Polyploidy**

### **Implications of Whole-Genome Duplications on Dioecy and Sex Chromosomes in Mosses**

**Sarah Carey**, University of Florida, Gainesville, FL

The mating system is one of the key factors shaping patterns of variation within species and potentially contributing to long-term macroevolutionary patterns. Whole-genome duplications (WGDs) may promote the evolution of dioecy because duplicate genes can specialize for sex-specific function or WGDs may cause transitions from dioecy to hermaphroditism because male and female sex-determining factors are

reunited within an individual. Here we examine the link between WGDs and mating system in the mosses, a group in which ~60% of the 13,000 species are dioecious but many independent transitions between sexual system have occurred. We first generated a chromosome-scale assembly of the genome of the dioecious moss *Ceratodon purpureus*. To find evidence of WGD, we used Ks-based and synteny-based analyses. Using Ks, we found a WGD in *C. purpureus*, independent of the two in the hermaphroditic moss, *Physcomitrella patens*. This WGD is more recent than the evolution of the sex chromosomes in mosses, which arose before the common ancestor of *C. purpureus* and *P. patens*, providing genomic evidence that WGDs do not necessarily lead to a transition to hermaphroditism. In self-synteny analyses, we found evidence that two chromosomes in *C. purpureus* lack collinear copies. Further phylogenomic analyses indicate these “missing” chromosomes have fused onto the *C. purpureus* sex chromosomes. These results demonstrate the important role WGD can play in the ongoing evolution of sex chromosomes and the potential to influence mating system.

## **W849: Population and Conservation Genomics 1**

### **Population and Conservation Genomics: Challenges and Opportunities**

**Om P. Rajora**, University of New Brunswick, Fredericton, NB, Canada

Advances in genomics technologies and bioinformatics tools have revolutionized several disciplines of biology, including population, evolutionary and conservation genetics. Population and conservation genomics employ large-scale, genome-wide and/or transcriptome-wide genetic information and novel bioinformatics and statistical methods to address various fundamental and applied questions in biology. Population and conservation genomics provide unprecedented opportunities to address novel and long-standing intractable questions with unprecedented power and accuracy in various disciplines of biology, especially population, evolutionary, ecological and conservation genetics, plant and animal breeding, and conservation and genetic management of planet’s bioresources. However, there are still several challenges in realizing the full potential of population and conservation genomics across the disciplines and organisms. I will briefly outline the challenges and opportunities of population and conservation genomics research and applications in addressing fundamental and applied questions in various disciplines of biology.

## **W850: Population and Conservation Genomics 1**

### **Detecting Genomic Variation Underlying Phenotypic Characteristics of Reintroduced Coho Salmon (*Oncorhynchus kisutch*)**

**Rebekah L. Horn**, Columbia River Inter-Tribal Fish Commission, Hagerman, ID

For species that have been extirpated from parts of their range, conservation managers often reintroduce individuals to these areas in hopes of restoring populations to pre-decline conditions. Coho salmon (*Oncorhynchus kisutch*) were listed as extirpated in the mid-1990s in the interior reaches of the Columbia River watershed. Starting in the late 1990s, the Columbia River Treaty tribes were successful in starting a re-introduction program that has established a Mid-Columbia River Coho salmon stock. Fish are preferentially selected for broodstock to incorporate naturally occurring phenotypic characteristics to facilitate local adaptation. On the Wenatchee River in Washington, broodstock are preferentially selected at a lower and upper river dam, however, only ~32% of fish successfully ascend a 15km high-gradient reach to the upper river dam. Fish that successfully ascend the reach generally arrive early in the season and have a better overall body condition. In other salmonids, phenotypic traits such as return timing has been shown to be under genetic control. To determine if there are genomic regions that underly the phenotypic traits found to impact migration success up a high-gradient reach, low-coverage whole genome re-sequencing (lcWGR) was performed on adult fish returning to the system. Genome-wide association tests revealed three genomic regions that are associated with fish return location. Results of the lcWGR can be incorporated as a genetic screening during broodstock selection to preferentially breed fish that have the phenotypic characteristics that confer greater potential for steeper and longer migration distances.

## **W851: Population and Conservation Genomics 1**

### **Massive Haplotypes Underlie Ecotypic Differentiation in Sunflowers**

**Marco Todesco**<sup>1</sup>, Gregory L. Owens<sup>2</sup>, Natalia Bercovich<sup>1</sup>, Jean-Sébastien Légaré<sup>1</sup>, Shaghayegh Souidi<sup>3</sup>, Dylan O. Burge<sup>1</sup>, Kaichi Huang<sup>1</sup>, Katherine Ostevik<sup>4</sup>, Emily Drummond<sup>1</sup>, Ivana Imerovski<sup>1</sup>, Kathryn Lande<sup>1</sup>, Mariana Pascual<sup>1</sup>, Winnie Cheung<sup>1</sup>, S. Evan Staton<sup>1</sup>, Stéphane Muñoz<sup>5</sup>, Rasmus Nielsen<sup>6</sup>, Lisa A. Donovan<sup>7</sup>, John M. Burke<sup>7</sup>, Sam Yeaman<sup>8</sup> and Loren H. Rieseberg<sup>1</sup>, (1)University of British Columbia, Department of Botany and Biodiversity Research Centre, Vancouver, BC, Canada, (2)University of California, Berkeley, Berkeley, CA, (3)University of Calgary, Department of Integrative Biology, Calgary, AB, Canada, (4)Duke University, Department of Biology, Durham, NC, (5)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRAE/CNRS/Université de Toulouse, Castanet-Tolosan, France, (6)Department of Integrative Biology and Statistics, Berkeley, CA, (7)University of Georgia, Athens, GA, (8)University of Calgary, Calgary, AB, Canada

Species experiencing different environments across their range often include multiple ecotypes, with locally adaptive morphological and/or physiological characteristics. This typically requires the concerted effect of alleles at multiple loci. Much remains to be understood about how ecotypes are established and maintained; in particular, a longstanding question in evolutionary biology concerns how these adaptive allelic combinations are kept together in the face of hybridization with neighbouring, non-adapted populations.

To better understand the genetic basis of local adaptation and ecotypic differentiation, we re-sequenced 1506 wild sunflowers from three species that include several well-characterized ecotypes. We identified 37 large (1-100 Mbp), non-recombining haplotype blocks (haploblocks) associated with numerous ecologically relevant traits, and soil and climate characteristics. Limited recombination in these regions allows adaptive alleles to be inherited together, and we find that haploblocks differentiate several sunflower ecotypes. For example, in *Helianthus argophyllus* (silverleaf sunflower) a 30 Mbp haplotype introgressed from its sister species *H. annuus* controls a 77 day difference in flowering between coastal and inland ecotypes, likely through deletion of a *FLOWERING LOCUS T* homolog; several other haploblocks are associated with seed size, flowering time and soil fertility in dune-adapted ecotypes of *H. petiolaris*.

These haploblocks are generally associated with polymorphic structural variants, which provide a straightforward mechanism for suppressing recombination between haplotypes, and they are highly divergent, often appearing to represent introgressions from other, possibly extinct, congeners. These results highlight a pervasive role of structural variation and introgression in maintaining complex ecotypic adaptation

## **W852: Population and Conservation Genomics 1**

### **Population Genomics and Improved Reference Genomes of African Rice (*Oryza glaberrima*) and its Wild Ancestor allow Comparisons of 44 Characterized Domestication Genes to Demystify Longstanding Assumptions of Parallel Evolution**

**Rachel Meyer**<sup>1</sup>, Briana L. Gross<sup>2</sup>, Jing Ning<sup>3</sup>, Min Hu<sup>3</sup>, Jae Young Choi<sup>4</sup>, LeQin Chang<sup>3</sup>, Michael D. Purugganan<sup>5</sup> and Zuofeng Zhu<sup>6</sup>, (1)University of California Santa Cruz, Santa Cruz, CA, (2)University of Minnesota Duluth, Duluth, MN, (3)China Agricultural University, China, (4)New York University, (5)New York University, New York, NY, (6)China Agricultural University, Beijing, China

Sub-Saharan Africa has its own domesticated rice species that previous work demonstrated was gradually domesticated across multiple areas of West Africa. A population bottleneck was found to start in the early Holocene, just as in Asian rice. Reference genome assemblies for *Oryza* species and whole genome resequencing of numerous cultivars and wild accessions have enabled researchers to test for parallel evolutionary trajectories between African rice and Asian rice. These rice species provide an ideal testable framework for theoretically advancing Nikolai Vavilov's Law of Homologous Series, a pillar of genetics theory. However, reference genome quality needed to be improved in African wild and

domesticated rice species before functional mutations in orthologous domestication and improvement genes could be readily compared. Here we present new reference genomes for *Oryza glaberrima* and *O. barthii* and their structural differences from Asian rice genomes for *O. sativa* and *O. rufipogon*. We then used these genomes to assess the extent of parallel functional mutations affecting 44 genes characterized in Asian rice to have been part of domestication and improvement phenotypic changes. Analyses were performed on whole genome sequences of 267 *O. glaberrima* landraces and 132 *O. barthii* accessions. We found that despite published reports of major traits exhibiting underlying parallel orthologous gene loss of functions or deletions, that few genes exhibit functional mutations in African rice despite shared phenotypes with Asian rice, and that functional mutations are often part of the standing variation in the wild. This work provides insight into the overlap of both natural and artificial selection and the Law of Homologous Series.

## **W853: Population and Conservation Genomics 1**

### **Inferring Plant Genome Evolution of an Invasive Species (*Ambrosia artemisiifolia* L.) using Historic Herbarium Specimens**

**Vanessa Carina Bieker**, Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, Kathryn A. Hodgins, Monash University, Clayton, Australia and Michael D. Martin, Department of Natural History, NTNU University Museum, Norwegian University of Science, Trondheim, Norway

A vast number of plant specimens from various taxa, times and locations are preserved in herbaria. Due to improvements in DNA extraction and sequencing technologies, these specimens have become readily available for genetic studies. The study of invasive plants offers an excellent opportunity to answer fundamental evolutionary questions about whether parallel adaptation to similar introduced environments affects the same genomic regions. The diploid annual weed *Ambrosia artemisiifolia* (common ragweed) is native to North America and became invasive upon introduction in Europe during the late 19<sup>th</sup> century. Its highly allergic pollen is the main cause of hay fever, and the plant produces high annual costs (control, health care, reduced crop production). To infer temporal changes in population structure and gene-linked allele frequencies at specific geographic locations, we resequenced more than 300 modern and 400 historic herbarium specimens collected up to 180 years ago from both the native and the introduced ranges. We map low-depth shotgun sequencing data (mean coverage=1) against a 1.4-Gbp reference genome assembly, infer genotype likelihoods using ANGSD, and population genetic structure using NGSadmix. To identify genomic regions putatively under selection, we perform an outlier analysis based on the genomic distribution of Wright's fixation index  $F_{ST}$  in sliding windows. We expect to find repeated patterns of local adaptation and a higher frequency of adaptation from standing variation (as opposed to *de novo* mutations) in the introduced range. We will test different introduction scenarios in order to identify the source population(s) of European common ragweed. Preliminary results will be reported.

## **W854: Population and Conservation Genomics 1**

### **Local Adaptation in *Populus trichocarpa* Torr. & Gray**

**Hari Chhetri**, West Virginia University, Morgantown, WV

Because of its rapid growth, hybrid vigor, broad geographic distribution, transformation potential, and the availability of tremendous genetic resources and wide phenotypic variation, *Populus* is a highly desirable species for biofuel production and other wood products. Understanding the genetic mechanisms underlying local adaptation is key for the sustainable management and domestication of forest trees like *Populus*. Here we report on the possible mechanisms underlying local adaptation in *Populus trichocarpa* using whole genome re-sequencing, phenotypic and geo-climate data for 869 trees. First, we show that morphological and physiological traits are strongly correlated with the geo-climate variables of the source locations in *P. trichocarpa*. Second, using Genotype-Environment Association Analysis (GEA) and Redundancy Analysis (RDA) we identified several outlier loci that occur within and near genes related to important plant physiological functions and cuticular wax formation. A total of 32 genes were shared

between RDA and GEA methods. Third, using RDA, we decomposed the among population variance of 869 trees into climate and geography. While climate and geography predictors together explained 7.6% of the total variation in the SNP matrix, climate alone explained 2.6% of the total variation. Partitioning the variance components in the response matrix of phenotypic traits into the explanatory matrices of SNPs, climate and geography, explanatory matrices altogether explained 22% of the total variation, whereas SNPs alone explained 3.9% of the total variation. These findings have important implications for developing management and conservation strategies and sustainability of forest resources in the face of climate change.

## **W855: Population and Conservation Genomics 1**

### **Population Genetics and Hybridization of Mallards and Mallard-like Ducks in North America**

**Philip Lavretsky**, University of Texas at El Paso, El Paso, TX

I present genomic data used to characterize population structure among five recently diverged New World mallard-like ducks, estimate gene flow, as well as identify signatures of selection across their genomes. In contrast to previous studies, coupling landscape-level sampling efforts and the thousands of ddRAD-seq loci successfully assigned individuals to their respective taxon or hybrid class. First, I find limited evidence of contemporary gene flow, but find support for ancient gene flow between some of the species pairs. Among the species, I report that the most complicated genetic relationship between mallards and American black ducks is due to ancestry and not recent gene flow, and that American black ducks are not on their path to becoming a “hybrid swarm.” Thus, despite recurring cases of hybridization in this group, I conclude that the overall genetic similarity reflects retained ancestral variation and not gene flow. In fact, I report previously unknown outlier regions across the Z-chromosome and several autosomal chromosomes that may be important in the diversification of this group. Conversely, I provide evidence from century-old (1842-1915) and contemporary (>2009) mallard comparisons, and confirm that intensive stocking practices of domestic game-farm mallards conducted across the last century has fundamentally changed the genetic integrity of North America’s wild mallard population, including the establishment of a feral x wild mallard hybrid swarm in eastern North America. It becomes of great interest to ask whether the iconic North American mallard is declining in the wild due to introgression of maladaptive traits from these domesticated forms.

## **W856: Population and Conservation Genomics 2**

### **Introduction**

**Om P. Rajora**, University of New Brunswick, Fredericton, NB, Canada

## **W857: Population and Conservation Genomics 2**

### **Genomic Impact of Founder History on Populations of the Critically Endangered Dama Gazelle (*Nanger dama*)**

**Klaus-Peter Koepfli**, Center for Species Survival, Smithsonian's National Zoological Park and Conservation Biology Institute, Washington, DC

Dama gazelles are the world’s largest and rarest gazelle species, with only about 200 animals remaining in the wild. They are native to the Sahara Desert and Sahel, with only remnant populations remaining in Chad, Mali, and Niger. However, more than 2,300 dama gazelles are managed *ex situ* in zoos and private collections around the world, with the largest number of animals found on private ranches in North America, mostly in Texas. Three geographic subspecies have been recognized based on color patterning differences among populations: addra (*Nanger dama ruficollis*), the nominate dama (*N. dama dama*), and mhorr (*N. dama mhorr*). The *ex situ* population of mhorr gazelles was founded by only a small number of individuals, before this subspecies had become extinct in the wild. In contrast, the addra gazelle *ex situ* population was founded by a larger group of individuals. We examined the genome-wide effects of these different founding histories by generating whole genome sequences of addra and mhorr gazelles, which included a chromosome-scale reference genome assembly from one addra gazelle.

Mhorr gazelles had almost 50% less heterozygosity, a genome occupied by up to 45% of runs of homozygosity, and about three times the number of putatively loss-of-function mutations compared to addra gazelles. We also analyzed addra gazelle populations managed in zoos and private ranches in North America using SNPs identified with ddRAD and found differences in levels of inbreeding and admixture among individuals. These genomic data help inform the conservation management, genetic rescue, and reintroduction back into the wild of this critically endangered antelope.

#### **W858: Population and Conservation Genomics 2**

##### **Mountain Lion Genomes Provide Insights into the Genomic Consequences of Inbreeding and Genetic Rescue**

**Megan A. Supple**<sup>1</sup>, Nedda F Saremi<sup>2</sup>, Richard E. Green<sup>2</sup> and Beth Shapiro<sup>3</sup>, (1)University of California, Santa Cruz, Santa Cruz,, CA, (2)University of California, Santa Cruz, Santa Cruz, CA, (3)Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA

Mountain lion populations across much of North and South America have become increasingly isolated due to human persecution and habitat loss. To explore the genomic consequences of this isolation, we analyzed whole genomes and found signatures of close inbreeding in isolated North American mountain lion populations. However, tracts of homozygosity were rarely shared among mountain lion populations, suggesting that assisted gene flow would restore local genetic diversity. The genome of a Florida panther from a recently admixed population had long tracts of homozygosity, suggesting that although translocations may introduce diversity into a population, sustaining diversity in small and isolated populations will require either repeated translocations or restoration of landscape connectivity. Mountain lions provide an opportunity to examine the potential to restore diversity through genetic rescue and to observe the effects of translocations, informing the management of small and isolated populations.

#### **W859: Population and Conservation Genomics 2**

##### **What We Can Learn from Wild Emmer Wheat in Response to 28 Years of Global Warming**

**Yong-Bi Fu**, Plant Gene Resources of Canada, Saskatoon, SK, Canada

The realized threats of global warming to biodiversity have catalyzed the search for solution to protect and conserve extant plant genetic resources. Part of the solution, however, is dependent on the knowledge of how plant populations respond genetically to these threats, which is largely lacking. This presentation will highlight the major findings from our unique genomic investigation to characterize genetic responses in 10 wild emmer wheat populations in Israel that were sampled twice in 1980 and 2008. After the 28 years of global warming, these populations displayed elevated selection, reduced diversity and temporal divergence, and carried increased mutational burdens forward. However, some populations still showed the ability to acquire beneficial alleles for future adaptation. The patterns of genetic response to rainfall and temperature varied and were complex.

#### **W860: Population and Conservation Genomics 2**

##### **Modelling Local Adaptation and Migration Potential Predict the Vulnerability of Populations of Wild Maize to Climate Change**

**Jonas Aguirre**, University of California, Irvine, Irvine, CA

Climate change is an important threat to biodiversity. Populations can respond to climate change by adapting locally, migrating to suitable areas or by going extinct. Unfortunately, the majority of studies evaluate the impact of climate change based on niche modelling, omitting the importance of local adaptation and migration potential. In our study, we analyzed how populations of two wild relatives of maize, the teosintes *Zea mays* ssp. *mexicana* and *Z. mays* ssp. *parviglumis*, will respond to climate change. First we used gradient forest to model adaptive and neutral allelic frequencies across the teosintes landscape. These models were used to project allelic changes in 8 future scenarios of climate change. Based on predicted allelic changes we defined the vulnerability of populations. Second we used

niche modelling to predict the current and future distribution of adaptive alleles. Third, we used these allelic models and circuit theory to predict the probability of teosinte's population to migrate to new areas. We found that based on local adaptation information, teosinte populations will be vulnerable to climate change and only few populations will be able to migrate by 2070. This contrast with previous niche modelling that found that climate change would affect around 30% of teosintes populations. Since teosintes and maize are genetically close, we searched adaptive SNPs in maize and found that our set of outlier SNPs predict the temperature at which maize landraces grow, suggesting that our approximation might help mitigate the impact of climate change in this important crop.

## **W861: Population and Conservation Genomics 2**

### **Genomics of Sorghum Local Adaptation to a Parasitic Plant**

Emily S Bellis<sup>1</sup>, Elizabeth Kelly<sup>1</sup>, Claire Lorts<sup>1</sup>, Huirong Gao<sup>2</sup>, Victoria Deleo<sup>1</sup>, Germinal Rouhan<sup>3</sup>, Andrew Budden<sup>4</sup>, Govinal Bhaskara<sup>5</sup>, Zhenbin Hu<sup>6</sup>, Robert Muscarella<sup>7</sup>, Michael P. Timko<sup>8</sup>, Baloua Nebie<sup>9</sup>, Steven Maina Runo<sup>10</sup>, Doane Chilcoat<sup>11</sup>, Thomas Juenger<sup>12</sup>, Geoffrey P. Morris<sup>13</sup>, Claude W. dePamphilis<sup>1</sup> and **Jesse Lasky**<sup>14</sup>, (1)The Pennsylvania State University, University Park, PA, (2)Corteva Agriscience, (3)Muséum national d'Histoire naturelle, France, (4)Royal Botanic Gardens, Kew, United Kingdom, (5)University of Texas at Austin, (6)Department of Agronomy, Kansas State University, Manhattan, KS, (7)Uppsala University, Sweden, (8)University of Virginia, Charlottesville, VA, (9)ICRISAT, Mali, (10)Kenyatta University, Nairobi, Kenya, (11)Corteva Agriscience, Johnston, IA, (12)University of Texas at Austin, Austin, TX, (13)Kansas State University, Manhattan, KS, (14)Pennsylvania State University, University Park, PA

Host-parasite coevolution can maintain high levels of genetic diversity in traits involved in species interactions. In many systems, host traits exploited by parasites are constrained by use in other functions, leading to complex selective pressures across space and time. Here, we study genome-wide variation in the staple crop *Sorghum bicolor* (L.) Moench and its association with the parasitic weed *Striga hermonthica* (Delile) Benth., a major constraint to food security in Africa. We evaluate the hypothesis that sorghum landraces are subject to geographic selection mosaics for resistance to parasites by analyzing variation at sorghum *LOW GERMINATION STIMULANT 1 (LGS1)*, a gene thought to impact resistance. *LGS1* influences stereochemistry of strigolactones, hormones that control plant architecture and belowground signaling to mycorrhizae and are required to stimulate parasite germination. Multiple independent *LGS1* loss-of-function alleles are broadly distributed among African landraces and geographically associated with *S. hermonthica* occurrence, suggesting a role in local adaptation to parasite pressure. However, low frequency of these alleles within *S. hermonthica*-prone regions and their absence elsewhere implicate potential tradeoffs restricting their fixation. Consistent with tradeoffs, we find signatures of balancing selection surrounding *LGS1* and other candidates from analysis of genome-wide associations with parasite distribution. Experiments with CRISPR-Cas9 edited sorghum indicate the benefit of *LGS1*-mediated resistance strongly depends on parasite genotype and abiotic environment and comes at the cost of reduced photosystem gene expression. Our study demonstrates long-term maintenance of diversity in host resistance genes across smallholder agroecosystems, providing a valuable comparison to both industrial farming systems and natural communities.

## **W862: Population and Conservation Genomics 2**

### **Leveraging Genotype-by-Environment Interactions across Discrete Environmental Regions to Select More Sustainable Beef Cattle**

**Harly J. Durbin**<sup>1</sup>, Troy N. Rowan<sup>1</sup>, Jared E. Decker<sup>1</sup> and Stephen P. Miller<sup>2</sup>, (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Angus Genetics Inc., Saint Joseph, MO

Compared to other livestock systems, beef cattle are produced in a wide range of climates with minimal management interventions. While environmental heterogeneity is a strength of the beef cattle industry, it also increases the potential for genotype-by-environment interactions. This may present issues for

sustainable beef production, especially for breeders in stressful environments purchasing genetics from outside of their region.

Random regression models using a single environmental variable may not capture all stressors unique to a region (i.e., local pathogens and forages) and their interactions. Therefore, we generated environmental region-specific genetic predictions. Using k-means clustering on 30-year-normals for temperature, rainfall, and elevation, we assigned growth trait records from cattle registered in the American Angus Association to one of 7 discrete climate regions based on herd zip code.

Data was collected by American Angus breeders from 1990-2018 and included 6,310,534 weaning weight and 3,189,162 post-weaning gain measurements. We then calculated breeding values and variance components from univariate (single region) & multivariate (multiple regions) restricted maximum likelihood models. We find that while genetic trends are similar between regions, there is significant re-ranking of breeding values. Genetic correlations between regions ranged from 0.77 to 0.93 for weaning weight direct, 0.66 to 0.95 for the maternal component of weaning weight, and 0.63 to 0.91 for post-weaning gain. Using single-step BLUP combining genotyped and non-genotyped animals, we calculate SNP effects and find potentially adaptive variants contributing to re-ranking between regions.

In the future, these approaches could be incorporated into genetic evaluations to help beef cattle producers identify animals better suited to their environment.

### **W863: Potato Genomics**

#### **Genome Sequencing of 12 Taxonomically Diverse Potato Species of Varying Ploidy**

Maria Kyriakidou<sup>1</sup>, Noelle Anglin<sup>2</sup>, Helen H. Tai<sup>3</sup>, David Ellis<sup>2</sup> and **Martina Stromvik**<sup>4</sup>, (1)Department of Plant Science - McGill University, Montreal, QC, Canada, (2)International Potato Center, Lima, Peru, (3)Agriculture and Agri-Food Canada, Fredericton, NB, Canada, (4)Department of Plant Science - McGill University, Ste. Anne de Bellevue, QC, Canada

While there is an immense depth of genetic diversity in the tuber-producing *Solanum* clade, because their genomes are very complex, much of the potato genomics work to date has been carried out in simple lab clones and inbred diploids. We have sequenced and *de novo* assembled twelve taxonomically diverse potato (*Solanum* sp.) species ranging in ploidy from diploid to pentaploid from the genebank at the International Potato Research Center (CIP) in Lima, Peru. Structural variation and presence-absence analyses were used to identify regions with high variation. These regions proved to contain genes involved in abiotic and biotic stress tolerance. Furthermore, a draft genome of a cultivated diploid potato species (*S. stenotomum* subsp. *goniocalyx*), was achieved using hybrid sequencing and assembly. This was used with the other five diploid genomes in this study and three previously published potato genomes to construct a pan-genome model for diploid potato. This pan-genome includes novel genes, among which are for example those involved in fruit, flower and tuber shape, self-incompatibility and biotic stress resistance, all with great potential for applications to potato breeding programs.

### **W864: Potato Genomics**

#### **Developing Pathways for Practical Deployment of DNA-Based Selection Strategies in Potato Breeding**

**Dan Milbourne**, Fergus Meade, Stephen Byrne, Maria de la O Leyva Perez, Colum Kennedy, Francesca Mesiti and Denis Griffin, Teagasc, Crop Science Department, Carlow, Ireland

Marker Assisted and Genomic Selection (MAS & GS) offer great potential to improve the efficiency of potato breeding, but practical, logistical and economic factors can often hinder their deployment. We describe a series of experiments designed to address some of these constraints. A wealth of information exists on molecular markers for loci involved in resistance to biotic stresses of potato but these markers come in many different forms. Ideally a breeding programme would routinely perform annual genotyping on a single unified platform. To address this, we developed an approach to consolidate various types of

molecular markers to a single SNP-based platform using whole genome sequencing of carefully constructed pools of DNA using, as examples, markers for the *H1* nematode and *R2* blight resistance genes. We show the use of this process to drive the cost effective routine deployment of MAS for disease and pest resistance in our breeding programme. Whilst MAS is useful for dominant disease resistance loci, GS is more suited to polygenic traits. To facilitate GS for processing traits, we accumulated fry-colour data on a large population of lines under selection in our breeding programme and combined this with ~50K SNPs identified with reduced representation sequencing to evaluate accuracy of genomic prediction. We were able to predict fry-colour with moderate accuracy (and to identify a major QTL on chromosome 10). We also demonstrated that it was still possible to achieve moderate (but lower) prediction accuracies for fry colour with hundreds of markers. Based on this, we outline the first steps in the development of an inexpensive, multi-allelic, and adaptable genotyping platform that can be used for simultaneous marker and genomic assisted selection with minimal changes to the current structure of the breeding programme.

## **W865: Potato Genomics**

### **Insect Resistance in Potato**

**Natalie R Kirkwyland**<sup>1</sup>, Christina DiFonzo<sup>2</sup>, Norma Manrique<sup>3</sup>, Joseph Coombs<sup>1</sup> and David Douches<sup>1</sup>, (1)Michigan State University, East Lansing, MI, (2)Michigan State University, (3)CIP, Peru

The diploid potato species *Solanum chacoense* has been used for over three decades to introgress leptine glycoalkaloid mediated Colorado potato beetle (*Leptinotarsa decemlineata*) host plant resistance into cultivated potato (*Solanum tuberosum*). Investigation and exploitation of the complex genetic basis conferring Colorado potato beetle resistance in *S. chacoense* has been obscured by interspecific crosses, low recombination within small populations, and a dearth of marker density to provide sufficient resolution. As a consequence, no cultivars resistant to this devastating insect defoliator pest have been achieved. To combat these challenges, we generated a self-compatible, diploid F<sub>2</sub> mapping population of 223 individuals from the *S. chacoense* lines USDA8330-1 and M6, a self-compatibility donor, segregating for Colorado potato beetle field resistance and accumulation of secondary metabolites implicated in Colorado potato beetle resistance. We employed bi-parental linkage mapping and whole genome bulk segregant analysis in this F<sub>2</sub> population to identify a major QTL on the long arm of chromosome 2 explaining 43.7% and 30.7% of the variance in Colorado potato beetle field resistance and leptine biosynthesis, respectively. Validation of the resistance region on chromosome 2 was accomplished in an additional 406 F<sub>2</sub> progeny using custom InDel markers. Integrated RNAseq and glycoalkaloid profiling of resistant and susceptible individuals point to previously uncharacterized uridine diphosphate-glycosyltransferases and ethylene response transcription factors within the QTL region. Identification of beetle resistant F<sub>2</sub> lines with low leptine content may indicate close linkage between genes responsible for leptine accumulation and another mechanism of resistance on chromosome 2.

## **W866: Potato Genomics**

### **3-R-Gene Potato for Late Blight Resistance**

**Eric M Magembe**, International Potato Center, Nairobi, Kenya

#### **3R Gene Potato for Late Blight Resistance**

Considered responsible for one million deaths in Ireland and widespread famine in the European continent during the 1840s, late blight, caused by *Phytophthora infestans*, remains the most devastating disease of potato (*Solanum tuberosum* L.) with about 15-30% annual yield loss in sub-Saharan Africa, affecting mainly smallholder farmers. We show here that the transfer of three resistance (R) genes from wild relatives [RB, Rpi-blb2 from *Solanum bulbocastanum* and Rpi-vnt1.1 from *S. venturii*] into potato provided complete resistance in the glasshouse and field over several seasons. We observed that the stacking of the three R genes produced a high frequency of transgenic events with resistance to late blight. In the field, 16 resistant transgenic events with the 3R-gene stack from the potato varieties 'Desiree' and 'Victoria' grew normally without showing pathogen damage and without any fungicide

spray, whereas their non-transgenic equivalent varieties were rapidly killed. Yields of two transgenic events from 'Desiree' and 'Victoria' grown without fungicide to reflect small-scale farm holders were estimated to be 29 t/ha and 45 t/ha, respectively. This represents a three to four-fold increase over the national average. T-DNA insertion characterization by next generation sequencing has identified 4 lead transgenic events that are being tested in multi-location confined field trials for regulatory studies. Thus, these late blight resistant potato varieties, which are the farmers' preferred varieties, could be rapidly adopted and bring significant income to smallholder farmers in sub-Saharan Africa.

## **W867: Potato Genomics**

### **The Origin and Widespread Occurrence of *Sli* based Self-Compatibility in Potato**

**Corentin C.R. Clot**, Clara Polzer, Charlotte Prodhomme, Christel J.M. Engelen, Ronald C.B. Hutten and Herman J. van Eck, Plant Breeding, Wageningen University and Research, Wageningen, Netherlands

As self-compatible (SC) diploids have become central to innovative potato breeding, it appears essential to characterize SC sources and to develop predictive markers for this trait. In this study, we used a k-mer based bulk segregant analysis to map SC in two *S. tuberosum* F1 populations. Bulks were created based on concordant berry set and pollen tube growth data upon selfing. Genomic DNA of bulks and parental lines were sequenced and k-mers were generated. Set algebra and depth filtering were used to identify k-mers linked in coupling and repulsion phase with a SC locus. In both population, the SC phenotype mapped to the distal end of chromosome 12, and this was supported both by coupling and repulsion phase k-mers. Intersection between the k-mers in coupling phase with SC in the two populations confirmed the discovery of a shared haplotype of approximately 1.5 Mb. We subsequently screened a collection of read archives of potatoes and wild relatives for the presence of k-mers specific to this haplotype. Simultaneously, we evaluated a panel of diploid clones with SC haplotype specific KASP markers. The well-known SC clones M6, G254, US-W4 and RH89-039-16 were found positive for this haplotype. We conclude that all those clones have the *Sli* locus. Interestingly, k-mers specific to the SC haplotype were also commonly found in several tetraploids cultivars. Reconstruction of the breeding history with pedigree data points to Rough Purple Chili as origin of *Sli*, thus explaining its widespread occurrence in *S. tuberosum* breeding germplasm.

## **W868: Potato Genomics**

### **Joint QTL Analysis of a Tetraploid Potato Diallel Population**

**Jeffrey B. Endelman** and Maria V. Caraza-Harter, University of Wisconsin, Madison, WI

The genetic mapping of quantitative trait loci (QTL) in tetraploid potato has contributed to our understanding of traits and breeding progress. Initially, QTL mapping relied on biparental F1 populations; more recently, genotyping-by-sequencing and SNP arrays have been used in genome-wide association studies (GWAS) to identify QTL across a broader set of germplasm. Although the apparent simplicity of GWAS makes it attractive, the high genetic diversity of cultivated potato makes it economically challenging to genotype large populations with enough haplotype-specific markers. Following the trend already established in diploid species, an alternative approach is to perform QTL mapping with founder genotype probabilities in multi-parent populations. To explore this strategy, a diallel population of 435 progeny was established by crossing three parents from the US red market class in all pairwise combinations. The population was genotyped with the potato V3 SNP array and phenotyped for several quantitative traits, including vine maturity, tuber shape (length/width ratio), and skin color. Genotype probabilities were calculated for each F1 population separately and then combined for joint QTL analysis. Significant QTL were detected at the genomic positions corresponding to previously characterized genes for each trait (*CDF1* for vine maturity; *OFP20* for tuber shape; *an2* for skin color), but a new locus affecting red skin color was also detected on chromosome 12. According to the Deviance Information Criterion (DIC), which is a penalized measure of model fit, for every trait the founder allele model was superior to the best SNP detected by GWAS. For vine maturity and skin color, the additive model was selected based on the DIC; for tuber shape, digenic dominance was also included. Software for public

release is being developed with collaborators at Wageningen University and the University of Florida, which will enable others to pursue this promising new strategy for QTL mapping in tetraploid crops.

## **W869: Poultry 1**

### **Major Advances in Defining Variability and Function of Chicken MHC-Y Region Genes**

**Marcia M. Miller**<sup>1</sup>, Jibin Zhang<sup>1</sup>, Ronald M. Goto<sup>1</sup>, Christa F. Honaker<sup>2</sup>, Paul B. Siegel<sup>2</sup>, Robert L. Taylor, Jr.<sup>3</sup> and Henk Parmentier<sup>4</sup>, (1)Beckman Research Institute, City of Hope, Duarte, CA, (2)Department of Animal and Poultry Sciences, Virginia Tech, Blacksburg, VA, (3)West Virginia University, (4)Wageningen University, Netherlands

MHC-Y is a second region of major histocompatibility complex (MHC) genes in the chicken genome. MHC-Y contains distinctive, polymorphic MHC class I-like genes, along with members of four additional gene families and hundreds of interspersed repetitive sequences. We have developed a simpler, more efficient method for distinguishing MHC-Y haplotypes. Southern hybridizations are no longer needed. The new method is based on capillary electrophoresis of products generated by PCR from a short tandem repeat (STR) region located immediately upstream of the class I-like genes in MHC-Y. The chromatograms generated are highly reproducible and easily scored. Hundreds of samples can be genotyped efficiently. We have begun to study large populations of birds for association between MHC-Y genotype and immune function. We first typed White Leghorn chickens at Virginia Tech that have been selected 46 generations for high and low antibody responses observed following i.v. challenge with sheep red blood cells (SRBC). Results clearly show highly asymmetric distributions of MHC-Y haplotypes in the high antibody selected (HAS) and low antibody selected (LAS) lines. To gain insight into whether asymmetry in haplotype distribution might have occurred by chance, we analyzed a second pair of (brown layer) lines at Wageningen University similarly selected for over 30 generations for high and low antibody responses following i.m. challenge with SRBC. The MHC-Y genotypes are also asymmetrically distributed in the Wageningen lines. Similar findings in these two independent studies increase the likelihood that the association between antibody response and MHC-Y genotype is genuine rather than a chance event.

Supported in part by USDA NIFA Grant 12233246.

## **W870: Poultry 1**

### **Role of Humoral Immunity in Clearance of Lentogenic Newcastle Disease Virus in Chickens**

Karen Tracy<sup>1</sup>, R.a. Gallardo<sup>2</sup>, Emily Aston<sup>1</sup> and **Huaijun Zhou**<sup>3</sup>, (1)University of California, Davis, (2)University of California, Davis, School of Veterinary Medicine, Population Health and Reproduction, Davis, CA, (3)Department of Animal Science, University of California, Davis, CA

The vulnerability of poultry to viral diseases, including velogenic Newcastle disease virus (vNDV), poses a serious threat to human food security and health. vNDV causes very high mortality in unvaccinated domestic chicken flocks, but low-virulence lentogenic NDV strains are commonly used for vaccination. Despite the existence of these effective vaccines, the relative importance of innate, cellular, and humoral immunity in clearing the virus is poorly understood. We therefore investigated the specific role of the humoral response in clearing an infection with a high titer of the lentogenic La Sota strain of NDV. We ablated B cells in commercial Hy-line W-36 chickens by targeting the bursa, the site of avian B cell development. Birds were surgically or chemically bursectomized starting on the day of hatch and subsequently infected with  $10^7$  or  $4 \times 10^8$  EID50 NDV at 3 weeks of age. Surgical bursectomy significantly reduced production of IgY specific for NDV antigens at 14 days post-infection (DPI), and chemical bursectomy completely eliminated production at 7, 14, and 21 DPI. However, neither treatment had a large effect on viral titers in the tears or slowed viral clearance. Clinical signs observed after infection with  $4 \times 10^8$  EID50 NDV were not worsened by cyclophosphamide treatment. We conclude that humoral immunity is not required for clearance of a primary infection with a lentogenic strain of NDV in chickens.

## **W871: Poultry 1**

### **New Insights into Genome Structural and Somatic Variation in Chicken**

**Giridhar Athrey**, Texas A&M University, College Station, TX

The chicken, *Gallus gallus*, was the first domestic animal to have its genome assembled, and has been at the forefront of discoveries in genome biology of avian species. Even as we have gathered a tremendous amount of knowledge about the associations between the chicken genome and phenotypes, much remains to be uncovered. The advancement of genome sequencing and assembly technologies is enabling us to learn new aspects of the chicken's genome biology.

Genome structural variants, and somatic mutations in chicken are two of the most fundamentally important and interesting characteristics relevant to understanding the relationship of genotype to phenotype. While the interest in chicken genome structural variants has surged, and new details have been revealed in the past few years, their associations with domestication traits, or commercially important traits remain unresolved to a large extent. The causes and extent of somatic mutations in chicken is mostly unknown.

In this talk, I will present results from our studies based on both short- and long-read technologies that is providing new details about the functional role of structural variants, and association with phenotypes - based on comparison of the ancestral Red Junglefowl with commercial broilers. Furthermore, I will present the first report (to my knowledge) of somatic mutations in chicken skeletal tissue experiencing oxidative stress. Our results suggest new hypotheses that can link high genome variants and high metabolic rate of modern broilers to oxidative stress.

## **W872: Poultry 1**

### **Non-Coding Circular RNA Expression in Turkey Skeletal Muscle**

**Kent M. Reed**, University of Minnesota, St. Paul, MN

Circular RNAs (circRNAs) are novel, single-stranded RNAs and an emerging area of investigation in regulation of gene expression. These RNAs are generated from exonic/intronic sequences joined head to tail (back spliced). Although their functions are not fully understood, circRNAs are hypothesized to function as modifiers of gene expression by acting as microRNA sinks and regulators of splicing and transcription. Conventional RNA-seq analyses discount the presence and significance of these molecules but analysis with new algorithms have found circRNAs to be widely expressed with modest sequence conservation. This study was designed to mine turkey RNA-seq data to test for the presence of circRNA species. Sequencing data generated from a study of turkey poult muscle tissues exposed to thermal challenge were used in this study. RNA-seq libraries were prepared from a comparatively slower-growing (randombred control 2, RBC2) line, and a faster-growing body-weight selected (F) line. RNA-seq reads (average 18,788,823 reads/library) from the 28 paired-end libraries were mapped to the turkey genome (UMD5.0). Circular RNAs were predicted with CIRI software and over one thousand potential circRNAs were identified in the dataset. Targeted PCR was used to confirm presence of a subset of the predicted circRNAs within the RNA pool. Differential expression analysis suggests that the abundance and relative expression of these non-coding RNAs have a potential role in muscle to the biological response to thermal challenge.

## **W873: Poultry 1**

### **Integrating Genetics and "Omics" Data to Dissect Chicken Resistance to Infectious Disease in Low and Middle Income Countries (LMICs).**

**Androniki Psifidi**<sup>1</sup>, Georgios Banos<sup>2</sup>, Pete Kaiser<sup>3</sup>, Olivier Hanotte<sup>4</sup>, Joshi Chaitanya<sup>5</sup>, Fiona Tomley<sup>1</sup> and Damer Blake<sup>1</sup>, (1)Royal Veterinary College, University of London, Hatfield, United Kingdom, (2)The Roslin Institute and Scotlands' Rural College, Midlothian, Scotland, (3)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (4)LiveGene – CTLGH, International

Livestock Research Institute, Addis Ababa, Ethiopia, (5)College of Veterinary Science and Animal Husbandry, Anand, India

Poultry play an important role in the agriculture of many Low and Middle Income Countries (LMICs). Many chickens in sub-Saharan Africa and South/South East Asia are indigenous or tropically –adapted, and often raised under scavenging conditions. Although these birds are well adapted to local conditions, vaccination and optimal biosecurity measures are rarely applied. Infectious diseases remain a major cause of mortality and reduced productivity. Moreover, the risk of zoonotic disease outbreaks is increased in these parts of the world. Breeding for increased resistance to infectious diseases as well as for beneficial enteric microbiome composition, with relevance to competitive exclusion, offers a potentially sustainable solution to improve animal and human health and welfare. However, genomic studies of immune responses, infectious disease resistance and gut microbiome composition have been limited in LMICs. Moreover, successful genomic selection programmes require large numbers of individual birds with genomic and phenotypic information, which may be a challenge to achieve in the numerically small populations of indigenous chicken ecotypes. Currently, we are investigating the feasibility of genomic selection programmes and the best approaches to enhance antibody response and resistance to major infectious diseases, and breed for beneficial microbiome composition. Our results so far have been encouraging; however further studies are needed to validate them and establish the best strategy to genetically improve chickens in LMICs.

### **W874: Poultry 1**

#### **Infectious Bronchitis Virus Infection Affects Chromatin Accessibility and RNA Differential Expression in a Tissue-Specific Manner**

**Yvonne Drechsler**, College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA and **R. David Hawkins**, University of Washington, Seattle, WA

The goal of this study is to elucidate the genetic and epigenetic mechanisms and gene regulatory networks that control immune responses to coronavirus infection, which will yield a better understanding of processes that are essential to mounting the appropriate immune response. Coronavirus infections are highly contagious and associated with severe respiratory and gastrointestinal diseases in humans and animals. A lack of understanding of the molecular mechanisms governing pathogenesis of coronaviruses has severely hampered progress in designing effective strategies for prevention and treatment. Developing new vaccine strategies and treatment require a better understanding of the underlying causes of pathogenesis and viral molecular biology. The avian coronavirus, infectious bronchitis virus (IBV), infections result in respiratory, enteric and reproductive disease that lead to significant morbidity and mortality. Serotype M41 causes severe respiratory disease while IBV Gray is nephropathogenic. We analyzed chromatin accessibility via ATAC-seq and RNA expression in tissues of M41 and Gray infected birds, specifically lungs and kidneys. We found epigenetic changes specific to the strains and their target tissues. M41 infected birds showed more changes in lungs, whereas Gray infected birds showed more changes in the kidney, corresponding to the tissues that display more pathogenesis specific to the serotype. There were also differential gains and loss of chromatin accessibility associated with the two serotypes. RNA sequencing analysis showed differential gene expression as well, specific to strain or tissue investigated. In conclusion, there is evidence of the virus causing serotype-specific epigenetic changes in the target tissues affected by M41 and Gray.

### **W875: Poultry 1**

#### **Role of the Chicken T Cell Receptor-Beta Repertoire in Genetic Resistance to Marek's Disease**

**Cari Hearn** and **Hans Cheng**, USDA, ARS, Avian Disease and Oncology Laboratory, East Lansing, MI

Marek's disease (MD) is a T cell lymphoproliferative disease of chickens induced by the Marek's disease virus (MDV), an oncogenic alphaherpesvirus. MDV has repeatedly shown increased virulence in MD-vaccinated flocks, prompting continued efforts in improved vaccines and enhanced genetic resistance. Model pairs of genetically MD-resistant and susceptible chicken lines that are either MHC-matched or MHC-congenic has allowed the study of T cell receptor (TCR) repertoires associated with MDV infection.

Chickens resistant to MD showed higher usage of Vbeta-1 TCRs than susceptible chickens, in both the CD8 and CD4 subsets in the MHC-matched model, and in the CD8 subset only in the MHC-congenetic model; and Vbeta-1+ CD8 cells expanded during MDV infection. The TCR locus was found to be divergent between MD-resistant and susceptible chickens in the MHC-matched model, with MD-resistant chickens expressing a greater number of Vbeta-1 TCRs and an increased representation of Vbeta-1 CDR1 loops with an aromatic residue at position 45. TCR Vbeta-1 CDR1 usage in resistant x susceptible F1 birds indicated that the most commonly used CDR1 variant was present only in the susceptible line, suggesting that selection for resistance in the MHC-matched model has optimized the TCR repertoire away from dominant recognition of one of the B2 haplotype MHC molecules. Finally, TCR downregulation during MDV infection in the MHC-matched model was observed most strongly in the MD-susceptible line, and TCR downregulation due to viral reactivation in a tumor cell line could be demonstrated to be virus-specific and not due to apoptosis induction.

### **W876: Poultry 1**

#### **Differential, Allele Specific, and Co-Expression in the Chicken Spleen Transcriptome under Avian Pathogenic *Escherichia coli* Infection**

Melissa S. Monson, Michael G. Kaiser, Anna Wolc and **Susan J. Lamont**, Department of Animal Science, Iowa State University, Ames, IA

Avian pathogenic *Escherichia coli* (APEC) decreases growth and egg production in chickens and can cause high levels of mortality. Our long-term goal is to reduce the negative impact of APEC on poultry welfare and performance by facilitating improved veterinary and animal breeding strategies; this will require understanding host functional responses to infection, such as changes in gene expression, molecular pathways and regulatory elements. In this study, F1 progeny from reciprocal crosses between Fayoumi (disease-resistant) and broiler (disease-susceptible) lines were inoculated with APEC O1:K1:H7 or sterile PBS at 14 days of age. Spleen tissue was collected 1 or 2 days post infection (DPI) for bacteriology and RNA-seq. Reads were generated using the Illumina HiSeq 3000 (n = 5-6 samples/F1 cross/inoculation/DPI) and were mapped to the chicken genome (Galgal6a). Differential expression (DE) analysis (edgeR) identified 580 significant genes responding to APEC at 1 DPI and 157 genes at 2 DPI, with greatest increases in pro-inflammatory genes (such as *IL22*, *IL17A*, and *PTX3*). Co-expression patterns (WGCNA) revealed 10 significant modules, which include immune response and cell division genes. Allele specific expression (ASE; GATK4 pipeline) provided markers for *cis*-acting regulatory elements impacted by APEC infection, including APEC-specific ASE within genes with significant DE (such as *SERPING1* and *MGST1*). Overall, transcriptome analysis provided insight on early immune responses to APEC in the spleen and targets to investigate for reducing colibacillosis in chickens. Support: USDA-NIFA-AFRI #2015-67015-23093 as part of the joint NIFA-BBSRC Animal Health and Disease program and Hatch project #5424 and #5458.

### **W877: Poultry 1**

#### **Prophylactic Reduction of Bacterial Chondronecrosis with Osteomyelitis in Broilers**

**Adnan Alrubaye**, University of Arkansas/Fayetteville, Fayetteville, AR

We have determined that Zinpro Availa-ZMC is effective in reducing the incidence of lameness by 20-25% (P<0.05) in broilers trials. Bacterial Chondronecrosis with Osteomyelitis (BCO) lameness is a major animal welfare issue that affects the commercial broilers industry in the United States and many other countries across the World. Lameness due to BCO is caused by a variety of pathogenic microorganisms as well as stress factors, which degrade the growth plates of the femur and tibia. Our work demonstrates that bacteria leak across epithelia of the respiratory and gastrointestinal systems into the bloodstream. Bacteria that colonize the growth plates lead to necrosis leading to an inability to stand or walk due to pain resulting from severe damage. Thus, BCO is considered a major animal welfare issue where outbreaks in broiler operations may reach lameness of 10%. There are many factors that can stress birds like rapid growth, handling, transportation, heat, and respiratory infections, leading to increased BCO lameness. We have used the wire flooring model developed by Dr. Wideman, University of Arkansas Department of Poultry Science, to create stress experimentally in order to induce lameness.

We also developed a bacterial challenge with *Staphylococcus agnetis*, that mimics the lameness epidemic for birds on litter flooring. Challenged birds will experience lameness at 70%, while the unchallenged birds in the same facility experience BCO lameness of 50% lameness. The Wire-Flooring model induces lameness by inducing a physical stress on long bones, which will ultimately lead to microfractures in the proximal femoral and tibial heads. The litter flooring with *S. agentis* challenge is a better system for evaluation of epithelial integrity and immune function, using conditions that mimic real field conditions.

We have tested several probiotics and prebiotics in the past years and determined that the Zinpro trace minerals Availa-ZMC – zinc, manganese and cobalt – was effective in reducing the incidence of lameness in our broilers trials. The Availa-ZMC improved gut integrity by strengthening the tight junctions in the epithelia reducing bacterial leakage, enhanced bactericidal activity of peripheral blood monocytes, and reduced overall lameness. Improving gut health, strengthening tight junctions, and improving bone strength are all key to preventing BCO lameness in chickens and improving the animal welfare of broilers.

### **W878: Poultry 1**

#### **Host Response to *Eimeria* spp Infection in Meat-Type Chickens**

**Samuel E. Aggrey**<sup>1</sup>, Marie C. Milfort<sup>2</sup>, James C Foutz<sup>2</sup>, Alberta L Fuller<sup>1</sup> and Romdhane Rekaya<sup>3</sup>, (1)University of Georgia, Athens, GA, (2)University of Georgia, (3)University of Georgia, Department of Animal and Dairy Science, Athens, GA

A study was conducted to identify transcriptome differences between two chicken genotypes infected with *Eimeria* (*E.*) *acervulina* and to understand the underlying gene expression changes and further delineate genotype-specific effects. Fourteen day old chicks of an unimproved (ACRB) and improved (COBB 500) genotype were orally infected with  $2.5 \times 10^5$  sporulated *E. acervulina* oocysts. At 6 day-post infection, the jejunum of 5 birds from each genotype and their controls were used for global transcriptome analysis. There were 5,262 differentially expressed genes (DEGs) between the ACRB infected and controls, and 2,884 DEGs between COBB infected and controls. There were common pathways between ACRB and COBB genotypes in response to *E. acervulina* infection. Among the common pathways were actin cytoskeleton, MAPK signaling, focal adhesion and cell adhesion molecules. However, there were also genotype-specific pathways in response to the infection. Whereas Toll-like receptor and retinol pathways were specific to the COBB genotype, oxidative phosphorylation and calcium signaling pathways were specific to the ACRB genotype. Both ACRB and COB genotypes have shared pathways in responding to *Eimeria* spp infection, however, each genotype also show distinct transcriptome signature.

### **W879: Poultry 1**

#### **Hypothalamic Structures Containing Corticotropin Releasing Hormone and Vasotocin Neurons and Receptors Involved in Sustaining the Stress Response in Poultry**

**Wayne Kuenzel**, Hakeem Kadhim and Seong W. Kang, University of Arkansas, Fayetteville, AR

Recently we have discovered that a nucleus within the septal brain region, the nucleus of the hippocampal commissure (NHpC), contains CRH neurons that appear to show the first response to the stressor, food deprivation (FD), with increased CRH gene expression that peaks at 2h of FD. CRH neurons in the hypothalamic paraventricular nucleus (PVN) show a different expression pattern with a steady, significant increase in CRH mRNA that peaked at 8h of FD. An experiment was therefore designed to determine the gene expression pattern of the other major stress neuropeptide, arginine vasotocin (AVT) and its two major receptors, V1aR and V1bR, following the same stressor, FD. Male Cobb 500 chicks, at 14 days of age, were divided into 6 treatment (trt) groups (n = 10 birds/trt): controls (CON, fed *ad libitum*) and 5 groups subjected to 1h, 2h, 3h, 4h and 8h of FD. Blood, brain and anterior pituitary were sampled from each bird. The NHpC and PVN were micro-dissected from brain and total RNAs were extracted for qRT-PCR. Plasma from blood samples was utilized to determine stress

hormone levels of corticosterone (CORT) by radioimmunoassay. In the PVN no change in AVT gene expression occurred during the first 2h of FD. At 3h of FD a significant increase in AVT expression occurred and remained significantly elevated at 4h and 8h of FD. Of interest, both the V1aR and V1bR showed significantly elevated expression during the same time periods as AVT expression within the PVN, thereby displaying positive feedback of both receptors with AVT expression. In contrast, within the NHPc, AVT mRNA was significantly lower at 1h FD and showed steadily decreasing levels thereafter with lowest expression of AVT at 8h of FD. Gene expression of the V1bR in the NHPc significantly increased at 2h of FD and peaked at 8h of FD. The V1aR showed significantly reduced levels of mRNA at 1h and 2h of FD, returned to control levels at 3h of FD and displayed significantly elevated mRNA at 8h FD stress. Overall in contrast to the PVN, the NHPc displayed a negative feedback of both the V1aR and V1bR with the neuropeptide AVT. In summary, following a response to FD stress, CRH neurons in the NHPc appear to be first responders quickly followed by CRH neurons in the PVN. Thereafter, a delayed activation of AVT expression in the PVN occurred beginning at 3h of FD, that sustained the stress response based upon a positive feedback of increased gene expression of its two receptors, the V1aR and V1bR, as well as the pattern of plasma CORT beginning at 1h of FD that displayed a steady and significant rise with a plateau shown at 4h to 8h of FD. Taken together, results suggest that the NHPc in the septal brain region containing CRH neurons and an AVT terminal field might play a critical role in the neuroendocrine stress response of poultry. Supported by a grant from the Arkansas Biosciences Institute, Division of Arkansas and HCED fellowship from Iraq to HK.

### **W880: Poultry 1**

#### **Gene Networks Expressed in the Endocrine System Controlling Egg Production Rates in Turkey Hens**

**Tom E. Porter**, University of Maryland, College Park, MD

Ovulation and egg production in birds are controlled by the hypothalamo-pituitary-gonadal (HPG) endocrine axis. Within flocks of commercial turkey hens, high egg-producing hens (HEPH) show increased hypothalamic, pituitary, and ovarian gene expression related to HPG axis stimulation, when compared to low egg-producing hens (LEPH). Moreover, pituitary and ovarian cells from HEPH show increased *in vitro* responsiveness to hormonal stimulation. RNAseq analysis was performed on hypothalamic, pituitary, and ovarian cell samples from LEPH and HEPH, both during and outside of the preovulatory hormonal surge, to identify novel regulators of HPG axis function. HEPH displayed up-regulation of gene expression in the hypothalamus and pituitary gland related to stimulation of gonadal function and up-regulation of steroidogenic genes in the granulosa cells of the largest preovulatory follicle (F1G) and cells of the small white follicles (SWF). Pathway analysis of the RNAseq results indicated differential regulation between LEPH and HEPH of thyroid hormone (T3) and estradiol (E2) signaling in all tissues. In LEPH, up-regulated genes showed enrichment of T3 signaling. In HEPH, E2 was identified as an upstream regulator that was activated relative to LEPH. In pituitary cell cultures from LEPH and HEPH, T3 pretreatment decreased gonadotropin mRNA levels in cells from both LEPH and HEPH, with the effect being more prominent in HEPH cells. E2 pretreatment increased gonadotropin mRNA levels in LEPH cells. Treatment with T3 of SWF cells decreased E2 production from HEPH cells to the levels seen in LEPH cells, whereas T3 treatment did not impact E2 production in LEPH cells. In conclusion, up-regulation of T3 activity in LEPH and E2 activity in HEPH may play a role in regulating HPG axis function and ultimately account, in part, for differences in egg production rates in turkey hens.

### **W881: Poultry 1**

#### **Delayed Access to Feed Affects Goblet Cell Distribution in Chickens**

**Eric A. Wong**, Virginia Tech, Blacksburg, VA and Kuan-Ling Liu, Virginia Tech

Newly hatched chicks are often subjected to delayed access to feed (DAF) and water up to 48 hours posthatch, which impacts small intestinal development. Goblet cells arise from the stem cell pool present in the intestinal crypt and secrete mucin2 (Muc2), which contributes to the mucus layer. The objective of this study was to determine the effect of DAF on Muc2 mRNA abundance and the number and distribution of goblet cells. Cobb-500 broiler chicks hatching within a 12-hour window were randomly

distributed into three groups: control with no feed delay (ND), 24-hour feed delay (D24), and 36-hour feed delay (D36). Duodenum, jejunum, and ileum were collected at 0, 24, 36, 72 and 120 hours posthatch. Muc2 mRNA abundance was quantified by qPCR. Stem cells expressing olfactomedin 4 (Olfm4) and goblet cells expressing Muc2 mRNA were identified by in situ hybridization. Statistical analysis was performed using JMP Pro 14 and significant differences were determined by t-test and one-way ANOVA. DAF caused a decrease in Muc2 mRNA abundance, which was lower in the duodenum of D24 and D36 chicks compared to ND chicks at 72 hours and lower in the jejunum of D36 chicks compared to ND and D24 chicks at 36 hours. Cells expressing Muc2 mRNA were detected in both the crypt and along the villi. The crypt cells expressing Muc2 mRNA also expressed the stem cell marker Olfm4 mRNA. These Olfm4 and Muc2 mRNA-expressing crypt cells have only been reported in chickens. The number and distribution of goblet cells in both the upper and lower half of the villi were determined and expressed as a ratio (VU/VL). The VU/VL for goblet cells was greater in D24 chicks compared to ND chicks at 24 hours in the jejunum and ileum and greater in D36 chicks compared to ND and D24 chicks at 36 hours in the jejunum. This result showed that there were fewer goblet cells in the lower half compared to the upper half of the villus following DAF, which may be due to a reduction in the number of goblet cells emerging from the crypt. The combined reduction in Muc2 mRNA expression and number of goblet cells could result in a decrease in the mucus layer and an increase in the risk of infection from pathogens.

## **W882: Poultry 1**

### **Impact of High Dietary Selenium on the Transcriptome and Selenometabolites in Turkeys**

**Roger A Sunde**, University of Wisconsin, Madison, WI and **Rachel M Taylor**, University of Wisconsin-Madison, Madison, WI

Selenium (Se) has long been known as a toxic element for animals, and yet we lack good biomarkers for assessing high Se status, and little is known about mechanism(s) of Se toxicity at a molecular level. In contrast, the levels of selenoproteins and of a subset of selenoprotein transcripts decrease dramatically in Se deficiency. (Sunde et al. 2016).

To assess the effect of Se status on the full transcriptome, RNA-Seq was conducted in collaboration with KM Reed and KM Mendoza (U Minn). Day-old turkeys were fed truly Se-deficient diets ( $<0.005 \mu\text{g Se/g}$ ) supplemented with graded levels of inorganic Se (as selenite) up to  $5 \mu\text{g Se/g}$ . In response to high Se intake (1, 2 and  $5 \mu\text{g Se/g}$ ) vs. Se-adequate ( $0.4 \mu\text{g Se/g}$ ), there were only a limited number of significant differentially-expressed transcripts, all with only relatively small fold-changes; no transcript showed a consistent pattern of altered expression across high-Se treatments, and no associated metabolic pathways and biological functions were significant and consistently found with high-Se. A comparison with parallel studies in rats also failed to identify common differentially-expressed transcripts. Collectively, this study indicates that turkeys do not alter gene expression as a homeostatic mechanism to adapt to high Se.

Liver Se in turkeys fed  $5 \mu\text{g Se/g}$  accumulates to 6X the level in poult fed  $0.4 \mu\text{g Se/g}$ . To identify and characterize turkey liver selenometabolites, HPLC-ICP-MS and HPLC-ESI-MS/MS was conducted in collaboration with K Bierla, J Szpunar, and R Lobinski (CNRS-UPPA, Pau, France). In Se-deficient, Se-adequate and high-Se livers, no selenomethionine (SeMet) was detected showing that turkeys do not synthesize SeMet *de novo* from inorganic Se. Selenoprotein Se (selenocysteine, Sec) content in high-Se liver only doubled compared with Se-adequate liver, indicating that the 6-fold increase in liver Se was not due to increased selenoprotein. What increased dramatically were the low molecular weight (LMW) selenometabolites: glutathione-, cysteine- and methyl- conjugates of the selenosugar, seleno-N-acetyl galactosamine (SeGalNac). In addition, more Se was present as selenosugars decorating general proteins via mixed-disulfide bonds than was present in Sec-containing selenoproteins in Se-adequate liver. In high-Se liver, these "selenosugar-decorated" proteins plus LMW selenosugars comprised ~90% of the Se in the water-soluble fraction. These analyses show that increased selenosugar formation occurs with high Se supplementation, further increasing selenosugar-decorated proteins, but also

increasing selenosugars linked to LMW thiols, and leading to formation of methyl-SeGalNac. This pathway appears to underlie how avians and other animals adapt to high Se.

Funding: USDA Hatch 1013496, and Wisconsin Alumni Foundation 12046295

### **W883: Poultry 1**

#### **Genetics of Ascites Revealed by Whole Genome Resequencing in Two Distinct Broiler Lines**

**Douglas D. Rhoads**, University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR

We used whole genome resequencing (WGR) to identify candidate regions for ascites in broilers in two distinct broiler lines. The two lines were the Relaxed line (REL) representing a commercial elite line of the late 1990s, and the other a commercial product (CP) from a different breeding company. Both lines were phenotyped for ascites in the hypobaric chamber. For the REL we mapped determinants in both genders while in CP we only analyzed males. For the REL we sequenced pooled DNAs (n=10) for each gender and phenotype with two biological replicates and each pool was sequenced to 66 Gigabases. For CP we barcoded each DNA individually and sequenced to >5 Gbases. SNP frequency and read depth were used to filter the data which was then plotted by chromosomal position to identify regions with groups of SNPs which were statistically differentially represented between phenotypes. For REL we identified 31 regions and have validated two of these regions by genotyping a larger collection of DNAs for specific SNPs in those regions. These two regions represent the first verified loci for ascites in broilers and we are evaluating these regions for production traits using Marker Assisted Selection. However, these same two regions are not related to ascites in CP and the regions associated with ascites in CP appear to be distinct. WGR represents a straightforward, cost-effective method for mapping the genetics of complex traits in broilers with no *a priori* assumptions. WGR allows us to analyze the segregation of tens of millions of SNPs per genome for as little as \$50 per genome. WGR has demonstrated that the genetics of ascites appears to be highly polygenic when one considers different genetic stocks.

This project was supported by Agriculture and Food Research Initiative Competitive Grants no. 2015-67015-22960 and 2018-67015-28244 from the USDA National Institute of Food and Agriculture

### **W884: Poultry 1**

#### **Probing the Capacity for Synthesis of Long Chain Omega-3 Polyunsaturated Fatty Acids in the Developing Broiler Chick**

Robert Mihelic<sup>1</sup>, Heather Winter<sup>1</sup>, Kurt Lamour<sup>2</sup>, Josh Powers<sup>3</sup>, Shawn Campagna<sup>3</sup> and **Brynn Voy**<sup>1</sup>, (1)University of Tennessee, Department of Animal Science, Knoxville, TN, (2)University of Tennessee, Department of Entomology and Plant Pathology, Knoxville, TN, (3)University of Tennessee, Department of Chemistry, Knoxville, TN

Eicosapentaenoic acid (EPA; 20:5 n-3) and docosahexaenoic acid (DHA; 22:6 n-3), the omega-3 fatty acids found in marine oils, suppress adipogenesis and inflammation, increase bone strength, bolster the immune system, and enhance energy utilization in chickens. Unlike most species, chickens can synthesize considerable amounts of EPA and DHA by elongating and desaturating the essential fatty acid alpha-linolenic acid (ALA; 18:3 n-3). However, despite the potential production benefits of enhancing this pathway in broilers, the elongase and desaturase genes in the *Gallus gallus* genome are not well-characterized. The objective of this study was to define the expression patterns of fatty acid elongase and desaturase genes during adipose development in broiler chicks, and to associate expression with tissue fatty acid abundance. Expression patterns were analyzed in the context of a larger set of adipogenic and metabolic genes to enrich functional annotation. We utilized targeted RNA sequencing to efficiently quantify expression of a large gene set across a large sample set. Primers were designed to query expression of genes with known roles in metabolism and adipose development, including six elongation of very long chain fatty acid genes (*ELOVLs* 2-7), three fatty acid desaturase genes (*FADS1*, 2, and 6), and two stearyl-CoA desaturases (*SCD* and *SCD5*). Adipose tissue was collected from broiler

chicks (Ross 308; n=5-7/age) at embryonic days 13, 15 and 17 (subcutaneous (SQ)) and at 7 and 14 d post-hatch (SQ, abdominal (AB) and neck) and snap-frozen. This age span is a time of dynamic metabolic adaptation as chicks transition from almost complete reliance on lipids *in ovo* to a carbohydrate-rich diet after hatch. Total RNA was isolated from each depot, quantified, quality-checked, and used to synthesize cDNA. Raw sequencing reads were aligned to the chicken genome (*Gallus\_gallus*-5.0) and reads were counted using CLC Genomics Workbench (Qiagenbioinformatics.com). Reads for each gene were normalized to total counts within sample and analyzed for effects of age and depot in SAS (V 9.4) using ANOVA and least squares means for contrasts ( $\alpha=0.05$ ). Relative fatty acid abundance was measured in subcutaneous adipose samples using GC/MS. Multivariate analyses were used to interrelate gene expression and fatty acid abundance. Expression levels of five of the six ELOVL genes (*ELOVL2*, 3, 4, 5 & 7) and each of the five desaturase genes varied significantly with age in SQ adipose ( $p<0.05$ ). Correlations between gene expression and fatty acid profiles suggests that two pathways for EPA and DHA synthesis may exist in broiler chick adipose tissue. They also indicate that *ELOVL7*, which has not been characterized in chickens, may play a significant role in synthesizing DHA in chick adipose tissue, both during embryonic development and after hatch. Novel roles for *ELOVL3* and for *SCD5* in fatty acid metabolism and adipocyte function are also suggested by interrelationships between chick age, gene co-expression, and fatty acid abundance. Collectively, these data provide new insight into pathways that control the synthesis of bioactive lipids that influence metabolism, growth and immunity in chickens.

#### **W885: Poultry 1**

##### **Blood Plasma Biomarkers Representing Woody Breast in Broiler**

**Byungwhi C. Kong**, Wayne Kuenzel, Casey M. Owens, Walter Bottje, Seong W. Kang, Kentu Lassiter and Sara Orłowski, University of Arkansas, Fayetteville, AR

Breast myopathies including white striping and woody breast (WB) are meat quality defects that result in an annual loss of \$200 million or more to the U.S. poultry industry. WB is the increased hardness of raw breast fillets. Degenerative myopathies result in increased fat and decreased protein content in broiler breast meats. To gain insights into the etiology, studies have evaluated the tissue changes associated with WB. Additionally, less invasive biomarkers contained in blood samples, such as plasma proteins, can be effective for pre-diagnosis of breast myopathies during the normal growth phase of broilers. Thus, the present study used proteomics approaches to analyze blood plasma samples with different degrees of myopathies to identify differentially abundant proteins at 4- and 8 wks of age. To explore identifying early-age-biomarkers, that can represent the potential onset of WB, blood samples were collected from 100 ROSS 708 male broilers. At 8 wks of age, WB conditions were scored by palpitation. A total of 32 plasma samples (8 for each group of CON and WB birds; at 2 time points, 4 wks and 8 wks) was subjected to Tandem Mass Tag labeling with Orbitrap Fusion analysis, that were performed by the Proteomics Core, University of Arkansas Medical Sciences. Results showed protein amounts of CCL26, COL1A1, and COL1A2 contained less in WB plasma at both 4- and 8 wks of age.

#### **W886: Poultry 1**

##### **Genetic Basis of Wooden Breast and White Striping in Commercial Broiler**

**Behnam Abasht**, University of Delaware, Newark, DE

#### **W887: Poultry 1**

##### **Prediction of Wooden Breast in the Liver and Pectorals Major using Machine Learning**

**Christopher M. Ashwell**, North Carolina State University, Raleigh, NC

#### **W888: Poultry 1**

##### **The Chicken Wingless-2 Mutation Takes Flight**

Ingrid A. Youngworth, Stanford University, Stanford, CA and **Mary E. Delany**, Animal Science, University of California, Davis, CA

The autosomal recessive lethal chicken wingless-2 (*wg-2*) mutation results in a developmental syndrome characterized by absent wings, truncated legs, craniofacial defects as well as skin and feather defects, and kidney malformations. Previously, mapping and genotyping utilizing a *wg-2* congenic inbred line established that the mutation resides within 227 kilobases (kb) of chromosome 12 encoding three verified/validated (*RAF1*, *CNBP*, *RAB43*) and three predicted (*ISY1*, *CERC5*, *EFCC1*) genes. Here we report on the results of a capture array designed to target and sequence the candidate region along with flanking DNA in 24 birds from the congenic inbred line. The results led to establishment of a smaller *wg-2* linked region of 119 kb. The variation detected in the region included point mutations and insertions or deletions. Interestingly, a point mutation predicted to cause a premature stop codon was detected in the *RAF1* gene of *wg-2* (versus the wild-type GGA 12 chromosome). We studied mRNA expression (RT-qPCR, Taqman) of *RAF1* in embryos of +/+, +/*wg-2*, *wg-2/wg-2* (mutants) genotypes as well as the expression of the other genes in the candidate region. Of the genes found to be expressed (*RAF1*, *CNBP*, *ISY1* (noting this gene is now validated by our finding of transcription)) we employed Western blotting to detect their protein expression. Interestingly, *RAF1* transcription was elevated in the mutants (relative to the other genotypes) yet there was no *RAF1* protein detected in the mutants and reduced protein in the heterozygotes (relative to wild type). *RAF1* encodes a protein integral to the Ras/Raf/MAPK signaling pathway controlling cellular proliferation and with known involvement in human diseases. Notably, human RASopathies are developmental syndromes caused by germline mutations in genes of this pathway. Our work suggests *RAF1* as a high priority candidate causative gene for chicken *wg-2* and further advances an animal model which could be useful in studying an important signaling pathway.

#### **W889: Poultry 2**

##### **NRSP-8 Update**

**James M. Reecy**, Department of Animal Science, Iowa State University, Ames, IA

#### **W890: Poultry 2**

##### **Genetic Analyses of African Local Chicken Ecotypes Challenged with Newcastle Disease Virus to Improve their Disease Resilience**

**Muhammed Walugembe**, Animal Science Department - Iowa State University, Ames, IA

Newcastle disease (ND) is a global threat to poultry, especially in low and middle-income countries, where entire small holder flocks are often lost to the disease. Local chicken ecotypes are important to rural family households through provision of high-quality protein in the form of eggs and meat and serve as a source of income. Studies were conducted in two countries, Ghana and Tanzania. In each country, three popular chicken ecotypes were challenged with a lentogenic (vaccine) strain of NDV. Various host response phenotypes, including viral load at 2 and 6 dpi, anti-NDV antibody levels (pre-infection and 10 days post-infection, dpi), and growth to 38 days of age, were measured. All birds were genotyped using a 600K Single Nucleotide Polymorphism (SNP) panel. We estimated genetic parameters and performed genome-wide association study (GWAS) analyses, using data on 1399 and 1440 birds from Tanzania and Ghana, respectively. Heritability estimates for the various traits ranged from moderate to high (0.18 – 0.55). Six and twelve quantitative trait loci (QTL) were identified by single-SNP analyses for growth and/or response to NDV for Tanzania and Ghana, respectively. Several locations of these QTL corresponded in location with genomic regions explaining >1% of the genetic variance identified by the Bayes B GWAS analysis method. Immune related genes were located in the QTL regions for some response traits. Significant SNPs from GWAS and other important SNPs from separate studies, along with SNPs spread across the genome were used in the development of a 5K SNP panel for use in imputation. The moderate estimates of heritability and identified QTL suggest that host response to NDV can be improved through selective breeding of African local chicken ecotypes to enhance NDV resilience and vaccine efficacy.

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Feed the Future Innovation Lab for Genomics to Improve Poultry (cooperative agreement number AID-OAA-A-13-00080).

## **W891: Poultry 2**

### **Poultry Genetic Resources: Catalyst for Common Use**

**Rosemary Walzem**, Texas A&M University, College Station, TX

Genetic methods used to modify mammals do not work for a major branch on the tree of life: birds. New science from several labs has made work with bird “germ lines” possible but exact approaches vary and make adoption of techniques daunting. One goal of our project is to make gene manipulations in birds a highly accessible technique for conservation, health and agriculture purposes. Genetic changes in chickens are possible; our team members can isolate culture and reintroduce the reproductive primordial germ cells (PGCs) needed to make such modifications. However, wild species of birds require differing and variable conditions for PGC applications, which has slowed expansion of PGC genetic technologies. Models using other bird species need development. Indeed our first project has funding from Revive & Restore, a non-profit group that seeks to enhance biodiversity through new techniques of genetic rescue for endangered and extinct species. The target species in this case is the Heath Hen (*Tympanuchus cupido cupido*), a grouse that formerly occupied scrubby heath-land barrens along the US eastern seacoast but went extinct in 1932. The strategy to revive the Heath Hen requires isolation of viable PGCs from its closest genetic relative, the Greater Prairie Chicken (GPC, *Tympanuchus cupido*) and subsequent gene editing to restore Heath Hen sequence. Viable GPC-PGCs were obtained by others but failed to go germline. A primary aim of this stage of the project is to determine whether there are cell phenotype markers such as gene expression patterns that can predict likelihood of germline transmission and so reduce the labor, cost and time of gene manipulations in birds.

## **W892: Poultry 2**

### **A Genetic Duplication Is Associated with Ectopic Expressions of HOX Genes and Alters Body Region Identity in Crested Chickens**

**Jingyi Li<sup>1</sup>**, Brian W. Davis<sup>1</sup>, Mi Ok Lee<sup>2</sup> and Andersson Leif<sup>1</sup>, (1)Texas A&M, College Station, TX, (2)Texas A&M University, college station, TX

Determination of regional body plan is fundamental to developmental biology. In animals, HOX genes are known to play a key role in this process. The ‘Crest’ trait in chicken, is a dominant mutation where the feather present on the head is replaced by those normally originating from dorsal skin. The association between Crest and the ectopic expression of the *HOXC8* gene suggests it is a direct modification of the standard body plan. In this study, by taking advantage of an improved genome assembly and publicly available WGS data, the causal mutation was identified as a small duplication in the intron of *HOXC10*. We found this variant perfectly associated with Crest in a panel of over 150 chicken populations. Additionally, we found that this duplication is associated with the ectopic expression of at least 6 flanking genes in the cranial skin of Crested chickens. Interestingly, we also see altered expression in the skulls of chickens with cerebral hernias, a phenotype involves the malformation of skull and is closely associated with Crest. We propose that the duplication disrupts the normal collinearity of HOX gene expression essential for standard body plan development. This, in turn, transforms the perceived body region identity of the feather follicles, resulting in dorsal-like feather production and potential perturbation of cranial structure. Our study suggests that these phenotypes are novel models to understand the unique mechanisms influencing HOX gene regulation.

## **W893: Poultry 2**

### **A Multiomics Toolbox to Advance Avian Research**

**Wesley Warren**, Animal Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO

## **W894: Poultry 2**

### **Chromosome-Level Genome and Comparative Analyses of Duck Uncover Genome and Chromatin Architecture Evolution of Birds**

**Jing Li**<sup>1</sup>, Jilin Zhang<sup>2</sup>, Zhengkui Zhou<sup>3</sup> and Qi Zhou<sup>1</sup>, (1)Life Sciences Institute, Zhejiang University, Hangzhou, China, (2)Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden, (3)Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

Duck is one of the most important poultry species and also a key model to study avian influenza viruses, yet its current genome quality is much lower than those of most birds. Here we combined diverse technologies (PacBio, 10X, BioNano and Hi-C) and RH linkage map, and produced a high-quality chromosome-level genome for Peking duck. The new genome (ZJU1.0) has been improved for contiguity by 70-fold greater than the previously published version (BG11.0) and dramatically refines the gene annotation. We found that the specific burst of certain subfamilies (CR1 and ERVL) of repetitive elements in duck is the major cause of fragmentation of the old genome. We also identified the putative telomere and centromere regions, and specific CR1 family sequences that are enriched at the centromeres of microchromosomes. Comparison of chromatin architectures between chicken and duck revealed strong natural selection against the inversions that disrupt the topologically associated domains. Sex chromosome analysis indicated that the duck W chromosome is not fully degenerated. In total, we identified 84.5 Mb Z chromosome which including 2.2 Mb pseudoautosomal regions (PAR) and 16.8 Mb W chromosome. We demarcated three times of recombination suppression based on the Z/W sequence similarity, forming a pattern of 'evolutionary strata' along the duck Z chromosome. Overall, the new duck genome provides an important resource to study avian genome evolution and future improvement of domesticated traits.

## **W895: Proteomics**

### **TurboID-Based Proximity Labeling for *in planta* Identification of Protein Interaction Networks**

**S.P. Dinesh-Kumar**<sup>1</sup>, Yuanyuan Li<sup>2</sup>, Ugrappa Nagalakshmi<sup>2</sup> and Yongliang Zhang<sup>3</sup>, (1)Department of Plant Biology and The Genome Center, University of California Davis,, Davis, CA, (2)UC Davis, (3)China Agriculture University, China

Protein-protein interactions (PPIs) play an important role in various biological processes. The yeast two-hybrid assay and affinity purification followed by mass spectrometry (AP-MS) has been routinely used to identify PPIs. However, these methods have some shortcomings especially it fails to capture weak or transient interactors. The recently developed TurboID-based proximity labeling approach overcomes many of the drawbacks of AP-MS and other PL methods. We will discuss comparative analyses of TurboID-based PL compared to other PL methods in plants. In addition, we will present our findings on the identification of novel interactors of a nucleotide-binding leucine-rich repeat (NLR) class of immune receptor that confers immunity to a viral pathogen.

## **W896: Proteomics**

### **Getting to the Edge of Plant-Pathogen Interaction Networks**

**Shahid Mukhtar**, University of Alabama at Birmingham, Birmingham, AL

Networks consist of systems' components, referred to as 'nodes', and interactions between them, termed 'edges'. The understanding of how networks function in the schema of an entire system is fueling the development of network based approaches. Such a framework is pertinent to evaluation of diverse biological networks including protein-protein interactions. Network analysis has been a recent focus in biological sciences due to its ability to synthesize global visualizations of cellular processes and predict functions based on inferences from network properties. Understanding these topological features and deciphering the network architecture can provide insights into the identification of new community structures, unknown signaling pathways, and novel relationships between genes and their products. The work in my laboratory is focused on how macromolecular networks control biological processes and how

environmental perturbations in such networks can explain diverse phenotypes. Here, I'll discuss our two recently generated Arabidopsis interactomes and their network analyses.

### **W897: Proteomics**

#### **Quantitative Proteomics Reveals Tissue-Specific Preferences for a Splicing Regulator, Serine/Arginine-RICH 45, in *Arabidopsis thaliana***

**Xiao-Ning Zhang**, St Bonaventure University, St Bonaventure, NY

In eukaryotic cells, mRNA metabolism is highly regulated to provide a productive pool of mature mRNAs for translation. Alternative splicing is one such regulatory step to increase transcript diversity during mRNA metabolism. Current understanding is limited on how alternative splicing is coordinated with other regulatory mechanisms and ultimately influences the dynamic gene expression spatially and temporally. More importantly, regulations in RNA metabolism are accomplished via complex and dynamic protein-protein interactions and protein-RNA interactions. Our latest findings suggest that the Arabidopsis Serine/Arginine-rich 45 (SR45) functions as a scaffold in splicing regulation. It is associated with RNAs and proteins with diverse functions. The *sr45-1* null mutant exhibits pleiotropic phenotypes in both vegetative and reproductive stages. In order to understand how SR45 influences proteome at different developmental stages, a Tandem Mass Tag-based quantitative mass spectrometry was used to compare the protein abundance from both inflorescence and leaf between Arabidopsis wild-type (Col-0) and *sr45-1* mutant plants. Data analysis yielded 206 inflorescence-specific SR45-dependent differential accumulation (SDA) proteins, 216 leaf-specific SDA proteins and 21 SDA proteins that were common between leaf and inflorescence. PANTHER GO enrichment analysis showed that the inflorescence-specific SDA proteins had a strong preference for chloroplastic and nucleolar components, galactosidase-catalyzed reactions and starch catabolism, while the leaf-specific SDA proteins had a strong preference for non-nuclear& non-chloroplastic components, intracellular redox reactions and ER response. This suggests that SR45 may target different sets of RNAs between inflorescence and leaf to achieve this outcome. Although only a small percentage of these SDA proteins overlapped with differentially expressed or alternatively spliced RNAs from our prior study, evidence suggests that additional SR45-regulated mechanisms, such as DNA methylation, may also contribute to the detected proteomic changes. SR45 is a part of the conserved eukaryotic apoptosis and splicing-associated protein (ASAP) complex. Interestingly, the *sr45-1* mutant inflorescence had dramatically reduced levels of all ASAP components – SR45 (0%), the Sin3-associated protein 18 (SAP18, 13%) and ACINUS (66%). ASAP regulates RNA metabolism at multiple levels. Recent studies have been focused on its repressive function on transcription via SAP18 recruiting histone deacetylase to target loci. However, only a differential accumulation of ACINUS protein (37%) was detected in leaf, but not for SAP18. This is most likely due to the low abundance of SAP18 protein in the leaf. The fact that SR45 affected the protein accumulation, but not the RNA level, of SAP18 and ACINUS suggests that SR45 may regulate translation and/or turnover of these two proteins via additional players. In summary, our findings suggest a possibility that, in addition to pre-mRNA splicing, SR45 utilizes different gene regulatory pathways in a tissue-specific manner to promote growth and reproduction in Arabidopsis.

### **W898: Proteomics**

#### **Evolution of Mitochondrial Proteome in Animals**

Viraj Muthye, Iowa State University and **Dennis V. Lavrov**, Iowa State University, Ames, IA

Mitochondria – membrane-bound organelles present in most eukaryotic organisms – are involved in multiple cellular processes, including oxidative phosphorylation, Fe/S cluster biosynthesis, amino-acid and lipid metabolism, and apoptosis. In humans, these and other functions require more than mitochondrial 1500 proteins, all but a few of which are encoded in the nuclear genome and imported into mitochondria. We investigated the evolution of animal mitochondrial proteome by comparing its composition in species with available experimental data. Furthermore, we used existing and developed new bioinformatics approaches to infer mitochondrial proteins in species for which experimental data were not available. Our analyses revealed a large extent of mitochondrial proteome variation in animals and determined some major evolutionary processes responsible for it. In addition, we identified several

animal species with highly unusual mitochondrial proteomes, which might be interesting targets for experimental research. The new bioinformatics tools created for this project should be useful for future analysis of mitochondrial proteomes.

#### **W899: Proteomics**

##### **Efficacy of Phytochemicals on *Campylobacter jejuni* Biofilms on Common Food Processing Surfaces and their Effect on Proteome of *C. jejuni***

**Abhinav Upadhyay**, UNIVERSITY OF CONNECTICUT, Storrs,, CT and **Basanta Wagle** (Univ of Arkansas), **Indu Upadhyaya** (Univ of Connecticut), **Komala Arsi** (Univ of Arkansas), **Sandip Shrestha** (Univ of Arkansas), **Rohana Liyanage** (Univ of Arkansas), **Dan Donoghue** (Univ of Arkansas), **Annie Donoghue** (USDA-ARS Fayetteville, AR)

*Campylobacter jejuni* is a major foodborne pathogen that causes severe gastroenteritis in humans and is strongly linked with the consumption of contaminated poultry products. Often, *C. jejuni* survives in the processing environment by forming biofilms and recent investigations have highlighted the role of biofilms in the environmental persistence of *Campylobacter* for contaminating poultry products. This study investigated the efficacy of three phytochemicals, *trans*-cinnamaldehyde (TC), eugenol (EG) or carvacrol (CR) in inhibiting or inactivating *C. jejuni* biofilms on common food contact surfaces. All phytochemicals reduced *C. jejuni* biofilm formation as well as inactivated mature biofilm on polystyrene and steel surfaces at 20°C and 37°C (P<0.05). All the phytochemicals downregulated the genes encoding for motility systems (*flaA*, *flaB*, *flgA*). In addition, the expression of stress response (*cosR*, *ahpC*) and cell surface modifying (*waaF*) genes was reduced by 0.01% EG. Additionally, the effect of the aforementioned phytochemicals on the proteome of *C. jejuni* (NCTC 11168) biofilms has been investigated. Proteins were extracted from biofilms and subjected to SDS-PAGE followed by in-gel tryptic digestion and LC-MS/MS based protein quantification. A total of 100 proteins were identified which contribute to cellular and metabolic process, biological regulations and membrane integrity. The expression of 27 proteins was significantly modulated (fold change ~ 4.6 to 20) in the biofilms compared to planktonic cells (P<0.05). The TC, EG and CR significantly downregulated NapA (required for signaling pathway during oxidative stress). Moreover, TC and CR reduced the expression of chaperone protein (DnaK; required for oxidative stress response). The results suggest that a subset of *C. jejuni* proteome changes during biofilm formation, and phytochemicals modulate key proteins contributing to *C. jejuni* biofilm formation and could be used as a natural disinfectant for controlling *C. jejuni* biofilms.

#### **W900: Proteomics**

##### **Fast Neutron Mutagenesis Alters Seed Protein Content in Soybean**

**Savithiry Natarajan**, USDA-ARS, Beltsville, MD, **Nazrul Islam**, USDA, Beltsville, MD, **Robert M. Stupar**, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, **Wesley Garrett**, USDA,ARS, BELTSVILLE, MD, **Hari B. Krishnan**, USDA-ARS-PGRU, Columbia, MO and **Wesley Garrett**, USDA-ARS, Beltsville, MD

To produce value added seed traits, soybean has been subjected to genetic alteration by various approaches. Among the approaches, mutagenesis through fast neutrons (FN) radiation is intriguing because it yields a variety of mutations, including single/multiple gene deletions and/or duplications which alters seed composition. Characterizing the seed composition of the fast neutron mutants and its relationship with gene mutation is useful towards understanding protein traits in soybean. We selected ten mutants based on a screening of total protein content using near infra-red spectroscopy from a large population of fast neutrons mutagenic plants. The mutant 2R29C14Cladecr233cMN15 (nicknamed as L10) showed the highest protein content compared to wild type, followed by three other mutants (L03, L05, and L06). We have physically mapped the position of the deletion or duplications of genes in each mutant using comparative genomic hybridization (CGH). All ten mutant lines had one or more deletions and/or duplications. We selected the L03 mutant for detailed proteomic analysis because it exhibited 55% protein while only showing a homozygous deletion encompassing few genes. We used mass

spectrometry and measured ~3,500 proteins in the seeds and found alterations in a network of proteins that control metabolic pathways, including protein synthesis. The deletion of a transcription factor, along with other several genes, may have altered the negative regulation of protein syntheses processes, resulting in an increase in the overall protein content of the seed. These findings will facilitate the ongoing efforts of scientists and breeders to improve both the quantity and quality of soybean proteins.

## **W901: QTL Cloning**

### **Map-Based Cloning of a QTL for Spikelet Number per Spike in Wheat**

**Jorge Dubcovsky**<sup>1</sup>, Saarah N. Kuzay<sup>1</sup>, Junli Zhang<sup>1</sup>, Yunfeng Xu<sup>2</sup>, Qiujie Liu<sup>1</sup>, Francine Paraiso<sup>1</sup>, Juan M. Debernardi<sup>1</sup>, Chengxia Li<sup>1</sup>, Huiqiong Lin<sup>1</sup>, Eduard Akhunov<sup>3</sup> and Guihua Bai<sup>4</sup>, (1)University of California, Davis, CA, (2)Department of Agronomy, Kansas State University, Manhattan, KS, (3)Department of Plant Pathology, Kansas State University, Manhattan, KS, (4)USDA, Hard Winter Wheat Genetics Research Unit, Manhattan, KS

A better understanding of the genes controlling differences in wheat grain yield can accelerate the improvements required to satisfy future food demands. Spikelet number per spike (SNS) is an important yield component that is determined early in the reproductive phase, resulting in a relatively high heritability. In a genome-wide association study (GWAS) including 262 photoperiod insensitive spring wheat accessions grown under full and limited irrigation, we identified a major QTL for SNS on chromosome arm 7AL and validated it in a biparental population in multiple locations. This SNS QTL was not associated with differences in heading time or plant height and was associated with significant differences in total grain yield. Based on these favorable characteristics, we decided to clone the underlying gene using a map-based cloning approach. We screened > 10,000 segregating plants from heterogeneous inbred families (HIFs) and identified recombination events that delimited an 87-kb candidate gene region between 674,019,191 and 674,106,327 bp on chromosome arm 7AL (CS RefSeq v1.0) including two complete and two partial genes. Exome capture data for this region revealed three major haplotypes that we designated as H1, H2 and H3. The H2 haplotype was associated with high-SNS in both H1 × H2 and H2 × H3 segregating populations, and only one of the four candidate genes, *TraesCS7A02G481600*, showed a non-synonymous polymorphism that differentiated H2 from both H1 and H3 haplotypes (C47F). We designated this gene as *WHEAT ORTHOLOG OF APO1 (WAPO1)* because it is orthologous to rice *ABERRANT PANICLE ORGANIZATION 1 (APO1)*. In rice, the APO1 protein interacts with LFY, and mutants for both genes result in reduced number of spikelets per panicle. Similarly, the wheat WAPO1 and LFY proteins interact with each other in wheat protoplasts and loss-of-function mutations of either genes result in reduced SNS. Using *in situ* hybridization, we observed expression of both genes in the inflorescence and spikelet meristems. The high-SNS allele *Wapo-A1b* (H2 haplotype) showed a rapid increase in frequency from hexaploid wheat landraces (45%) to old cultivars (62%), and from those to more recent North American spring wheat cultivars (83%), suggestive of positive selection for this allele. Taken together, our results point to *WAP0-A1* as the causal gene for the 7AL SNS QTL.

## **W902: QTL Cloning**

### **Novel Insights into Physiological and Molecular Mechanisms of Drought Tolerance in Chickpea**

**Rajeev K Varshney**, ICRISAT, Hyderabad, India

The genetic improvement of drought tolerance in chickpea (*Cicer arietinum* L.) is essential for enhancing its productivity in the semi-arid regions of the world. A genomic region, popularly called the “QTL-hotspot” harbouring 12 QTLs explaining 58.20% of phenotypic variance for drought tolerance related traits was identified on CaLG04 of chickpea. Introgression of the “QTL-hotspot” region into leading chickpea varieties by marker assisted backcrossing enabled identification and release of high yielding drought-tolerant chickpea varieties in India and Ethiopia. Genotyping-by-sequencing followed by bin mapping based QTL analysis delimited this region from ca. 7.74 Mb to a 292.58 kb segment and splitted the “QTL-hotspot” region containing 26 genes into two sub-regions “QTL-hotspot\_a” (139.22 kb, 15 genes) and “QTL-hotspot\_b” (153.36 kb, 11 genes). Sequence analysis of these genes in 3,000

chickpea accessions facilitated us to prioritize five candidate genes associated with yield component traits. Agrobacterium-mediated transformation of *Medicago* roots with these genes demonstrated role of two “QTL-hotspot” genes- *QhGene1* and *QhGene2* in altering root architecture and enhancing drought tolerance. Details about characterization and function of these genes will be presented. In parallel, phenotyping of fine mapping population (50 BC<sub>6</sub>F<sub>4</sub>) on four different phenotyping platforms showed increase in plant vigour due to “QTL-hotspot” region that facilitates better transpiration efficiency, root system architecture and pre-anthesis water extraction, resulting in higher seed yield under drought conditions. In brief, our study provides novel insights into physiological and molecular mechanisms of drought tolerance and highlights the role of “QTL-hotspot” for improving yield in chickpea.

### **W903: QTL Cloning**

#### **Genetic and Molecular Analysis of Rice Grain Yield under High Temperature Stress**

**Harkmal Walia**, University of Nebraska, Lincoln, Lincoln, NE

Rising global temperatures during cropping seasons are resulting in yield losses. These yield losses are emerging as a major obstacle for important cereal crops such as wheat and rice and hence global food security. Rice grain development is very sensitive to high temperatures. Given the heat sensitivity of developing grains, we aimed to explore the natural variation for high temperature tolerance using a suite of phenotyping approaches including an image-based, non-destructive platform for panicle imaging. We have analyzed these data using image analysis pipelines for 3D-feature extraction. To associate genomic linkages with these image-derived features, we have performed genome-wide association analysis (GWAS) to identify loci for high temperature tolerance that explain the natural variation within the rice germplasm. We will present results from this genetic analysis at panicle level over spatial and temporal scale in response to high temperatures.

### **W904: QTL Cloning**

#### **Map-Based Cloning of a Major QTL Associated with Waterlogging Tolerance in Soybean Involved in Regulation of Root System Architecture**

**Heng Ye**<sup>1</sup>, Li Song<sup>1</sup>, Lijuan Zhou<sup>1</sup>, Liakat Ali<sup>2</sup>, Chengjun Wu<sup>3</sup>, Leandro A. Mozzoni<sup>3</sup>, William Schapaugh<sup>4</sup>, Pengyin Chen<sup>2</sup> and Henry T. Nguyen<sup>1</sup>, (1)University of Missouri, Columbia, MO, (2)University of Missouri Fisher Delta Research Center, Portageville, MO, (3)University of Arkansas, Fayetteville, AR, (4)Kansas State University, Manhattan, KS

Continuous climate changes are making flood-caused waterlogging stress more severe with a predicted 30% increase in heavy precipitations by 2030. Waterlogging is also confounded by an increased application of irrigation that could result in excess water. In our group, soybean is used as a model crop to address the urgent demands of improving waterlogging tolerance and understanding the underlined molecular mechanism in plants. A quantitative trait locus (QTL) on chromosome 3 was found to be associated with waterlogging (soil-flooding) tolerance with donor allele from an exotic line (PI 561271). This QTL was isolated into near-isogenic background and confirmed to regulate waterlogging tolerance through modification of root length density during the stress. The QTL underlying gene, *WLT1*, was map-based cloned and transgenic analysis confirmed *WLT1* was responsible for this natural variation in waterlogging tolerance. The favorable allele of *WLT1* was found to have an insertion of a R-motif in the 5'-untranslated-region leading to suppression of its own translation, which indicated the involvement of R-motif regulated protein translation in plant adaptation to abiotic stress for the first time. *WLT1* was further identified to regulate auxin levels in pericycle cells to control root plasticity during the waterlogging stress. Introducing *WLT1* from the exotic parent into the elite parent enabled the resulting lines to yield significantly more relative to the near-isogenic lines without the gene introgression under waterlogging, drought or low-phosphorus field conditions. These findings suggest that auxin-regulated RSA is essential for improving waterlogging and drought tolerance and nutrient uptake in dryland crops.

### **W905: QTL Cloning**

## **A Receptor-like Kinase Enhances Sunflower Resistance to *Orobanche cumana* by Preventing Plant-Plant Interaction**

**Pauline Duriez**<sup>1</sup>, Sonia Vautrin<sup>2</sup>, Marie-Christine Auriac<sup>3</sup>, Julia Bazerque<sup>3</sup>, Marie-Claude Boniface<sup>4</sup>, Caroline Callot<sup>2</sup>, Sébastien Carrère<sup>3</sup>, Stéphane Cauet<sup>5</sup>, Mireille Chabaud<sup>3</sup>, Fabienne Gentou<sup>6</sup>, Marta Lopez-Sendon<sup>6</sup>, Clémence Paris<sup>6</sup>, Prune Pegot-Espagnet<sup>3</sup>, Jean-Christophe Rousseaux<sup>6</sup>, Begoña Pérez-Vich<sup>7</sup>, Leonardo Velasco<sup>7</sup>, Hélène Bergès<sup>2</sup>, Joël Piquemal<sup>6</sup> and Stéphane Muñoz<sup>4</sup>, (1)UBC, Vancouver, BC, Canada, (2)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (3)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (4)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRAE/CNRS/Université de Toulouse, Castanet-Tolosan, France, (5)French Plant Genomic Center CNRGV - INRAE, Castanet-Tolosan, France, (6)Syngenta, Saint-Sauveur, France, (7)Instituto de Agricultura Sostenible - CSIC, Córdoba, Spain

*Orobanche Cumana* is a non-chlorophyllous plant parasiting sunflower roots. Resistant sunflower cultivars play an important role for the control of *O. cumana* in fields. Some of them carry the *HaOr7* locus, a major resistance gene which confers resistance to *O. cumana* race F, the main race in South of Spain. We used a high-throughput genotyping tool to map the locus to the chromosome 7 by GWAs on 55 lines and by a segregating population of 355 RILs. *In silico* analyses and genetic data identified a genomic region on the chromosome 7 of 852 kb containing the *HaOr7* gene. We then restricted the genomic region of the *HaOr7* gene by genotyping a large segregating population of 14,281 F2 individuals, identifying 271 F2 recombinant plants between the two markers surrounding the 852kb genomic region. After selfing each of the 271 F2 plants, almost all F3 and F4 families were phenotyped in field. Finally the *HaOr7* locus mapped in a window of around 55kb, containing one full coding sequence, predicted to code for Leucine Rich Repeat receptor-like kinase protein. We performed a molecular diversity analysis of the region, in a panel of 170 wild relatives, wild and cultivated sunflower accessions. The analysis showed only one haplotype on *HaOr7+* lines on this candidate gene, while all *HaOr7-* lines carried a stop codon leading to a truncated protein lacking the intracellular kinase domain which might explain the susceptibility. To obtain the parental genomic sequences, we created and screened two BAC libraries from susceptible and resistant lines. The genomic sequences show polymorphisms with large structural variation, suggesting a wild origin of *HaOr7*. Furthermore, physiological and microscopic analyses showed *HaOr7* acts at early stages of the interaction, preventing the connection to the sunflower vascular system. Finally, the genetic, molecular and physiological characterizations allowed us to suggest a model explaining the resistance but also the susceptibility.

## **W906: QTL Cloning**

### **Identification of Candidate Resistance Genes Against Blackleg in *Brassica napus***

**Soodeh Tirnaz**, Yueqi Zhang, Ting Xiang Neik, Anita Severn-Ellis, Philipp E. Bayer, David Edwards and Jacqueline Batley, University of Western Australia, Perth, WA, Australia

The *Brassica* genus contains the greatest diversity of agriculturally important species. There has been a rapid increase in the global *Brassica napus* (canola/oilseed rape) production over the last four decades. Blackleg disease caused by the fungal pathogen *Leptosphaeria maculans* is one of the most devastating diseases of canola in Australia, Europe and America. Identification of resistance genes in *Brassicaceae* and breeding resistant lines is the most effective and sustainable approach to combatting the disease. Seventeen resistance genes that have a gene-for-gene interaction with *L. maculans* avirulence genes have been genetically mapped while only three resistance genes have been cloned. *Rlm4* and *Rlm7*, located in a QTL region on Chromosome A07 specifically interact with the avirulence gene *AvrLm4-7*, suggesting that *Rlm4* and *Rlm7* may be allelic variants of the same gene. In this research, genome wide identification of resistance genes was performed in different *Brassica napus* cultivars and resistance genes within the QTL region for *Rlm4-7* were characterised. Sequence comparison of resistance genes amplified from over 100 *Rlm4-7* susceptible and resistant lines were analysed to identify allelic diversity and for further understanding of the gene for gene interaction at this locus, leading to identification of the candidate gene. These causative *Rlm4-7* alleles will facilitate the breeding for *Rlm4-7* lines for managing blackleg disease in canola.

## **W907: Quinoa and close relatives**

### **An Improved Genome Assembly for the Highland Ecotype of Quinoa**

**Zhang Heng**, Shanghai Center for Plant Stress Biology, CAS, Shanghai, China

We previously reported a high-quality genome draft, termed Cq\_real\_v1, for an accession of the highland ecotype of quinoa (Zou et al. 2017). Here we updated the assembly using a combination of high-depth single-molecule sequencing, a high-density genetic map and high-quality optical mapping data. The resulting Cq\_real\_v2 assembly exhibit significantly improved synteny towards related species. We demonstrate that this assembly can be used for the mapping of genetic loci underlying important agronomic traits.

## **W908: Quinoa and close relatives**

### **Developing an EMS Quinoa Population**

**Brian Cox**, Brigham Young University, Provo, UT

Quinoa is a pseudocereal commonly grown in the Andean highlands of South America, but attempts to produce quinoa in other regions around the world have met varying degrees of success. Traits such as low heat tolerance, downy mildew susceptibility, low harvest index, and high seed-saponin levels need improvement in order to further the production of quinoa outside of its natural environment. This project focuses on understanding the effects of mutagenized genes on quinoa phenotypes. Following a forward genetics procedure, quinoa seeds were soaked in 2% EMS, a mutagen known for inducing point mutations, before being planted. Further generations were bred through self-fertilization, and new phenotypes such as variegated leaves and more complex branching patterns were later observed. DNA from families of interest was analyzed by sequencing and identifying novel SNPs. The expected EMS mutation rate was recorded, and various quinoa families presented mutations that may have had a significant effect on their phenotypes. Additional whole-genome sequencing of select mutant lines was performed to identify candidate mutations that caused their mutant phenotypes.

## **W909: Quinoa and close relatives**

### **Ancient Chloroplast and Nuclear Genomes Provide Insights into the Evolutionary History of Quinoa (*Chenopodium quinoa* Willd.)**

**Oscar Estrada**<sup>1</sup>, James Breen<sup>2</sup>, Kieren J. Mitchell<sup>3</sup>, Pilar Babot<sup>4</sup>, Salomon Hocsman<sup>4</sup>, David E Jarvis<sup>5</sup>, Mark A. Tester<sup>6</sup> and Alan Cooper<sup>7</sup>, (1)The University of Adelaide, Adelaide, SA, Australia, (2)Bioinformatics Hub, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia, (3)The University of Adelaide, Australia, (4)National University of Tucuman, Argentina, (5)Brigham Young University, Provo, UT, (6)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, (7)The University of Adelaide, Adelaide, Australia

Quinoa (*Chenopodium quinoa*), an important crop for the global food security, was domesticated in the Andean highlands of South America about 7,000 years ago, where wild relatives still grow today. Archaeological studies have reported quinoa remains dating back to 4,000 years before present, but the ancient genetic diversity remains uncharacterised. Recent sequencing studies of the quinoa nuclear genome suggested domestication took place independently in highland and coastal environments. However, the origin and exact timing of domestication, and the genomic relationships among the tetraploid relatives (*C. hircinum* and *C. berlandieri*) of quinoa are not fully understood. Here, we report the nuclear and chloroplast genome sequences of four ~1,400 years old archaeological samples of *C. quinoa* from the highlands of northern Argentina. Phylogenetic analyses show that wild, weedy and cultivated accessions from different species cluster in a monophyletic clade, indicative of historical hybridisation of closely related tetraploid species of *Chenopodium*. Two of the identified chloroplast haplogroups in the ancient samples appear to be absent from modern quinoa accessions, indicating a genetic bottleneck in the recent past. Further, sequence alignment to the nuclear and chloroplast reference genomes of *C. quinoa* identified a large number of nucleotide polymorphisms in the ancient

specimens, providing resources for further analysis of domestication history as well as genetic changes in loci of breeding value.

### **W910: Quinoa and close relatives**

#### **Cytonuclear Evolution in Allopolyploid Quinoa**

**Joel Sharbrough**, Colorado State University, Fort Collins, CO

Whole-genome duplications (WGDs), in which the number of nuclear genome copies is elevated as a result of autopolyploidy or allopolyploidy, are a prominent process of diversification in eukaryotes. The genetic and evolutionary forces that WGD imposes upon cytoplasmic genomes are not well understood, despite the central role that cytonuclear interactions play in eukaryotic function and fitness. In particular, cellular respiration and photosynthesis depend upon successful interaction between the 3000+ nuclear-encoded proteins destined for the mitochondria or plastids and the gene products of cytoplasmic genomes in multi-subunit complexes such as Rubisco, OXPHOS, Photosystems I and II, and organellar ribosomes. Allopolyploids are thus faced with the critically important task of maintaining successful interactions and coordinated gene expression between nuclear and cytoplasmic genes that were inherited from different species. Because maternal homoeologs are expected to be more closely "matched" to cytoplasmic genomes than are paternal homoeologs, incompatibilities between organelle genomes and paternal subgenomes of allopolyploids may lead to relaxed selection on paternal vs. maternal homoeologs of genes targeted to the mitochondria or plastids. To test this hypothesis, we compared rates of molecular evolution in maternal vs. paternal homoeologs of organelle-targeted genes in the allotetraploid *Chenopodium quinoa* (Amaranthaceae). This global assessment of cytonuclear coevolution in diploid vs. polyploid angiosperms provides insights into the molecular dynamics of cytonuclear incompatibilities that are likely to influence the success and evolution of hybrid polyploids.

### **W911: Quinoa and close relatives**

#### ***C. berlandieri* Genome**

**Thomas Jordan Kerby**, Brigham Young University, Provo, UT

*Chenopodium berlandieri* is a close relative of *C. quinoa* and is found throughout North America. *C. berlandieri* has adapted to biotic and abiotic conditions that adversely affect quinoa growth, making it an ideal target for the identification of genetic variation that can be used to improve the growth of quinoa outside its native habitat. Here we report the first high-quality reference genome assembly and annotation for *C. berlandieri* *supsp. nuttaliae*, a cultivated Mexican variety commonly known as huauzontle. The assembly was produced using PacBio data and was polished using Arrow and Pilon. The assembly was scaffolded using in vivo Hi-C into 1,194 scaffolds spanning 1.3 Gb. The scaffold N50 is 4.9 Mb. 19.28% of the assembly is in the largest 18 scaffolds, corresponding to the haploid number of chromosomes in *C. berlandieri*. We report genome annotation using RNA-seq data from ### tissues and make a genomic comparison between the genomes of *C. berlandieri* and quinoa. This comparison sheds light on potential genetic variation in *Chenopodium* species that can be used to improve its capacity to be cultivated in lower elevations and higher temperatures.

### **W912: Quinoa and close relatives**

#### **Deciphering/Comparing Genomes of a few Landraces of Grain Amaranths from India**

**Subhashini Srinivasan**, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Grain amaranths are species producing nutritious grains with especially high lysine, which is one of the major limiting amino acid among the malnourished population of underprivileged world. India is one of a few countries that has enjoyed cultivation of this crop for more than a few centuries since after the Columbian Exchange through a ban on these crops for cultivation in the West. The cultivation of all three grain amaranths species in India for the last several centuries has resulted in many landraces that are optimized for high yield and diverse environmental conditions. It is of interest to study their genomes and

compare them with other accessions in order to introduce desirable traits via breeding. Here, we present the genomes of a few landraces and compare them to existing accessions.

### **W913: Resources and Programs for Undergraduate Education in Genomics**

#### **Introductory Remarks**

**Scott Woody**, UW-Madison, Madison, WI

Session overview and speaker introductions

### **W914: Resources and Programs for Undergraduate Education in Genomics**

#### **The Genomics Education Alliance (GEA)**

**Vincent Buonaccorsi**, Juniata College, Huntingdon, PA, Anne Rosenwald, Georgetown University, Washington, DC, Rochelle Tractenberg, Georgetown University, Jason Williams, CyVerse, Tucson, AZ and Douglas L. Chalker, Washington University in St Louis, St Louis, MO

The Genomics Education Alliance represents a group of life science educators who have experience engaging students in Course-based Undergraduate Research Experiences (CUREs) in genomics and bioinformatics. Because we are convinced that CUREs are effective for students to learn both key concepts and the practice of science, we have come together to identify and overcome common barriers to put such experiences within the reach of all life science faculty and students. To achieve these goals, the GEA will: 1) host core bioinformatics tools, 2) curate and/or develop curriculum and faculty training resources, and 3) curate CURE assessment materials. We plan to curate a wide variety of freely available materials both from our existing genomics CUREs and new resources we create. The GEA will utilize the cyberinfrastructure provided by CyVerse to ensure sufficient compute capacity for faculty to use GEA resources in the classroom. Ultimately, we aim to facilitate efforts by faculty who build their own genomics CURE using our optimized resources. We are now recruiting faculty to pilot a set of stand-alone lessons meant to support CUREs in three areas: lessons on examining gene sequence similarities using BLAST, understanding eukaryotic gene structure by using a genome browser, and investigating gene expression by using basic tools for RNA-seq analysis. To learn more about these lessons and to sign up for this pilot, please visit the GEA web site at <https://gea.qubeshub.org/lessons>. The GEA is supported by National Science Foundation Research Coordination Network for Undergraduate Biology Education (NSF RCN-UBE) grant #DBI 1827130.

### **W915: Resources and Programs for Undergraduate Education in Genomics**

#### **Using Genome Browsers Constructed by G-OnRamp to Provide Students with a Course-Based Undergraduate Research Experience in Genome Annotation**

**Wilson Leung**<sup>1</sup>, Luke Sargent<sup>2</sup>, Yating Liu<sup>1</sup>, Nathan T. Mortimer<sup>3</sup>, David Lopatto<sup>4</sup>, Jeremy Goecks<sup>2</sup> and Sarah C. R. Elgin<sup>1</sup>, (1)Washington University in St. Louis, St. Louis, MO, (2)Oregon Health & Science University, Portland, OR, (3)Illinois State University, Normal, IL, (4)Grinnell College, Grinnell, IA

High-throughput sequencing has enabled researchers to sequence many eukaryotic genomes and to leverage comparative genomics in their scientific investigations (e.g., the Earth BioGenome Project). However, high quality annotations of protein-coding genes and other genomic features are a prerequisite for using these sequenced genomes in a variety of scientific investigations (e.g., differential expression analyses). Despite advances in computational algorithms, annotation of most eukaryotic genomes still benefits from manual curations. Course-based Undergraduate Research Experiences (CUREs) focused on genome annotation provide a cost-effective way for researchers to obtain high quality gene models, for educators to introduce students to big data and eukaryotic genes/genomes, and for students to benefit from the research experience. Genome browsers provide annotators with a visualization platform for synthesizing the evidence provided by sequence alignments, RNA-Seq data, and gene predictions to construct gene models best supported by the available evidence. To enable educators to more easily integrate genome annotation into their curriculum, the Genomics Education Partnership (GEP);

<http://gеп.wustl.edu>) and the Galaxy Project (<https://galaxyproject.org>) have developed G-OnRamp (<http://g-onramp.org>), a web-based platform for constructing UCSC Assembly Hubs and JBrowse/Apollo genome browsers for collaborative genome annotations in research and education settings. The genome browsers produced by G-OnRamp have been used to engage 15 faculty and >200 students in CUREs focused on annotation of four parasitoid wasp species. Knowledge and attitude assessments of students who have participated show gains similar to those obtained by GEP students as a whole. Faculty interested in using G-OnRamp to develop a CURE can contact us at [http://gеп.wustl.edu/contact\\_us](http://gеп.wustl.edu/contact_us).

## **W916: Resources and Programs for Undergraduate Education in Genomics**

### **Tools and Approaches for Making Bioinformatics Work in the Classroom**

**Jason Williams**, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Genomics and other areas of biology rely heavily on the tools of bioinformatics, data science, and computing. Bringing these tools into the classroom can be daunting – both in determining which tools to use and managing their use. We will cover several important technologies that can make using bioinformatics in the classroom a more manageable experience.

CyVerse has developed infrastructure for delivering bioinformatics infrastructure including platforms for data storage and analysis. Through *DNA Subway* (<https://dnasubway.cyverse.org/>) we have made workflows available for genome annotation, DNA barcoding, RNA-Seq, and metabarcoding/microbiome broadly accessible. Our VICE applications also provide an easy interface to Jupyter notebooks and RStudio.

The Genomics Education Alliance (GEA) is a newly NSF-funded research collaboration network. This alliance has brought together several well-known educators and educational projects. Resources in development by GEA include a core set of computational tools and we will demonstrate how we have used CyVerse resources to make classroom workflows. We will review the benefits and value propositions of technologies such as Docker, Jupyter, and CyVerse and provide tips on how these and other resources can be adopted and adapted for faculty in any context.

## **W917: Resources and Programs for Undergraduate Education in Genomics**

### **Undergraduates Contain Multitudes: Course-Based Metagenomics Analysis using the DNA Subway Purple Line**

**Ray Enke**, James Madison University, Harrisonburg, VA

Undergraduate students learn about Next Generation Sequencing (NGS) technology in courses, but often have difficulty understanding the impact of these techniques without hands-on experience analyzing actual NGS data. Here I describe a classroom-tested set of course-embedded activities focusing on metagenomics analysis of microbial diversity in biological samples tailored for implementation into diverse undergraduate classroom settings. These modular workflows can be applied to a variety of novel or publicly available 16S microbial metagenome data sets. Bioinformatics modules feature QIIME 2 analysis implemented in the recently developed DNA Subway Purple Line, a user friendly web-based suite of tools designed for students and educators with a novice levels of experience in genomics analysis. Downloadable classroom modules and lesson plans for these activities are publicly available to educators ([https://works.bepress.com/raymond\\_enke/](https://works.bepress.com/raymond_enke/)). Additionally, an NSF-funded workshop will be hosted this upcoming June 2020 in Brooklyn, NY for educators interested in implementing DNA Metabarcoding course materials.

## **W918: Resources and Programs for Undergraduate Education in Genomics**

### **Engaging Students in Gene Mapping using GameteMaker**

**Ed Himelblau**, California Polytechnic State University, San Luis Obispo, San Luis Obispo, CA

Mendelian genetics, meiosis, linkage, and the application of molecular tools (such as PCR, gel electrophoresis) are core concepts in undergraduate genetics. Gene mapping incorporates all of these core ideas and can serve as an anchoring experience for genetics students. In particular, mapping reinforces genetic linkage which underpins almost all approaches to gene discovery in both plant and animal model systems, from GWAS and QTL analyses to map-based cloning of mutant alleles. Mapping encourages genome-wide thinking as students begin by analyzing markers spread across the genome, then narrow their search to a decreasing chromosomal interval, then to a candidate locus, and finally to a causative mutation. In our research programs, we have seen first hand how engaged students become as they “hunt for the gene” while simultaneously reinforcing their understanding of core genetics and molecular biology concepts. Wet lab gene mapping projects may be possible for small numbers of students in independent research programs but are difficult to implement for a class. We have developed GameteMaker ([sim.fpscgenetics.org](http://sim.fpscgenetics.org)), a virtual platform that enables large groups of students to benefit from the experience of gene mapping. Students select an organism and mutant phenotype for mapping, raise an F<sub>2</sub> mapping population, and ‘click’ individuals for DNA extraction. Students select molecular (microsatellite) markers, calculate PCR master mix conditions, and analyze a virtual gel to determine the genotype of each member of their population at that locus. Students test for linkage between the marker and the variant using the chi-square test. Upon discovery of statistically significant evidence of linkage, students perform additional marker assays and three-point cross analyses to identify a small genetic interval that must include the causative locus. At that point, students can seamlessly transition to a genome browser that presents gene models and corresponding functional annotations, as well as short-read DNA sequence data from real or simulated next-generation sequencing experiments. GameteMaker is freely available to genetics instructors and their students.

### **W919: Resources and Programs for Undergraduate Education in Genomics**

#### **Using NGS Technology to Sequence Amplicons *en masse* to Solve Two Problems: Cost and Heterozygosity**

**Matthew Collin**, University of California Riverside, Riverside, CA

Expanding authentic research experiences to large undergraduate laboratory courses can be logistically challenging and expensive. In Biology 20: The Dynamic Genome course, at the University of California, Riverside, up to 9 sections per quarter of 24 first-year students participate in guided research projects. As part of the course students extract, quantify, and amplify DNA to analyze polymorphisms. An exciting part of the class is when students analyze their own sequence data. However, Sanger sequencing DNA amplicons for up to 216 students is very costly. Especially if each student has multiple amplicons or the amplicons require reads in both forward and reverse directions to obtain useful data. Additionally, some projects involve sequencing DNA from polyploid or heterozygous organisms which impacts the quantity of Sanger sequencing. Here we present an *en masse* sequencing methodology using the Pacific Biosciences RS II. From one RS II SMRTcell, > 10,000 circularized consensus sequences are produced each representing a different amplicon, thus obviating the need for bacterial libraries. Primary amplification is done with M13 tagged primers. Subsequently, individual amplicons are re-amplified to with barcoded primers and all sections are pooled for library construction. The resulting sequence is disaggregated with custom software. Utilizing, Pacific Biosciences amplicon sequencing in combination with barcodes enables the course to obtain high quality DNA sequences regardless of ploidy, heterozygosity or amplicon length at scale and a reasonable cost.

### **W920: Rhinoceros Genomics Studies: tools for conservation**

#### **Genetic Analysis and Marker Development in the Sumatran Rhinoceros (*Dicerorhinus sumatrensis*)**

**Alfred L. Roca**, University of Illinois at Urbana-Champaign, Urbana, IL

The Sumatran rhinoceros is critically endangered, with fewer than 100 individuals surviving across its current range. Accurate census estimates of the remaining populations are essential for development and implementation of conservation plans. In order to enable molecular censusing, we developed microsatellite markers with amplicon sizes of short length, appropriate for non-invasive fecal sampling. A

bioinformatics routine identified Illumina MiSeq genomic reads with short tandem repeats, and screened for loci that were polymorphic within the dataset. Twenty-nine novel polymorphic microsatellite markers were characterized ( $A = 2.4$ ;  $HO = 0.30$ ). These were sufficient to distinguish among individuals ( $PID < 0.0001$ ), and to distinguish among siblings ( $PID(sib) < 0.0001$ ). Markers were tested using samples in Indonesia, including fecal samples collected from wild rhinoceros in Sumatra. A subset of markers was established as polymorphic and effective for genotyping DNA from fecal samples of wild rhinoceros

To aid in conservation planning, we sequenced 218 bp of control region mitochondrial (mt) DNA, identifying 17 distinct mitochondrial haplotypes across modern ( $N = 13$ ) and museum ( $N = 26$ ) samples. Museum specimens from Laos and Myanmar had divergent mtDNA, consistent with the placement of western mainland rhinos into the distinct subspecies *D. s. lasiotis* (presumed extinct). Haplotypes from Bornean rhinos were highly diverse, but dissimilar from those of other regions, supporting the distinctiveness of the subspecies *D. s. harrissoni*. Rhinos from Sumatra and Peninsular Malaysia shared mtDNA haplotypes, consistent with their traditional placement into a single subspecies *D. s. sumatrensis*. Modern samples of *D. s. sumatrensis* were genotyped at 18 microsatellite loci. Rhinos within Sumatra formed 2 sub-populations, likely separated by the Barisan Mountains, though with only modest genetic differentiation between them. There are so few remaining Sumatran rhinoceros that separate management strategies for subspecies or subpopulations may not be viable, while each surviving rhino pedigree is likely to retain unique alleles. Because rapid genetic erosion is inevitable, along with the potential for fixation of harmful genetic variants, we underscore two overriding priorities for the species: 1) translocation of wild rhinos to ex situ facilities, and 2) collection and storage of gametes and cell lines from every surviving captive and wild individual. (Additional co-authors will be listed in the talk or poster.)

#### **W921: Rhinoceros Genomics Studies: tools for conservation**

##### **Conservation Genomics of the Sumatran Rhinoceros**

**Love Dalén**, Swedish Museum of Natural History, Stockholm, Sweden

The Sumatran rhinoceros (*Dicerorhinus sumatrensis*) used to be widespread over Southeast Asia but has experienced a severe decline as a result of habitat destruction and poaching. Concerns over inbreeding depression have prompted calls for the remaining populations to be managed as a single unit. To investigate the genetic threats facing this species, we have sequenced 21 Sumatran rhinoceros genomes from Borneo, Sumatra and the Malay Peninsula. We examined genome-wide diversity, inbreeding and deleterious mutations to assess the genomic impacts of recent population declines in these populations. The results suggest a deep divergence between Borneo and other populations, but surprisingly low levels of inbreeding and mutational load within populations. Moreover, most of the deleterious mutations identified were not shared among populations. These results have important implications for the current conservation management of the species.

#### **W922: Rhinoceros Genomics Studies: tools for conservation**

##### **Deleterious Genetic Variation and the Prospects for Genetic Restoration of White Rhinos**

**Aryn Wilder**, San Diego Zoo Institute for Conservation Research, Escondido, CA

#### **W923: Rhinoceros Genomics Studies: tools for conservation**

##### **Omics Characterization of Northern White Rhinoceros Induced Pluripotent Stem Cells**

**Iñigo Valiente-Alandi**, San Diego Zoo Institute for Conservation Research, Escondido, CA

Earth is currently facing the sixth mass extinction, predicted to be the most devastating that we have ever experienced. In many cases, standard conservation approaches are not enough to save critically endangered species. However, new hope may come with the advancement of stem cell-based methodologies. We have pioneered the development of the functional extinct northern white rhinoceros (NWR) species induced pluripotent stem cells (iPSCs) to be utilized in gamete generation and assisted

reproduction with the goal of maintaining population viability and genetic diversity, and avoiding extinction.

We have performed transcriptomic analysis on NWR fibroblasts and iPSCs to evaluate the pluripotency landscape in NWR iPSCs. We found over 5,000 differentially expressed genes (DEGs) between fibroblasts and iPSCs populations. Principal component analysis and hierarchical clustering of DEGs showed distinct separation of the fibroblast and iPSCs compartments. In accord with the RNA-Seq data, POU5F1, SOX2, NANOG, LIN28 and DNMT3b, key pluripotency markers in other species such as humans and mice, were top DEGs in rhino iPSCs. Gene ontology analysis of DEGs confirmed the nature of both populations with ion transport and cell-cell adhesion terms characterizing iPSCs and collagen fibril organization and cell migration biological terms defining NWR fibroblasts. We have also performed comparative transcriptomics of the pluripotency state on iPSCs from different species, including the NWR, and studied the transcriptomic trends of the differentiation potential of NWR iPSCs using unsupervised hierarchical clustering approaches.

To examine the energetic requirements of NWR iPSCs we studied the metabolome of these cells and compared it to human iPSCs. We have found similarities between both species in terms of mitochondria viability, principal source of energy (breakdown of glucose) and ATP production.

The molecular characterization of NWR iPSCs is being utilized to elucidate the mechanisms underpinning the pluripotency of these cell lines as they represent the source for the generation of the primordial germ cells that hold the potential for *in vitro* germ cell maturation, fertilization and embryo transfer to surrogates, generating a self-sustaining population of NWR.

#### **W924: Rice Functional Genomics**

##### **Identification of Critical Amino Acid Residues in *Ptr* mediated Plant Innate Immunity**

**Yulin Jia**, USDA-ARS, Stuttgart, AR, Haijun Zhao, USDA ARS, 130, AR, Melissa Jia, USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR and Yan Liu, University of Arkansas Rice Research and Extension Center, Stuttgart, AR

The *Ptr* gene in cultivated rice varieties confers resistance to a wide range of races of *Magnaporthe oryzae*, which causes rice blast disease, except for US pathotypes IE-1k and IB33, the latter being one of the most virulent blast races. *Ptr* was identified approximately 200 kb from the *Pi-ta* locus encoding a protein with 4 armadillo repeats. Here we report that an allelic variant of *Ptr* with minor amino acid variation in a black hull, awned weedy rice strain is responsible for resistance to IB33. One black hull, awned weedy rice x cultivated rice mapping population of 200 individuals was used for mapping with IB33 under greenhouse conditions. The resistance factor to IB33 was mapped at the *Ptr* locus between single nucleotide polymorphic marker (SNP) 10.633,942bp and 10.820,033 bp with the closest SNP at 10.724,430 bp excluding both *Pi-ta* and *Pi-ta2* (another NLR protein). We then developed a gene specific marker for a portion of the *Ptr* gene and examined the existence of the *Ptr* gene in each individual of the mapping population along with their reactions to IB33. The presence of the gene specific marker of *Ptr* observed in individuals resistant to IB33 suggests that the haplotype of *Ptr* in black hulled awned weedy rice is responsible for resistance to IB33. This finding helps in the identification of critical amino acid residues of the *Ptr* protein for detecting the pathogen signal in triggering effective plant innate immunity. Ideas on how to utilize both *Pi-ta* and *Ptr* in preventing rice blast disease will be presented.

#### **W925: Rice Functional Genomics**

##### **Independent Gene Selection for Rice Awn Development in Asia and Africa**

**Kanako Bessho-Uehara**, Department of Plant Biology, Carnegie Institution, Stanford, CA

Human has selected various mutations caused loss of function contributing numerous beneficial traits for agriculture. The identified genes related to agricultural traits are common in cultivated rice species in Asia (*Oryza sativa*) and Africa (*Oryza glaberrima*) such as *Sh1* (seed shattering) and *Rc* (pericarp color).

On the other hand, awn which is long needle-like structure formed at the tip of lemma is a trait that has been lost in both cultivated species, caused by the selection of different sets of genes. We identified three loci named *REGULATOR OF AWN ELONGATION 1* (*RAE1*), *RAE2*, and *RAE3* regulating awn development in rice. Analysis of *RAEs* sequence diversity identified a deleterious, independent mutations underlying low expression or frame shift that disrupt the function of *RAEs* protein. Sequence comparison using diverse rice species and genetic approach revealed cultivated Asian rice keep dysfunctional *RAE1* and *RAE2* but functional *RAE3* allele. In contrast, cultivated African rice retained the functional *RAE1* and *RAE2* allele despite its awnless phenotype. Our findings illuminate the molecular function of three *RAE* genes in awn development and shed light on the independent domestication histories of Asian and African cultivated rice.

## **W926: Rice Functional Genomics**

### **Rice Transcription Factor Binding Atlas**

**Taiji Kawakatsu**, Institute of Agrobiological Sciences, NARO, Tsukuba, Japan

Gene expression regulation plays central roles in all biological processes, such as development and response to environmental signals. Recognition of *cis*-elements by transcription factors (TFs) fundamentally shapes the gene regulatory networks. Comprehensive collection of TFs and their binding sites is critical for understanding of gene regulatory network in an organism. Here, we present *in vitro* TF binding atlas in rice, using DNA affinity purification sequencing (DAP-seq). We transferred >1000 rice TF cDNAs from Gateway entry clones to pIX-Halo vector. DAP-seq identified *in vitro* binding sites of hundreds of TFs. We also identified distinct recognition motives between closely related TFs. Collectively, we provide the novel resource for deciphering gene regulatory network in rice.

## **W927: Rice Functional Genomics**

### **Diversity and Function of Triterpenoid Pathways in Rice**

**Xiaoquan Qi**, Institute of Botany, Chinese Academy of Sciences, China, Beijing, China

Plants produce a large array of steroid and triterpenoid compounds via isoprenoid pathways. Cyclisation of 2,3-oxidosqualene to diversified sterol and triterpenoid skeletons are conducted by members of the oxidosqualene cyclase (OSC) family. We have identified 11 OSC gene homologues in the rice (*Oryza sativa* L.) genome and 7 of them encoding functional OSCs that produce different triterpenes, leading specific triterpenoid pathways. Interestingly, *OsOSC12/OsPTS1*, a key gene required for the poaceatapelol triterpene biosynthesis, is a grass conserved triterpene synthase. Mutants in this gene led to deficiency in formation of pollen coat. The mutant pollen grains overly dehydrate and rapidly lose viability at ambient or low relative humidity (RH<60%), and its fertility is restored at a high RH (>80%). The mutants' pollen coat lacks three major fatty acids and dehydrates rapidly. Mixtures of linolenic acid and palmitic acid or stearic acid are sufficient to prevent mutant pollen grains to over-dehydration. The humidity-sensitive genic male sterility (HGMS) encoded by *OsOSC12/OsPTS1* has potential for the production of hybrid varieties in rice and other grass.

## **W928: Rice Functional Genomics**

### **An Asymmetric Genome Interaction Drives Allele Transmission Bias in Interspecific Rice Hybrids**

**Yongyao Xie, Jintao Tang, Xianrong Xie, Jianle Huang, Yue Fei, Yao-Guang Liu and Letian Chen**, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, China

Hybrids between *Oryza sativa* (Asian rice) and *O. glaberrima* (African rice) show strong heterosis, but suffer from hybrid sterility (HS) that is mainly controlled by the *S1* locus. Three closely linked genes (*S1A4*, *S1TPR*, and *S1A6*) in the African *S1* allele (*S1-g*) constitute a killer-protector system that eliminates gametes carrying the Asian allele (*S1-s*). In Asian-African hybrids (*S1-g/S1-s*), the *S1A4-S1TPR-S1A6* interaction in sporophytic tissues causes male and female gamete abortion. *S1TPR*

rescues the gametes carrying *S1-g*, but gametes carrying *S1-s* lack S1TPR are aborted. Knockout of any of the *S1-g* genes eliminates the HS. Evolutionary analysis suggests that *S1* arose from newly evolved genes, multi-step recombination, and nucleotide variations, which might have aided speciation in *Oryza*. This research reveals how *S1* causes HS in African-Asian rice hybrids and provides a method to produce neutral *S1-g* alleles and thus bypass the interspecific reproductive barrier for creation of improved rice hybrids.

**Keywords:** Asian Rice, African Rice, Gamete Killer-Protector System, Hybrid Sterility, Segregation Distortion

### **W929: Rice Functional Genomics**

#### **WRKY53 Integrates the Classic BR Signaling and Mitogen-Activated Protein Kinase Pathway to Regulate Rice Architecture and Seed Size**

**Qingyun Bu**, Institute of Geography and Agroecology, Chinese Academy of Sciences, Haerbin, China  
Plant architecture and seed size are close related to rice yield. Brassinosteroids (BRs) signaling (GSK2-BZR1) and MAPK pathway (MAPKKK10-MAPKK4-MAPK6) are two major regulatory pathways that control plant architecture and seed size. However, the relationship and crosstalk node between BR signaling and MAPK pathway remain elusive.

Here, we identified WRKY53 as an important regulator of rice BR signaling. Phenotypic analyses showed that WRKY53 positively regulates rice BR signaling. We found GSK2 directly phosphorylates and destabilizes WRKY53, and WRKY53 genetically acts downstream of GSK2. Moreover, we show that WRKY53 can interact with and be phosphorylated by MAPKK4-MAPK6 cascade, and the phosphorylation is required for the biological function of WRKY53 in regulating BR responses. Genetic analysis indicated that WRKY53 acts downstream of MAPKK4-MAPK6 cascade. Therefore, we proposed that WRKY53 is also a direct target of MAPKKK10-MAPKK4-MAPK6 pathway. Together, our study not only reveals a critical role of WRKY53, but also discovers the sophisticated interplays between BR signaling and MAPK pathway in regulating rice architecture and seed size.

### **W930: Rice Functional Genomics**

#### **The International Oryza Map Alignment Project (IOMAP) from a Platinum Reference Genome Sequence (PSRefSeq) Perspective**

**Rod Wing**, Arizona Genomics Institute, University of Arizona, Tucson, AZ

Since 2003, IOMAP has aspired to create a genus-level comparative genomics platform that can be used to address both basic and applied questions in plant biology and agriculture. Fast forward 16 years: IOMAP is now on the verge of releasing 41 new or improved PSRefSeqs that represent a single accession from each of 25 wild *Oryza* species, plus African rice (*O. glaberrima*) and 15 *O. sativa* accessions that represent the 15 subpopulations of cultivated Asian rice. This 41 genome Pan-RefSeq dataset, when combined with the original *O. sativa* v.g. japonica c.v. Nipponbare “gold standard” RefSeq published in 2005, will constitute an unprecedented resource that can serve as a baseline platform to study 15 MY of evolutionary history aimed at solving the 10-billion people question. My presentation will discuss the IOMAP Pan-PSRefSeq data set, i.e. how it was generated and validated, and recent insights.

### **W931: Root Genomics**

#### **A Major Root Angle QTL in Durum Wheat Improves Yield in Drought and Crown Rot Environments**

**Samir Alahmad**, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Australia, Brisbane, Australia, Khaoula EL Hassouni, International Center for

Agricultural Research in the Dry Areas, Rabat, Morocco, Eric Dinglasan, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Australia, Brisbane, QLD, Australia, Jack Christopher, The University of Queensland - Leslie Research Facility, Toowoomba, QLD, Australia, Jason Able, The University of Adelaide, Adelaide, SA, Australia, Filippo Maria Bassi, ICARDA, Rabat, Morocco, Kai Voss-Fels, The University of Queensland, Brisbane, QLD, Australia and Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

Durum wheat (*Triticum turgidum* L. ssp. *durum*) production can experience significant yield losses due to crown rot (CR) disease. Losses are usually exacerbated when disease infection coincides with terminal drought. Durum wheat is very susceptible to CR and resistant germplasm is not currently available in elite breeding pools. Deploying physiological traits for drought adaptation such as enhanced root system architecture to reduce water stress, might minimise losses due to CR infection. A subset of lines from a nested association mapping population was evaluated for stay-green traits, CR disease incidence and yield in field experiments, as well as root traits under controlled conditions. Weekly measurements of normalized difference vegetative index (NDVI) in the field enabled modelling of the senescence pattern and calculation of stay-green traits for each genotype. Genome-wide association studies using 2,541 high quality polymorphic DArTseq markers identified a major QTL on 6A (*qSRA-6A*) and 6B (*qCR-6B*) underpinning seminal root growth angle and CR tolerance, respectively. Haplotype analyses identified allelic variants with favourable impact on yield under drought and CR environments. Results of this study highlight the value of combining above- and below-ground physiological traits to enhance yield potential. We anticipate these insights will assist breeders to design improved durum varieties that mitigate production losses due to water deficit and CR.

### **W932: Root Genomics**

#### **Characterization of a 1RS Chromosome with a 1BS Introgression Associated with Seminal Root Length and Root Development in Wheat**

**Gilad Gabay**<sup>1</sup>, Tyson R. Howell<sup>1</sup>, Junli Zhang<sup>2</sup>, Tzion Fahima<sup>3</sup>, Adam Lukaszewski<sup>4</sup> and Jorge Dubcovsky<sup>2</sup>, (1)University of California Davis, Davis, CA, (2)University of California, Davis, CA, (3)Institute of Evolution University of Haifa, Haifa, Israel, Israel, (4)University of California, Riverside, CA

We have previously shown that a 1BS introgression in 1RS (henceforth 1RS<sup>RW</sup>) is associated with differences in root length and development and water stress tolerance in field and hydroponic experiments. To identify the genomic differences between the 1RS and 1RS<sup>RW</sup>, we sequenced both arms using Illumina paired-end reads. The assembled space of scaffolds covered 145.3Mb for the 1RS arm and 153.8 Mb for the 1RS<sup>RW</sup> arm. Comparison of these assemblies with the 1BS pseudomolecule generated by the IWGSC identified a 4.6Mb segment of the 1BS arm in the 1RS<sup>RW</sup> arm. We performed additional sequencing of the coding regions of isogenic 1RS and 1RS<sup>RW</sup> and the donor of the 1BS segment Pavon. Surprisingly, the rye homologs of the wheat genes present in the 1BS introgressed segment were still present in the 1RS<sup>RW</sup> arm, indicating that this segment was not the result of two homeologous recombination events as originally proposed. To dissect the introgressed 1BS region, we generated radiation mutants. We identified one deletion affecting the proximal region of the 1BS insertion, designated as C1del. We sequenced the coding regions of C1del using exome capture and mapped the sequencing reads back to the reference wheat genome. Using this strategy, we determined that C1del has a 1.46Mb deletion of the 1BS introgressed segment. To test if the deleted region included the genes responsible for short roots, we backcrossed this deletion 4 times to Hahn-1RS<sup>RW</sup> to reduce background mutations. We genotyped and phenotyped the segregating BC<sub>4</sub>F<sub>2</sub> progeny for root length and found that plants homozygous for the deletion had significantly longer roots than the plants heterozygous or homozygous for the 1RS<sup>RW</sup> chromosome.

### **W933: Root Genomics**

## GWAS Reveals the QTLome Complexity Governing Different Root Types in Adult Durum Wheat Plants

Giuseppe Sciarra<sup>1</sup>, Marco Maccaferri<sup>2</sup>, Silvio Salvi<sup>3</sup>, Fabio Fiorani<sup>4</sup>, Kerstin A. Nagel<sup>4</sup>, Jonas Lentz<sup>5</sup>, Tania Gioia<sup>6</sup> and **Roberto Tuberosa**<sup>3</sup>, (1)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Italy, (2)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy, (3)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, Bologna, Italy, (4)Forschungszentrum Jülich GmbH, Jülich, Germany, (5)Forschungszentrum Jülich, Germany, (6)Università degli Studi della Basilicata, Italy

This study reports the high-throughput phenotyping and genetic dissection of 189 elite durum wheat for root system architecture (RSA) and shoot traits at an unprecedented level of details. Plants were grown up to the 7th leaf appearance (late tillering) using the GROWSCREEN-Rhizo platform (1-m deep rhizo-boxes) integrated with automatic root RGB imaging. The following primary RSA traits were measured: seminal, nodal and lateral total root length, root system convex hull, and root system width and depth distribution (twice/week). Measurements of leaf area, leaf number and tiller number were measured twice/week and SPAD was measured twice during the experiment. Root dry biomass and shoot fresh and dry biomass were collected at the end of the experiment. The time-course measurements of the above-mentioned traits allowed us to model root and shoot growth and thus to identify QTL describing the dynamic root growth, including the root/shoot. GWAS analysis was based upon the Illumina Infinium 90K SNP profiles and the Maccaferri et al. (2016) consensus map. The accessions are structured into five sub-groups representing the main breeding lineages worldwide, namely (i) ICARDA\_dry, with accessions bred for dryland areas, (ii) ICARDA\_temp, with accessions bred for temperate areas, (iii) ITALIAN, with accessions derived from crosses of Italian accessions with CIMMYT and Southwestern US accessions (Desert Durum®), (iv) CIMMYT\_70, with accessions adapted from CIMMYT germplasm introduced to Mediterranean countries, and (v) CIMMYT\_80, with accessions selected under irrigated conditions during the past three decades. In total, the 35 traits measured directly or computed as ratios, allowed for the identification of 75 significant QTL peaks, 13 of which affected both root and shoot traits, while 26 and 36 exclusively affected RSA or the shoot, respectively, hence indicating a limited functional overlap between the RSA and shoot QTLomes. Among the 39 QTL clusters shown to affect two or more traits, three were particularly noteworthy: (1) *QCl3subo* on chr. 1B affected ten traits, with a marked effect on root growth rate and root length density at depth, (2) *QCl25subo* on chr. 6A affected root system width and average leaf width, and (3) *QCl32* on chr. 7A affected nine traits, in particular root depth. Notably, this QTL cluster revealed a strong, contrasting selection pattern between the accessions of the rainfed and the irrigated breeding programs conducted at ICARDA and CIMMYT. The accessions used in this experiment were previously tested for yield and its components in 15 environments across the Mediterranean Basin. The joint analysis of field and platform data provides valuable insights toward a better understanding and deployment of the RSA QTLome to enhance durum wheat yield in different environmental conditions. The selection signatures evidenced in contrasting environments for water regimes suggest the possible role of the RSA QTLome in wheat adaptation and breeding in such conditions. Fine mapping and candidate genes analysis are underway for the major root growth angle on chr. 6A QTL using *in silico* TILLING. Additionally, the *QCl25subo-6A* haplotypes with contrasting effects on root growth angle are being introgressed in different genetic backgrounds.

### W934: Root Genomics

#### Monitoring Root System Dynamics

**Philip N. Benfey**, Duke University and Howard Hughes Medical Institute, Durham, NC

How roots explore their soil environment determines their ability to acquire nutrients and water as well as interact with microbes. Very few genes controlling root growth in soil are known, primarily because of the difficulty of observing the underground environment. We have identified the molecular mechanism underlying the circular movement of the root tip known as circumnutation. In collaboration with Dan Goldman (Physics, Georgia Tech) we have shown that circumnutation facilitates the root's ability to explore and penetrate soil horizons. We have also identified a compound that allows rice roots to grow through saline soil. At Hi Fidelity Genetics, we have developed a device that can monitor root growth in

soil over time. We have deployed thousands of RootTrackers and found striking differences in the response of different maize hybrids to water deficits.

### **W935: Root Genomics**

#### **Genomics of Root and Stay-Green Traits to Improve Wheat Adaptation to Late Season Drought**

**Jack Christopher**<sup>1</sup>, Cecile Richard<sup>2</sup>, Karine Chenu<sup>1</sup>, Mandy Christopher<sup>3</sup>, Valeria Paccapello<sup>3</sup>, Andrew K. Borrell<sup>4</sup> and Lee Hickey<sup>2</sup>, (1)The University of Queensland - Leslie Research Facility, Toowoomba, QLD, Australia, (2)Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia, (3)Department of Agriculture and Fisheries, Toowoomba, QLD, Australia, (4)University of Queensland, Warwick, Australia

In many wheat cropping regions, water limitation is the primary production constraint. This situation is predicted to be more frequent with future climate change. The stay-green phenotype allows crops to remain green and photosynthesize for longer than standard phenotypes after anthesis which can improve yields under late season drought. Root systems with more roots at depth can contribute by increasing access to deep soil moisture late in the season.

To study the genetics of root and stay-green traits in wheat, a multi reference parent nested association mapping (NAM) population was developed. Using the “speed breeding” system of rapid generation advance, over 1500 recombinant inbred lines were generated in approximately 18 months. A novel whole-genome NAM method (WG-NAM) was developed for genome-wide association mapping to identify markers associated with the target traits.

High-throughput techniques were developed and applied to the NAM lines to (i) phenotype seedling roots in controlled conditions, and (ii) characterize novel stay-green traits for hundreds of genotypes in standard yield plots in the field. NAM lines were phenotyped for yield and stay-green traits at multiple water-stressed and non-stressed environments during four seasons. Particular traits were associated with superior adaptation to certain environments.

Many lines with adaptive root and stay-green traits exhibited superior yield to the reference parent in relevant target environments and 54 such lines have been provided to commercial Australian wheat breeders for cultivar development.

This combination of technologies is improving understanding of the genetics and physiology of wheat adaptation to water-limited environments, accelerating genetic progress.

### **W936: Root Genomics**

#### **Novel Genetic Resources and Genes for Root Parasitic Nematodes**

**Jinrong Wan**, University of Missouri, Columbia, MO

### **W937: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops**

#### **Chromatin Accessibility Characterization of the Gene Regulatory Network Controlling Nodulation Factors Response in *Medicago truncatula***

**Daniel Conde**, University of Florida, Gainesville, FL

In legumes, perception of bacterial lipo-chitooligosaccharides (LCOs) initiates the developmental program that gives rise to the root nodule, which is colonized by nitrogen-fixing rhizobia. To gain a better understanding of the regulatory network involved in this developmental program, we measured the temporal transcriptome (RNA-seq) and genome-wide chromatin accessibility (ATAC-seq) of *Medicago truncatula* roots after treatment with *Sinorhizobium meliloti* LCOs. Roots were evaluated at eight time points between 15 minutes and 24 hours after LCO treatment. Analysis of the transcriptome data using a

time-series clustering algorithm identified groups of genes exhibiting drastic expression dynamics along the time course. In the ATAC-seq data, we observed a high correlation between chromatin accessibility and transcript abundance of the genes with significant expression changes in response to LCOs, including key regulatory genes of nodule formation such as *ERN1*, *CRE1*, *EPR3*, *DELLA1*, or *CYCLOPS*. Finally, we applied a novel algorithm to integrate RNA-seq and ATAC-seq data, which allow us to identify sets of induced and repressed genes at each time point and their upstream regulatory programs. Taken together our datasets and associated analyses are a new resource for understanding the response to LCOs in legumes, and potentially defines key regulators of root nodule symbiosis.

**Keywords:** *Medicago truncatula*; ATAC-seq; LCO; Regulatory Network

### **W938: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops**

#### **Insights & Engineering of Symbiotic Nitrogen Fixation in Poplar**

**Thomas Irving**<sup>1</sup>, Lucas Maia<sup>1</sup>, Sara Knaack<sup>1</sup>, Daniel Conde<sup>2</sup>, Matias Kirst<sup>2</sup>, Sushmita Roy<sup>1</sup> and Jean-Michel Ané<sup>1</sup>, (1)University of Wisconsin-Madison, Madison, WI, (2)University of Florida, Gainesville, FL

Current knowledge on plant symbioses with nitrogen fixing bacteria supports a single origin for the evolution of root nodules. Poplar is a close relative to the plants of this 'nitrogen fixing clade', and represents an excellent model for gain-of-nitrogen-fixation experiments. We show that poplar retains most genes considered necessary for nodulation, and describe a limited change approach to generate nodules capable of housing nitrogen fixing bacteria in a non-native host. We are working to alter promoter function to co-ordinate expression of two conserved signalling pathways to achieve this, focusing on the NIN/NLP transcription factor family and the hormone cytokinin. Insights gained from the poplar model can be used spread this agronomically useful symbiosis to more distantly related crop plants, and help increase yields in regions unable to access large-scale fertilization.

### **W939: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops**

#### **Evolution of Symbiotic Gene Regulatory Networks**

**Tatiana Vernie**<sup>1</sup>, Jean Florent Paul Keller<sup>2</sup>, Chloé Langlet<sup>2</sup>, Camille Girou<sup>2</sup> and Pierre-Marc Delaux<sup>1</sup>, (1)LRSV, Toulouse, France, (2)LRSV, France

Nitrogen is a life component available in the food chain thanks to plants. Air is full of dinitrogen, but this form is not accessible to plants. The Haber Bosch process uses fossil energy to fix dinitrogen in ammoniac, the nitrogen active form present in chemical fertilizers. This process has huge environmental and economical costs. One alternative is the biological nitrogen fixation by nitrogen fixing bacteria. Few plant species are able to establish symbioses with these bacteria. However this nitrogen-fixing root nodule symbiosis is limited to four plant orders that belong to a single clade. Recent phylogenomics studies have suggested that nitrogen-fixing root nodule symbiosis has been acquired once and lost multiple times. Only two genes have been shown to be lost in non-nodulating species and conserved in the nodulating ones. Evolution of these 2 genes does not correlate with the evolution of nitrogen-fixing root nodule symbiosis. We hypothesize that the evolution of nitrogen-fixing root nodule symbiosis is due to neofunctionalization and/or redirected gene expression. We are currently working on this second hypothesis by searching for *cis* element specific to nitrogen-fixing root nodule symbiosis. Our global approach is to genetically activate nitrogen-fixing root nodule symbiosis signalling in nodulating and non nodulating plant species to identify downstream genes expressed only in nodulating species and then look for conserved *cis* elements in their promoters. In parallel we have identified conserved *cis* elements in known nitrogen-fixing root nodule symbiosis -related genes. Validation of these *cis* elements in the legume *Medicago truncatula* is ongoing.

### **W940: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops**

## **Capturing the Conserved *cis*-Regulatory Elements and Gene Expression Complexity of Root Nodule Symbiosis within the N-Fixing Clade**

**Shifeng Cheng**, Agricultural Genome Institute at Shenzhen, CAAS, Shenzhen, China

There might be only a few of key genes involved to trigger the emergence of new adaptive traits parallelly recurring across different plant families, through co-option or rearrangement of the pre-existing genetic pathway. Here, based on our previous observation limited in the protein sequences to explore the origin and diversification of root nodule symbiosis within the N-fixing nodulation clade, we further extended our study to identify and catalogue the non-coding *cis*-regulatory elements as well as an extensive differential gene expression analysis across different ‘comparison pairs’ between different families. Combining the genome-wide approach and the target known gene family survey, we obtained a rich set of candidate non-coding elements and infection/nodulation genes specifically responded to varied species diversification and environmental conditions. We propose a co-expression, co-evolution scenario to explain the evolution of root nodule symbiosis that was recruited from the ancient existing genetic pathways.

## **W941: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops**

### **Parasponia As a Comparative System to Identify Genetic Changes Causal to the Evolution of Root Nodule Symbiosis**

**Rene Geurts**, Wageningen University, Wageningen, Netherlands

The nitrogen-fixing nodulation trait occurs in phylogenetically separated plant lineages, of which legume (Fabaceae)-rhizobium symbiosis is best known due to its agronomic importance. Evolution of nitrogen-fixing root nodules can be explained by two alternative hypotheses: (i) a single gain of the trait followed by massively parallel loss, or (ii) parallel evolution of the trait and only a few losses. For long the latter hypothesis was widely accepted. However, recent phylogenomic data using *Parasponia*-the only non-legume lineage that can establish nitrogen-fixing root nodules with rhizobium-revealed parallel loss of key nodulation genes in related non-nodulating species. These findings strongly support the alternative hypothesis; a single gain of the nitrogen-fixing nodulation trait followed by a massively parallel loss in most descendant lineages. The consequences of these findings will be presented.

## **W942: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops**

### **Challenging Soil Environments and the Origins of Nitrogen-Fixing Symbioses**

**Ryan A. Folk**, Mississippi State University, Mississippi State, MS

## **W943: Seed Genomics**

### **Gene Regulatory Networks of Maize Endosperm**

**Ramin Yadegari**, University of Arizona, Tucson, AZ

Endosperm is a major nutritive structure of the angiosperm seed and supports embryo development and the germinating seedling. In cereal grains, endosperm constitutes a large proportion of the mature seed, containing large amounts of carbohydrates and proteins. Domesticated cereal endosperm provides more than 50% of the calories in the human diet, either directly or indirectly through animal feed. To identify transcriptional processes that regulate endosperm cell differentiation, large-scale mRNA profiling of endosperm time series and individual cell types of the endosperm have been performed. In this talk, I will report on the results of our analyses identifying multiple gene-regulatory hubs associated with cell differentiation of the starchy endosperm, aleurone, and basal endosperm transfer layer. Early endosperm development is highly sensitive to drought stress which can decrease grain yield, even if favorable

moisture conditions are restored. I also will report on our efforts to understand the nature of the changes in endosperm development and the gene networks affected by drought stress.

#### **W944: Seed Genomics**

##### **Temporal Dynamics of Heat Stress Response during Early Seed Development in Rice**

**Jaspreet Sandhu**, Kan Liu, Jinyu Li, Larissa Irvin, Chi Zhang, Qi Zhang and Harkmal Walia, University of Nebraska, Lincoln, Lincoln, NE

In rice, early seed development is a key determinant of yield. The early seed development is highly sensitive to heat stress, especially impacting seed size. For instance, even a transient increase in ambient temperature negatively impacts the seed size by altering the developmental transitions of endosperm. To examine the phenotypic and molecular consequences of heat stress, we imposed a series of heat stress treatments and identified seeds at 1-2 Days after flowering (DAF) to be most sensitive to transient heat stress. To understand the molecular mechanism of rapid heat responses of 1 and 2 DAF, we performed a time-course transcriptome analysis of developing seeds under control (28°C) and heat stress (35°C). This study shows that early upregulated heat stress-responsive genes were enriched in endoplasmic reticulum responses while later heat responses were enriched in consequent cellular processes. Using differential co-expression network analysis, we identified a MYB-related gene that regulates the some of the early responses. Mutants deficient in this functional MYB-protein showed multiple seed defects under control conditions and exhibited hypersensitivity to heat stress. Our work provides insights into the temporal dynamics of molecular response of young rice seeds and identified key target pathways and genes for improving temperature resilience in rice.

#### **W945: Seed Genomics**

##### **A Novel Regulator of Seed Reserve Accumulation**

**Hiro Nonogaki**, Department of Horticulture/Oregon State University, Corvallis, OR

The major developmental programs of seeds include embryogenesis, reserve accumulation, dormancy, and maturation drying. These programs, which occur sequentially during seed development but in an overlapping manner, are controlled by both program-specific and common regulators. The analyses of the DOG1 (DELAY OF GERMINATION1) family proteins, some of which regulate abscisic acid (ABA) signaling and seed dormancy, started to highlight their functions also in other seed maturation programs than dormancy, including the induction of seed desiccation tolerance and reserve accumulation. DOG1-LIKE4 (DOGL4) was found as a novel regulator of seed reserve accumulation. Overexpression of *DOGL4* by the chemically induced gene expression system in developing and imbibed seeds caused the upregulation of a number of seed maturation-specific genes and proteins, including albumins, globulin, oleosins, LATE EMBRYOGENESIS ABUNDANT (LEA) proteins, and some of their upstream regulators also. Unlike DOG1, DOGL4 does not enhance seed dormancy. While DOG1 also induces seed desiccation tolerance-associated genes, only limited overlaps were found between the DOG1- and DOGL4-induced genes, suggesting their diverged functions in the seed maturation programs. Understanding the DOGL4 function in seed endosperm has potential to increase protein and lipid contents in grain crops. Decoding the evolutionary history of the DOG1 family proteins could also address an important question about the origin of the seed programs and the emergence of seed plants.

#### **W946: Seed Genomics**

##### **Vivipary in Mangrove: Germination with a Developing Embryo**

**Qingshun Quinn Li** and Xiaoxuan Zhou, Xiamen University, Xiamen, China

Vivipary refers to the phenomenon that a sexually reproduced seed germinates on the maternal plant. Relevant to precocious germination that reduces grain yield, viviparous mutants are sought for understanding this process. However, common existence of vivipary in mangroves can be a valuable resource in seed biology research. We conducted a spatial-temporal transcriptome analyses on 4 stages and 10 maternal and embryo tissues based on anatomical observation in *Kandelia obovata*, a

representative mangrove species with genetically programmed viviparous propagules. The results revealed that biological pathways related to seed development and/or germination and environment response are significantly regulated. Genes related to hormone biosynthesis and signal transduction, cell cycle, flavonoid biosynthesis were identified in 14 pair-wise comparison of 11,287 differentially expressed genes, suggesting a precocious changed state was set up before seed germination. Besides, gene expression network analyses revealed specific patterns of each tissue, where hydrolase activities are highlighted in seed coat, photosynthesis and rapid growth processes were highly related to cotyledon and axis, respectively. Furthermore, transporter genes especially glucosyltransferase that are highly expressed in seed coat are evidence of the nursing role of maternal tissues to the developing embryo. The trend of decreased ABA and increased GA from non-vivipary to vivipary transition were validated by RT-qPCR in 15 genes. A network containing 306 hormonal genes exhibited the central position of *LEC1*, *FUS3*, the role of which is the first reported in viviparous mangrove. Under regulation of *LEC1-FUS3-ABI3*, viviparous embryo is still maintained in an embryo developmental state. Therefore, viviparous embryo was in a biphasic status that germinates while developing. Our results shed light on embryo development and seed germination.

### **W947: Seed Genomics**

#### **Integrated Multi-Omic Framework of the Plant Response to Jasmonic Acid**

**Mathew G. Lewsey**, La Trobe University, Bundoora, Australia, Mark Zander, Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA, Natalie Clark, Iowa State University, Lingling Yin, La Trobe University, Australia, Elizabeth Hann, UC Riverside, Paola Saldierna Guzmán, UC Merced, Ziv Bar-Joseph, Carnegie Mellon University, Roberto Solano, Centro Nacional de Biotecnología-CSIC, Spain, Justin W. Walley, Iowa State University, Ames, IA and Joseph R. Ecker, Salk Institute & Howard Hughes Medical Institute, La Jolla, CA

Understanding the systems-level actions of transcriptional responses to hormones provides insight into how the genome is reprogrammed in response to environmental stimuli. Here, we investigate the signaling pathway of the hormone jasmonic acid (JA), which controls a plethora of critically important processes in plants and is orchestrated by the transcription factor (TF) MYC2 and its closest relatives in *Arabidopsis thaliana*. We generated an integrated framework of the response to JA that spans from the activity of master and secondary-regulatory TFs, through gene expression outputs and alternative splicing to protein abundance changes, protein phosphorylation and chromatin remodeling. We integrated time series transcriptome analysis with (phospho)proteomic data to reconstruct gene regulatory network models. These enable us to predict previously unknown points of crosstalk from JA to other signaling pathways and to identify new components of the JA regulatory mechanism, which we validated through targeted mutant analysis. These results provide a comprehensive understanding of how a plant hormone remodels cellular functions and plant behavior, the general principles of which provide a framework for analysis of cross-regulation between other hormone and stress signaling pathways.

### **W948: Seed Genomics**

#### **Discover Soybean Seed Quality Trait Genes, Allele and Germplasm using a Big-Data Driven Technology Platform**

**Yong-Qiang Charles An**, USDA-ARS Plant Genetics Research Unit, Donald Danforth Plant Science Center, ST. Louis, MO

Advent of next generation sequencing technologies offers an unprecedented opportunity to discover trait genes and infer gene regulatory network underlying seed reserve production for seed quality improvement. We analyzed and characterized genome resequencing data of over 1,500 diverse wild and cultivated soybean accessions, which are generated by our laboratory, our collaborators or available to the public. We further adapted/developed a suite of inter-disciplinary data mining strategies to integrate the genome sequencing data with their associated genetic and phenotypic data for discovering new trait genes, genotyping known alleles, identifying new alleles and germplasm containing desirable alleles.

The data driven platform increases the discovery efficiency significantly. We have integrated the platform with other -omics data to discover soybean oil QTL genes and gain insight into their underlying molecular mode of action.

#### **W949: Sequencing Complex Genomes**

##### **The *Oryza australensis* Genome: A Complex Rice Genome**

**Olivier Panaud**, University of Perpignan, Perpignan, France

The International *Oryza* Map Alignment Project (IOMAP) from a Platinum Reference Genome Sequence (PSRefSeq) Perspective

Since 2003, IOMAP has aspired to create a genus-level comparative genomics platform that can be used to address both basic and applied questions in plant genome evolution and agriculture. Fast forward 17 years: IOMAP is now on the verge of releasing 41 new or improved PSRefSeqs that represent a single accession from each of 25 wild *Oryza* species, plus African rice (*O. glaberrima*) and 15 *O. sativa* accessions that represent the 15 subpopulations of cultivated Asian rice. This 41 genome Pan-RefSeq dataset, when combined with the original *O. sativa* v.g. japonica c.v. Nipponbare “gold standard” RefSeq published in 2005, will constitute an unprecedented resource that can serve as a baseline platform to study 15 MY of evolutionary history aimed at solving the 10-billion people question. My presentation will discuss our new 1Gb long-read genome assembly of *O. australiensis* which survived a massive transposable element burst about 2 million years ago that essentially doubled its genome size without poloidization, i.e. how it was generated and validated, and recent insights. (IOMAP is: Mingsheng Chen, Bin Han, Robert Henry, Yue-ie Hsing, Scott Jackson, Ken McNally, Antonio Costa de Oliveira, O. Panaud, Doreen Ware and Rod A. Wing)

#### **W950: Sequencing Complex Genomes**

##### **Widespread Gene-Scale Structural Variants Revealed by Long-Range Sequencing**

**Harmeet Singh Chawla**<sup>1</sup>, Jenny HueyTyng Lee<sup>2</sup>, Julian Gabur<sup>1</sup>, Christian Obermeier<sup>1</sup>, Isobel Parkin<sup>3</sup> and Rod Snowdon<sup>1</sup>, (1)Department of Plant Breeding, Justus Liebig University, Giessen, Germany, (2)Justus Liebig University, Giessen, Germany, (3)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

There is increasing evidence that genome structural variation (SV) contributes strongly to trait variation in eukaryotic species. However, the impact of SV in complex, highly-duplicated plant genomes is difficult to assess without *de novo* genome assembly, due to an inability to effectively distinguish and assay small-scale SV events using second generation sequencing technologies such as Illumina or 454 sequencing. Third-generation long read sequencing technologies now make it possible to precisely detect small scale SV in a range of 30 to 1000 base-pairs. By analysing medium coverage long read sequencing data originating from 12 diverse *Brassica napus* genotypes, we were able to identify widespread, genome-wide, intragenic SV events, many of which were discovered in genes contributing to eco-geographical adaptation and other agronomically important traits. Our results suggest that revisiting complex plant genomes using long-read sequencing can reveal unexpected levels of functional gene variation, with major implications for trait regulation and crop improvement.

#### **W951: Sequencing Complex Genomes**

##### **Sequencing and Assembling Highly Heterozygous and/or Repeat-Rich Plant Genomes using Oxford Nanopore Technology**

**Victor A. Albert**<sup>1</sup>, Brad Abramson<sup>2</sup>, Nicholas Cho<sup>3</sup>, Kenji Fukushima<sup>4</sup>, Steven J. Fleck<sup>1</sup>, Nolan Hartwick<sup>5</sup>, Gillian Khew<sup>6</sup>, Charlotte Lindqvist<sup>1</sup>, Yee Wen Low<sup>6</sup>, Katie Nolan<sup>7</sup>, Sitaram Rajaraman<sup>3</sup>, Tanya Renner<sup>7</sup>, Franziska Saul<sup>4</sup>, Jarkko Salojärvi<sup>8</sup>, Mathias Scharmann<sup>9</sup>, Crystal Tomlin<sup>1</sup> and Todd Michael<sup>2</sup>, (1)University at Buffalo, Buffalo, NY, (2)J. Craig Venter Institute, La Jolla, CA, (3)Nanyang

Technological University, Singapore, (4)University of Wuerzburg, Germany, (5)J. Craig Venter Institute, (6)Singapore Botanic Gardens, Singapore, (7)Penn State University, State College, PA, (8)University of Helsinki, Helsinki, Finland, (9)University of Lausanne, Switzerland

The development of long read sequencing technologies has entirely changed the landscape of possibilities for sequencing and assembling complex plant genomes. While Pacific Biosciences SMRT sequencing has served admirably for a number of years now, Oxford Nanopore technology is highly portable and requires much less up-front investment. Concerns have arisen over Nanopore error rates compared to PacBio or Illumina, but at least using current versions of flow cells, chemistry, and base-calling, we no longer find these misgivings tenable regarding the construction of a highly contiguous genome. Still, a combined approach is required, at minimum including polishing using low-error - but cheap - Illumina reads. Assemblies at the chromosome scale often require further efforts, such as HiC scaffolding - but the work flow is now democratized to the extent that any university lab should be able to generate a high-quality genome of its choice. We use case-by-case workflows for generating chromosome-scale assemblies of various-sized plant genomes. Unfortunately, no one assembly approach (e.g., De Bruijn graph or overlap-layout-consensus method) works best for all species, given the various and sundry nature of their heterozygosities, ploidy levels, and transposable element blooms (the latter two also in terms of their event ages). Despite not yet achieving a truly pipeline approach, we are satisfied with our ability to generate excellent de novo genomes on unprecedentedly low time and cost scales. We will describe several of our recent projects and the individual challenges encountered and how they were overcome.

## **W952: Sequencing Complex Genomes**

### **The Importance of Sequencing Depth for a Complex Transcriptome**

**Adhini Sudhindra Kumar Pazhany**<sup>1</sup>, Virginie Perlo<sup>2</sup>, Frederik Botha<sup>3</sup>, Agnelo Furtado<sup>4</sup>, Angela O'Keeffe<sup>5</sup>, Ardy Kharabian Masouleh<sup>5</sup>, Robert Henry<sup>6</sup>, Karen Aitken<sup>7</sup>, Angeliqne d'Hont<sup>8</sup>, Adam Healey<sup>9</sup>, Jane Grimwood<sup>10</sup>, Kerrie Barry<sup>11</sup> and Jeremy Schmutz<sup>10</sup>, (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia, (2)University of Queensland, Brisbane, Australia, (3)QAAFI Queensland Alliance for Agriculture and Food Innovation - UQ University, St Lucia, QLD, Australia, (4)University of Queensland/QAAFI, Brisbane, QLD, Australia, (5)QAAFI (Queensland Alliance for Agriculture and Food Innovation), The University of Queensland, Brisbane, QLD, Australia, (6)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia, (7)CSIRO Agriculture and Food, St Lucia, Australia, (8)CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), Montpellier, France, (9)HudsonAlpha Institute For Biotechnology, Huntsville, AL, (10)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (11)Department of Energy Joint Genome Institute, Walnut Creek, CA

Sugarcane, a unique biological system which has evolved over time to suit the changing human needs is now being manipulated as a source of alternate energy and an array of platform chemicals. Sugarcane is a crop with diverse end uses and applications, a wealth of genetic resources and a rich breeding history. The daunting size and extreme complexity of the genome together with high heterozygosity and variable chromosome numbers has long hampered genomic research in sugarcane. Analysis of the transcriptome using long read sequencing has been reported. Normalization of libraries before sequencing has been widely employed in transcriptome analysis. In the complex sugarcane transcriptome, normalization was found to both reveal more rare sequences and result in the loss of many sequence variants. This suggested the need for deep sequencing to capture the diversity of sequences in sugarcane. Recently, a monoploid genome sequence was generated for the cultivar R570. We now report analysis of the transcriptome of this genotype and the impact of sequencing at greater depth on the recovery of transcripts. The results may guide future transcriptomic studies in sugarcane to make informed decisions on the required depth of sequencing.

## **W953: Sequencing Complex Genomes**

## **The Landscape of Large Chromosomal Inversions in Barley**

**Sudharsan Padmarasu**, Leibniz Institute for Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and "International Barley Pangenome Consortium"

Chromosomal inversions are one of the segregating structural variants (SVs), leading to the formation of new species, adaptive phenotypes and ecological clines by suppressing recombination between locally favoring alleles. Despite the prevalence of inversion in diverse taxonomic groups, knowledge on its establishment and maintenance in natural population is still lacking. Here, we demonstrate the use of chromosome conformation capture sequencing (Hi-C) to detect megabase-scale polymorphic inversions in a barley diversity panel comprising more than sixty accessions, mainly sampled from the IPK *ex situ* genebank. In this panel, we identified dozens of inversions relative to the Morex reference genome, most of them occurring at low frequency and non-recombining regions. Notably, we identified a frequent inversion of about 10 Mb size on chromosome 2H. Our survey of 2H frequent inversion in diverse barley accessions from the IPK genebank revealed its specific geographic origin and unique presence in domesticated barleys, but not in wild species. Precise delineation of the inversion breakpoints and resequencing of 308 barley accessions, encompassing domesticated and wild species revealed its robust association with late flowering, leading to the spread of single inverted haplotype in Europe, significantly to northern Europe. This study sheds light on the maintenance of large polymorphic inversion linked to adaptive traits.

Financial support of the project was provided by the German Ministry of Education and Research (BMBF, FKZ: 031B0190A).

## **W954: Sequencing Complex Genomes**

### **Comparison of Long Read Methods for Sequencing and Assembly of a Plant Genome**

**Valentine Murigneux**, Institute for Molecular Biosciences - The University of Queensland, St Lucia, QLD, Australia

We report a comparison of three long read sequencing technologies applied to the de novo assembly of a plant genome, *Macadamia jansanii*, with an estimated genome size of 780 Mb. This is a rare species that is a close relative of the macadamia nut recently domesticated in Hawaii and Australia. The species was discovered as a single population of about 30 plants in the wild in Eastern Australia. We have generated sequencing data using Sequel (Pacific Biosciences, 83x coverage), PromethION (Oxford Nanopore Technologies, 32x coverage) and stLFR (single-tube long fragment read, BGI, 77x coverage) technologies for the same DNA sample and evaluated the quality of the assemblies that can be generated directly from these data sets. Several assemblers were compared in the assembly of the PacBio and Nanopore reads. The resulting assemblies show a high contiguity (contig N50=35 kb-1.5 Mb) and a high degree of completeness (88-95% complete BUSCO genes). We also present results obtained from hybrid assemblies combining long reads technologies or short reads and long reads technologies and discuss the cost associated to the generation of each of the long read sequencing data.

## **W955: Sex Chromosomes and sex determination**

### **Convergent Origination of Dosage Compensation Mechanism**

**Diego Cortez**, Center for Genome Sciences, UNAM, Cuernavaca, EM, Mexico

Sex determination in vertebrates is a hot topic in evolutionary biology and reptiles offer a unique opportunity to test hypotheses on the origin and evolution of sex determination systems because this group presents a great variety of sex systems. During the seminar I will present what we have found out about the evolution of the dosage compensation mechanism between sex chromosomes in lizards, as well as some insights we have uncovered regarding the molecular basis of temperature-dependent sex determination in sea turtles.

## **W956: Sex Chromosomes and sex determination**

### **Haploid Selection Drives Transitions between Sex Determination Systems**

**Michael F Scott**, UCL, London, United Kingdom

Sex determination is remarkably dynamic; many taxa display shifts in the location of sex-determining loci or the evolution of entirely new sex-determining systems. Predominant theories for why we observe such transitions generally conclude that novel sex-determining systems are favoured by selection if they equalise the sex ratio or increase linkage with a locus that experiences different selection in males versus females. We use population genetic models to extend these theories in two ways: (1) We consider the dynamics of loci very tightly linked to the ancestral sex-determining loci, e.g., within the nonrecombining region of the ancestral sex chromosomes. Variation at such loci can favour the spread of new sex-determining systems in which the heterogametic sex changes (XY to ZW or ZW to XY) and the new sex-determining region is less closely linked (or even unlinked) to the locus under selection. (2) We consider selection upon haploid genotypes either during gametic competition (e.g., pollen competition) or meiosis (i.e., nonmendelian segregation), which can cause the zygotic sex ratio to become biased. Haploid selection can drive transitions between sex-determining systems without requiring selection to act differently in diploid males versus females. With haploid selection, we find that transitions between male and female heterogamety can evolve so that linkage with the sex-determining locus is either strengthened or weakened. Furthermore, we find that sex ratio biases may increase or decrease with the spread of new sex chromosomes, which implies that transitions between sex-determining systems cannot be simply predicted by selection to equalise the sex ratio. In fact, under many conditions, we find that transitions in sex determination are favoured equally strongly in cases in which the sex ratio bias increases or decreases. Overall, our models predict that sex determination systems should be highly dynamic, particularly when haploid selection is present, consistent with the evolutionary lability of this trait in many taxa.

## **W957: Sex Chromosomes and sex determination**

### **Differential Gene Expression between Fungal Mating Types is associated with Sequence Degeneration**

**Wen-Juan Ma**, Amherst College, Amherst, MA

Degenerative mutations in non-recombining regions, such as supergenes or sex chromosomes, can cause suboptimal expression, leading to differential expression between alleles if mutations occur stochastically in one or the other allele. Reduced allelic expression due to degeneration has indeed been suggested to have occurred on various sex chromosome systems. However, whether an association between specific features of degeneration and reduced allelic expression occurs has not been extensively tested and sexual antagonism can also account for differential expression on sex chromosomes. The anther-smut fungus *Microbotryum lychnidis-dioicae* is ideal for testing associations between specific degenerative signatures and differential expression because: 1) separate haploid cultures of opposite mating types help identify differential expression, 2) there are multiple evolutionary strata on its mating-type chromosomes, reflecting successive events of recombination suppression linking mating-type loci, and 3) there is no sexual antagonism that could select for differential expression as a confounding factor. We found that differentially expressed genes were enriched on the oldest evolutionary strata, and that several signatures of sequence degeneration were greater in differentially expressed than non-differentially expressed genes. Two particular degenerative signatures were significantly associated with lower expression levels within allele pairs: upstream insertion of transposable elements and acquisition of indels and/or early stop codons. Other degenerative mutations associated with differential expression included non-synonymous substitutions, altered intron and GC content. The association between differential gene expression and allele degeneration is relevant for a broad range of taxa where mating compatibility or sex is determined by genes located in large regions without recombination.

## **W958: Sex Chromosomes and sex determination**

## Identification of Sex Determination Region of the Y Chromosome in Spinach

Li'ang Yu, University of Illinois at Urbana Champaign, Urbana, IL

Spinach (*Spinacia oleracea* L.) is a dioecious leafy vegetable with a highly-repetitive genome, which is challenging for *De-novo* genome assembly. A segregating F<sub>1</sub> (pseudo-test cross) population from 'Viroflay' × 'Cornell-NO. 9' was used for genetic mapping by re-sequencing of 80 F<sub>1</sub> genomes. 40,324 SNPs from 'Viroflay' and 256,636 SNPs from 'Cornell-NO.9' were generated for construction of two high-density genetic maps. 2,378 bins (212,414 SNPs) were mapped to 6 linkage groups (LGs) with a total linkage distance of 476.83cM as the paternal 'Cornell-NO. 9' map; the maternal 'Viroflay' map consisted of 738 bins (29,282 SNPs) with 401.28cM total genetic distance from 6 LGs. Integration of two maps into a consensus map enabled additional 1,242 contigs to be anchored to 6 pseudomolecules from the published reference genome, which improved additional 237Mb (23.70%) assembly based on spinach estimated genome size. Screening of sex-co-segregating bin markers mapped the sex determination (SDR) region to LG1 at 45.18cM from 'Cornell-NO.9' map, which comprised of 5,678 sex co-segregating SNPs (39 bin makers) with an accumulative 18.4 Mb genomic size estimated. The region featured by reduced gene density, higher percentage of repetitive sequences and no recombination. 216 genes were identified in the SDR of the X chromosome counterpart in the reference genome. These resequencing-based linkage maps provide the resource for improving spinach genome *De-novo* assembly and identification of sex determination genes in spinach.

## W959: Sex Chromosomes and sex determination

### An Ancient Master Regulator Determines Sex in Poplars

Niels A. Müller<sup>1</sup>, Birgit Kersten<sup>1</sup>, Ana P. Leite Montalvão<sup>1</sup>, Niklas Mähler<sup>2</sup>, Carolina Bernhardsson<sup>3</sup>, Katharina Bräutigam<sup>4</sup>, Zulema Carracedo Lorenzo<sup>2</sup>, Hans Hoenicka<sup>1</sup>, Vikash Kumar<sup>2</sup>, Malte Mader<sup>1</sup>, Birte Pakull<sup>1</sup>, Kathryn Robinson<sup>2</sup>, Maurizio Sabatti<sup>5</sup>, Cristina Vettori<sup>6</sup>, Pär Ingvarsson<sup>3</sup>, Quentin C. B. Cronk<sup>7</sup>, Nathaniel R. Street<sup>2</sup> and Matthias Fladung<sup>1</sup>, (1)Thünen Institute of Forest Genetics, Grosshansdorf, Germany, (2)Umeå Plant Science Centre, Umeå University, Umeå, Sweden, (3)Department of Plant Biology, Linnean Centre for Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden, (4)Department of Biology, University of Toronto Mississauga, Mississauga ON, Canada, (5)Department for Innovation in Biological, Agro-food and Forestry systems, University of Tuscia, Viterbo, Italy, (6)Institute of Biosciences and BioResources, Division of Florence, National Research Council, Sesto Fiorentino, Italy, (7)Department of Botany, University of British Columbia, Vancouver, BC, Canada

Although hundreds of plant lineages have independently evolved dioecy, i.e. separation of the sexes, the underlying genetic basis remains largely elusive. Here we show that diverse poplar species carry partial duplicates of the *ARABIDOPSIS RESPONSE REGULATOR 17* (*ARR17*) ortholog within the male-specific region of the Y chromosome. These duplicates give rise to small interfering RNA and RNA-directed DNA methylation. Excitingly, CRISPR/Cas9-induced mutations demonstrate that *ARR17* functions as a sex switch triggering female development when 'on' and male development when 'off'. Despite repeated turnover events, including a transition from the XY to a ZW system, sex-specific regulation of *ARR17* is conserved across the poplar genus and likely beyond. Our data reveal how a single-gene-based mechanism of dioecy can enable highly dynamic sex-linked regions and contribute to maintaining sex chromosome recombination and integrity.

## W960: Sex Chromosomes and sex determination

### Elucidating Genetic Pathways of Sex Determination and Dimorphism in *Salix purpurea*

Brennan Hyden, Cornell University, Geneva, NY

Shrub willow (*Salix* section *Vetrix*), is a dioecious, short rotation coppice, bioenergy crop. While there are thousands of genes involved in sex dimorphism in shrub willow, there is an interest in mapping the gene networks associated with sex dimorphism in order to ultimately identify master regulator genes for sex. Here we report on expression QTL, differential expression analysis, and network analysis conducted in

*Salix purpurea*, an economically important and model shrub willow species with a ZW system and a sex determining region (SDR) mapped to 6.73 Mb on Chr15W. RNA-Seq data were obtained from 90 males and 90 females and small RNA data from 22 males and 22 females in an F<sub>2</sub> population and mapped to the recent version 5.1 genome assembly. Mapped reads were subsequently utilized for eQTL, DESeq, and network analysis. These data were used to identify gene associations and networks strongly associated with sex, and to develop hypotheses regarding master regulator genes of sex. We present these sex-associated network modules and hypothesized pathways and mechanisms of sex determination and dimorphism via candidate master regulator genes. This study is the first transcriptome-wide network analysis in *Salix* floral tissue and represents a significant step towards understanding sex determination in the genus *Salix*. Results from this study can also provide valuable insight and knowledge to sex determination in the related genus *Populus*, as well as other dioecious plant species.

## **W961: Simulation of Genetic and Genomic Systems**

### **QuLinePlus: A Versatile Module of QU-GENE for Plant Breeding Simulation**

Vivi N. Arief<sup>1</sup>, Wen-Hsi Yang<sup>1</sup>, Ian H. DeLacy<sup>1</sup>, Sai Krishna Arojj<sup>2</sup>, M. Z. Zulfi Jahufer<sup>2</sup> and Kaye E. Basford<sup>1</sup>, (1)The University of Queensland, Brisbane, QLD, Australia, (2)AgResearch Grasslands Research Centre, Palmerston North, New Zealand

QU-GENE is a computer simulation platform for plant breeding programs. It models the genotype-by-environment interaction using the E(N:K) model. There are two main components of QU-GENE, an engine that generate the genotype environment system and breeding modules that simulate different breeding strategies. Such structure gives QU-GENE capability and flexibility to compare various breeding modules created for specific purposes applied to the same breeding population. QuLinePlus is the latest addition to the QU-GENE breeding modules built to simulate breeding strategies for self, clonal, and open-pollinated crops. The default is now population crossing and hence the generation of half-sib (HS) and full-sib (FS) genetic families for developing synthetic varieties in forage breeding program. QuLinePlus, in conjunction with R, offers a versatile simulation module which can be used to simulate a wide range of breeding strategies, including both marker assisted and genomic selection. We demonstrate how QuLinePlus and R are used to compare the breeding strategies in perennial ryegrass (*Lolium perenne* L.). These breeding strategies include mass phenotypic selection; among HS phenotypic selection; among HS phenotypic selection and within HS genomic selection; and among and within HS genomic selection.

## **W962: Simulation of Genetic and Genomic Systems**

### **Simulation using *Crossword* Defines Gene-Discovery Potential of QTL-Seq**

Justin N Vaughn, USDA-ARS/University of Georgia, Athens, GA

Gene discovery is a central goal in modern plant breeding because it dramatically expands the utility of marker assisted selection and allows for targeted mutagenesis. QTL-seq combines bulk segregant analysis with next generation sequencing to identify QTL in segregating populations. Though conventionally QTL-seq is done with populations of <300 individuals, the method offers the possibility of using many thousands of individuals and thus achieving gene-scale recombination intervals. Would the genetic resolution expected from such large populations actually materialize in a real genetic context? We developed a simulation platform, called "Crossword", that uses empirical data as a starting point for simulation, allowing researchers to account for the actual genetic structure of founders as well as discontinuous recombination and genotyping artifacts that can substantially confound mapping resolution. Though Crossword has a range of applications, we focused on QTL-seq in small versus large populations. For traits controlled by 3 or 10 causal variants with variable effects, we generated a simulation database of 729 parameter sets. Results indicate the dominant role of population size in resolution strength and highlight the parameter's importance in false positive reduction, even in the context of single-plant (non-replicated) phenotyping. Parameters interact in non-linear ways, the most important being read-depth and population size. Extreme read depths, if achievable in reality, would exploit nearly all informative recombinations and lead to gene-scale mapping in minimal time.

## **W963: Simulation of Genetic and Genomic Systems**

### **AlphaSimR: An R-Package for Breeding Program Simulations**

**Robert Christopher Gaynor**, University of Edinburgh, The Roslin Institute, Edinburgh, United Kingdom

Stochastic simulation is a powerful tool for modeling long-term performance in plant or animal breeding programs. This type of modeling can be used to optimize existing breeding programs, evaluate novel breeding schemes, and test strategies for deploying new technology. Despite this potential, stochastic simulations are rarely, if ever, used to model many breeding programs. This is in part due to the difficulty in running these types of simulations. To make this task easier, AlphaSimR has been designed as a package for the R software environment. AlphaSimR provides its users with functions that match common operations undertaken in a breeding program, such as crossing and selection, so that users can build simulations by writing short scripts. This process of building simulations is relatively easy, highly intuitive, and very flexible. AlphaSimR has also been specifically designed and optimized for large scale simulations, with it easily handling simulations involving millions of individuals. This talk will provide a general overview of AlphaSimR and example of how the software has been used to model large breeding programs and teach students the basics of quantitative genetics.

## **W964: Simulation of Genetic and Genomic Systems**

### **GenoSim: A User-Friendly Simulation Tool for Sequence Reads and SNP Array Genotyping Data in Polyploid Species**

**Yanlin Liao**, Wageningen University and Research, Wageningen, Netherlands

Most recent linkage studies in polyploid crops are based on SNP array data. Genotype by sequencing (GBS) methods are gaining in popularity since they are more flexible and potentially cheaper. Genotype calling in polyploids is challenging due to multiple heterozygous classes that need to be discriminated, requiring high sequencing depth. Here we present GenoSim, a user-friendly software package for the simulation of data from sequencing or SNP arrays. The genotypes can be simulated to come from populations of arbitrary pedigree, size, ploidy level and mode of inheritance. Based on these simulated genotypes, GBS read counts or array intensities can be simulated with multiple sources of variation, to closely mimic real-life results for various crop species and genotyping technologies. The simulation of genotyping data can be useful to i) study the complexity of genotyping data by modelling the main sources of variation; ii) develop and test genotype calling software or any other software that uses SNP array intensities or read counts as input; iii) study the effect of disturbances in the genotyping data on downstream applications including genetic linkage mapping, QTL mapping and GWAS analyses for polyploids. We suggest parameters to use for approximating the data obtained from SNP array and bait capture sequence reads of potato, chrysanthemum and alstroemeria.

## **W965: Simulation of Genetic and Genomic Systems**

simplePHENOTYPES: SIMulation of Pleiotropic, Linked and Epistatic PHENOTYPES in R

**Samuel B Fernandes**, University of Illinois at Urbana-Champaign, Urbana, IL

Recent advances in genotyping and phenotyping techniques allowed for the acquisition of a great amount of data. Therefore, the deployment of powerful statistical methods such as multivariate analysis is becoming routine in many research groups. As the demand for multivariate software increases, we need better tools to validate and compare different implementations so they may grant solid results. To provide options to simulate different situations encountered by genetically correlated datasets, namely, pleiotropy and linkage disequilibrium (LD), we created an R package called simplePHENOTYPES. Using user-inputted marker data, simplePHENOTYPES will simulate phenotypes for a user-defined number of correlated traits under pleiotropy, partial pleiotropy or spurious pleiotropy. We implemented additive, dominance and epistatic models that may be simulated with different numbers of QTNs and allelic effects. Optionally, the user may input a specific allelic effect for each QTN. Otherwise, if a unique value

x is used, a geometric series will take place and the effect of the  $i$ th QTN will be  $x^i$ . To account for differential linkage disequilibrium and minor allele frequencies across experiments, we include the option for selecting a different set of QTNs in each replication. Replications are simulated assuming uncorrelated multivariate normally distributed residuals and the residual variance will be proportional to the inputted heritability. A different heritability may be assigned to each trait as well as a specific correlation for each pair of traits. Finally, output files are saved in an analysis-ready format for different software including plink/gemma, tassel and gapit.

## **W966: Single Cell Genomics**

### **Exploring New Frontiers for Systems Immunology at Single Cell Resolution**

Anthony Buzzai<sup>1</sup>, Hannah Newnes<sup>1</sup>, James Read<sup>1</sup>, Edgar Basto<sup>1</sup>, Katherine Audsley<sup>1</sup>, Emma DeJong<sup>1</sup>, Belinda Guo<sup>2</sup>, Parwinder Kaur<sup>3</sup>, Vanessa Fear<sup>1</sup>, Alistair Forrest<sup>4</sup>, Bree Foley<sup>1</sup>, **Anthony Bosco**<sup>5</sup> and Jason Waithman<sup>1</sup>, (1)Telethon Kids Institute, Australia, (2)School of Biomedical Sciences, The University of Western Australia, Australia, (3)UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA, WA, Australia, (4)Harry Perkins Institute of Medical Research, The University of Western Australia,, Australia, (5)Telethon Kids Institute, Perth, WA, Australia

The primary function of the immune system is to maintain tissue homeostasis and integrity, by providing protection against invading pathogens and cancerous cells. The immune system achieves this through the sensing of tissue damage and the subsequent activation of host defence mechanisms. The decision to respond (immunity) or not respond (tolerance) to a given challenge is mediated by three signals; the first signal (signal #1) is generated when dendritic cells acquire proteins derived from pathogens or tumors, and present these antigens for binding to specific receptors expressed on T cells. The second signal (signal #2) is provided by the engagement of costimulatory molecules expressed on the surface of dendritic cells with corresponding receptors on T cells. The third signal (signal #3) is delivered by soluble factors that are secreted in the local tissue microenvironment. Notably, the full repertoire and combination of signals that drive potent immune responses in diverse experimental settings remains to be discovered. The advent of single cell genomics together with novel computational tools that interrogate signalling paths from sender-to-receiver cells provides new opportunities to unlock the cellular and molecular drivers of immune function.

We developed a mouse model of murine melanoma, in which the B16 melanoma cell line was genetically manipulated and transplanted into mice to elicit a wide spectrum of immune responses. In these models, the disease course was characterised by either rapid progression, delayed progression, or complete elimination of the tumor. Employing single cell profiling, we demonstrated that disease progression versus elimination was associated with marked differences in the phenotype and functional markers of specific subpopulations of dendritic cells, monocytes, and macrophages. Moreover, employing computational analyses to decipher signalling paths between sender and receiver cells, we showed that the same populations of dendritic cells, monocytes, and macrophages transmit signals that elicit differential responses in effector T cell populations that mediate tumor control. This systems immunology approach has great potential to significantly expand our understanding of the cellular and molecular mechanisms that drive tumor elimination, and pave the way for the development of new therapies.

## **W967: Single Cell Genomics**

### **Tissue and Cell-Specific Regulation of the Genome during Seed Germination**

**Mathew G. Lewsey**<sup>1</sup>, Quentin Gouil<sup>2</sup>, Lim Chee Liew<sup>1</sup>, Yue You<sup>3</sup>, Marina Oliva<sup>4</sup>, Ryan Lister<sup>5</sup> and James Whelan<sup>1</sup>, (1)La Trobe University, Bundoora, Australia, (2)La Trobe University / WEHI, Australia, (3)WEHI, Australia, (4)University of Western Australia, Australia, (5)University of Western Australia, Perth, Australia

The distinct functions of individual cell types require cells to express specific sets of genes. The germinating seed is an excellent model to study genome regulation between cell types since the majority

of the transcriptome is differentially expressed in a short period, beginning from a uniform, metabolically inactive state. To better understand genome regulation in specific tissues of the seed we have conducted tissue-specific and single-cell experiments. We have used laser-capture microdissection RNA-sequencing to analyse small numbers of cells from the plumule, radicle tip and scutellum of germinating barley seeds over a 48 h time course. Tissue-specific gene expression was notably common; 25% (910) of differentially expressed transcripts in plumule, 34% (1876) in radicle tip and 41% (2562) in scutellum were exclusive to that organ. We determined that tissue-specific storage of transcripts occurs during seed development and maturation. Co-expression of genes had strong spatiotemporal structure, with most co-expression occurring within one organ and at a subset of specific time points during germination. We identified candidate transcription factors amongst these that may be regulators of spatiotemporal gene expression programs. In order to achieve greater resolution, we have recently assayed gene expression in thousands of single cells from *Arabidopsis* seeds across a replicated germination time series. Using mixtures of genetically divergent cells, we optimised methods that distinguish quiet cells from background. This facilitates the detection of gene expression changes in quieter cells, which are masked by more active cells in bulk RNA-seq. We also identify sub-tissue variation in the expression of germination-related transcription factors. Our findings contribute to the broader goal of generating an integrative model that describes the structure and function of individual cells within seeds during germination.

### **W968: Single Cell Genomics**

#### **Transdifferentiation of *Arabidopsis* Abscission Zone Cells at Single Cell Resolution**

**Jong Kyoung Kim**, DGIST, Daegu, South Korea

Abscission, shedding of plant organs, is regulated by developmental programs and environmental cues, which plays a crucial role in fruit production and crop yield in many plant species. Floral organ abscission in *Arabidopsis* occurs in the abscission zone (AZ) consisting of two neighboring cell types: residuum cells (RECs) that remain attached to the main plant body after abscission and secession cells (SECs) that are attached to the separated organs. RECs have been recently suggested to undergo transdifferentiation into epidermal cells. However, the intrinsic molecular mechanisms and extrinsic positional cues underlying this process are unknown. Here, using single-cell RNA sequencing, we show how RECs transdifferentiate into epidermal cells. Our study will help to understand molecular mechanisms of cellular plasticity in plants.

### **W969: Single Cell Genomics**

#### **A Multiomic Approach at the Single Cell Level to Better Understand the Transcriptional Regulation of Plant Genes**

Andrew Farmer<sup>1</sup>, Sandra Thibivilliers<sup>2</sup> and **Marc Libault**<sup>2</sup>, (1)National Center for Genome Resources (NCGR), Santa Fe, NM, (2)University of Nebraska Lincoln, Lincoln, NE

Similar to other complex living organisms, plants consist of different and highly specialized cell types (e.g., pollens for plant reproduction, root hair cells for water and nutrient absorption, etc.). The unique function of these different cell types is the consequence of the establishment of cell-type-specific transcriptional programs and their associated regulatory mechanisms (e.g., binding of transcription factors to regulatory elements of the genomic DNA, remodeling of the chromatin fiber). Recently, single cell transcriptomic approach has been applied on *Arabidopsis thaliana* root protoplasts allowing the accurate characterization of the transcriptional profiles of the cell-types composing seedling roots. As a first step in gaining a deeper understanding of the regulatory mechanisms controlling *Arabidopsis* gene expression, we report the use of single nuclei RNA-seq and ATAC-seq technologies from *Arabidopsis* roots and the integration of these datasets to reveal the impact of chromatin remodeling on gene transcription. The comparison of single nuclei and protoplast transcriptomes not only validated the use of nuclei as biological entities to establish biologically meaningful transcriptomes but also revealed the transcriptome of additional sub-cell types. Similarly, to our transcriptomic approach, the single nuclei ATAC-seq approach led to the distribution of the *Arabidopsis* nuclei in various clusters suggesting the differential remodeling of the chromatin between groups of cells. To better highlight the impact of

chromatin remodeling on gene transcription, we integrated single nuclei RNA-seq and ATAC-seq and revealed that cell-type-specific marker genes also display cell-type-specific pattern of chromatin remodeling. Our data suggest that the differential remodeling of the chromatin is a critical mechanism to regulate gene activity at the single cell-type level.

#### **W970: Single Cell Genomics**

##### **Accelerating Biology at True Resolution with Single Cell Analyses and Spatial Transcriptomics**

**Jason F Kim**, 10x Genomics, Pleasanton, CA

#### **W971: Single Cell Genomics**

##### **Expression Profiling Resources at the Bio-Analytic Resource: scRNA-Seq eFP Browser and More**

**Sylva Donaldson**, University of Toronto / CAGEF, Toronto, ON, Canada

The Bio-Analytic Resource (BAR) for Plant Biology integrates large scale data sets from plants and develops novel data visualization methods and easy-to-use tools for mining these data sets to enable hypothesis generation. The BAR is one of the top plant bioinformatics resources in the world, with 1.5M page views per month. ePlant is a zoomable user interface that allows the seamless viewing of large data sets across several levels of biological organization, and is now available for 16 different plants. One of the popular modules of ePlant is the eFP Browser for visualizing expression patterns of genes in different tissues and developmental stages. Cell-type-specific gene expression visualization has been available at the BAR for almost a decade, but we are now releasing an eFP Browser for single cell root RNA-seq data sets. This presentation will show this “scRNA-Seq eFP Browser” and how it integrates t-SNE plots and biological origin.

#### **W972: Single Cell Genomics**

##### **Accelerating Biology at True Resolution with Single Cell Analyses and Spatial Transcriptomics**

**Jason Kim**, 10x genomics, Pleasanton, CA

#### **W973: Small RNA**

##### **Unambiguous Deep Profiling of microRNA Targets with Chimeric-eCLIP**

**Eric Van Nostrand**<sup>1</sup>, **Brian Yee**<sup>2</sup> and **Gene W. Yeo**<sup>1</sup>, (1)University of California, San Diego, La Jolla, CA, (2)University of California, San Diego

MicroRNAs (miRNAs) are short RNAs that regulate RNA stability and translation, and are involved in a variety of human diseases. A key step towards the broad therapeutic usage of miRNAs is the unambiguous mapping of direct in vivo miRNA targets to assay both on- and off-target specificity. Previous approaches suffer from low specificity (pulldown of the entire miRNA targeting complex to obtain targets of miRNAs indiscriminately), low sensitivity (efforts to identify chimeric reads containing both miRNA and its target that yield ~1-2% recovery of chimeric fragments) or both (computational prediction algorithms), making it difficult to obtain the full regulatory map for a miRNA of interest. Here we describe how adaptation of enhanced CLIP-seq (eCLIP) to ligate miRNA with mRNA targets creates chimeric fragments that unambiguously map targets of individual miRNAs. We show that coupling chimeric ligation of microRNA with their mRNA targets (chim-eCLIP) with miRNA-specific enrichment enables unparalleled deep profiling of regulatory maps for individual miRNAs with dramatically improved chimera efficiency. chim-eCLIP has unique promise to map on- vs off-target interactions for shRNAs and other small RNAs that utilize the miRNA targeting machinery, enabling better understanding of undesired interactions for small RNA therapies.

#### **W974: Small RNA**

## **Dynamic Quantification and Annotation of Regulatory Elements and Gene Bodies using Total RNA**

**Sascha Duttke**<sup>1</sup>, Max W. Chang<sup>1</sup>, Suganya Viriyakosol<sup>1</sup>, Joshua Fierer<sup>2</sup>, Theo N. Kirkland<sup>1</sup>, Sven Heinz<sup>1</sup>, Aaron F. Carlin<sup>1</sup> and Christopher Benner<sup>1</sup>, (1)University of California, San Diego, La Jolla, CA, (2)Infectious Diseases Section, VA Healthcare San Diego

Capturing active transcription initiation is critical for studying gene expression and functional annotation of regulatory elements and genes in genomes. To facilitate such analysis in fresh or frozen tissues as well as organisms where physiological constraints or biosafety hinder nuclei isolation or genetic manipulations, we developed capped-small RNA-seq (csRNA-seq). csRNA-seq uses total RNA as starting material to detect transcription start sites of both stable and unstable RNAs including enhancer RNAs at single-nucleotide resolution. The method is highly sensitive to acute changes in transcription and identifies an order of magnitude more regulated transcripts than does RNA-seq. Interrogating tissues from species across the eukaryotic kingdoms identified unstable transcripts resembling enhancer RNAs, pri-miRNAs, antisense transcripts, and promoter upstream transcripts in multicellular animals, plants, and fungi spanning 1.6 billion years of evolution. Furthermore, we demonstrate that applying csRNA-seq together with total RNA-seq to study *Coccidioides immitis*, a BSL-3 pathogen that causes valley fever in humans, enables accurate annotation of genes and transcribed regulatory elements as well as the identification of candidate transcription factors and pathways critical for pathogenesis.

## **W975: Small RNA**

### **Rhizobial tRNA-Derived Small RNAs are Signal Molecules Regulating Plant Nodulation**

Bo Ren, Xutong Wang, Jingbo Duan and **Jianxin Ma**, Purdue University, West Lafayette, IN

Rhizobial infection and root nodule formation in legumes requires recognition of signal molecules produced by the bacteria and their hosts. Here we show that rhizobial tRNA-derived small RNA fragments (tRFs) are signal molecules that modulate host nodulation. Three families of rhizobial tRFs were confirmed to regulate host genes associated with nodule initiation and development via hijacking the host RNAi machinery that involves ARGONAUTE 1. Silencing individual tRFs with the use of short tandem target mimics or by overexpressing their targets represses root hair curling and nodule formation, whereas repressing these targets with artificial miRNAs identical to the respective tRFs or mutating these targets with CRISPR-Cas9 promotes nodulation. Our findings thus uncover a bacterial small RNA-mediated mechanism for prokaryote-eukaryote interaction and may pave the way for enhancing nodulation efficiency in legumes.

## **W976: Small RNA**

### **Quantitative, Super-Resolution Imaging of Small RNAs with sRNA-PAINT**

**Kun Huang**<sup>1</sup>, Feray Demirci<sup>2</sup>, Blake Meyers<sup>3</sup> and Jeffrey Caplan<sup>1</sup>, (1)University of Delaware, Newark, DE, (2)FiDoSoft Software Consulting, Redmond, WA, (3)Donald Danforth Plant Science Center, St. Louis, MO

Small RNAs are 21- to 24-nt non-coding RNAs that play important regulatory roles in the life of both animals and plants. Their small size and high diversity made it challenging to develop methods that have sufficient resolution and specificity. We created a method sRNA-PAINT, for the detection of small RNA with nanometer resolution. Our method utilizes the high-resolution and quantification advances in DNA-PAINT (DNA-based points accumulation in nanoscale topography) methodologies, and combines the specificity of locked nucleic acid (LNA) *in situ* detection of small RNAs. We applied sRNA-PAINT for detecting and quantifying small RNAs in different cell layers of early developmental stage maize anther that are important for male sexual reproduction.

## **W977: Small RNA**

## **Elucidating the Influence of RdDM during Reproduction through Comparative Small RNAseq in Two *Capsella* Species**

**Kelly Dew-Budd** and Mark A Beilstein, University of Arizona, Tucson, AZ

DNA methylation during seed development is partially controlled by the small RNA-directed DNA methylation (RdDM) pathway. In RdDM, small interfering (si)RNAs, which are abundant in gametophytes and developing seeds, are produced by processing transcripts from RNA Polymerase IV. These siRNAs interact with Pol V transcripts via complementary base-pairing, thereby targeting methylation machinery to proximal regions of the genome, a process critical for proper expression of imprinted genes. *Arabidopsis thaliana* RdDM mutants show no obvious reproductive defects; however, *Brassica rapa* RdDM mutants show a significant reduction in seed set and seed weight. One major difference between *A. thaliana* and *B. rapa* is breeding system, suggesting that RdDM may be more critical in *B. rapa* because methylation is mediating conflicts between maternal and paternal genomes (imprinting). To test this hypothesis, we characterized RdDM mutants in the self-compatible species *Capsella rubella* and its self-incompatible sister-species, *C. grandiflora*. We observed significant reproductive defects in both species, but *C. rubella* had more moderate defects when compared to the drastic reduction in seed set and seed weight observed in *C. grandiflora* RdDM mutants. To determine the molecular changes correlated with this phenotype, we sequenced small RNAs in reproductive and vegetative tissues in both species. The recent speciation of *C. rubella* from a *C. grandiflora* progenitor approximately 50 mya, the associated reduction in the effect of RdDM mutations on seed development in *C. rubella*, along with our small RNA sequencing results for both species, permit new insights into the genomic regions imprinted via RdDM that are relevant to the transition from outcrosser to inbreeder.

### **W978: Solanaceae**

#### **Contrasting N-Use Efficient Eggplants and Differential Expressed Genes in the N-Metabolism Pathway**

**Francesco Sunseri**, University Mediterranea of Reggio Calabria, Reggio Calabria, Italy

Nitrogen availability is one of the most limiting factors affecting crop yield. The identification of high Nitrogen Use Efficiency (NUE) genotypes may represent a valuable strategy to maintain high yield reducing N-supply. The genetic variation among eggplant genotypes in response to nitrate supply was exploited allowing us to identify two pairs of NUE-contrasting genotypes. RNAseq analysis was then assessed to identify differentially expressed genes (DEG) related to NUE pathway, after exposure to short- and long-term N-stress.

DEGs identification between genotypes in responses to low N supply were taking into account and comparisons were independently analyzed in both root and shoot. In particular, four up regulated genes were identified in high NUE genotypes by co-expression networks (GCN) analysis. Transcriptomics analysis highlighted the central role of some transcription factor (TF), which were up regulated in the N-use efficient genotypes. A TF belonging to WRKY family, involved in plant stress responses, showed a significant up-regulation. Interestingly, this WRKY33 triggered a higher expression of 21 genes including other TFs, many of which closely related to N-assimilation and remobilization.

These results fit well with the evidences of the key role of N-utilization component (NUE) to confer high NUE in eggplant suggesting the higher N-remobilization to the fruit, driven by GS enzyme, as a valuable strategy to enhance NutE.

### **W979: Solanaceae**

#### **Development of Introgression Lines Containing Segments from a Wild Pepper and their use for QTL Mapping**

**Ilan Paran**<sup>1</sup>, Vibha Mishra<sup>2</sup>, Noa Rona<sup>2</sup>, Yelena Borovsky<sup>2</sup> and Adi Doron-Faigenboim<sup>2</sup>, (1)Agricultural Research Organization, Rishon LeZion, Israel, (2)Agricultural Research Organization, Rishon LeZion, Israel

A set of pepper introgression lines (ILs) was developed using the wild *Capsicum frutescens* accession BG-2816 as a donor parent and the cultivated blocky type Cv. Maor (*C. annuum*) as a recurrent parent. Fixed ILs containing a single or few introgressions were genotyped by Genotyping By Sequencing (GBS) markers to define the genomic boundaries of each line. The ILs as well as their hybrids with Maor were subjected to phenotypic evaluation of yield and fruit traits. Numerous fruit weight QTLs were detected and we chose to focus on IL2-8 that had the greatest effect on this trait in the population. High-resolution mapping of the QTL was done by Bulk segregant analysis (BSA) using two bulks representing the largest and smallest fruits from an F<sub>2</sub> population of the cross IL2-8 x Maor. A 10Mb region in chromosome 2 was determined as containing the QTL. Phenotypic analysis of fixed recombinants within the QTL region indicated the occurrence of multiple tightly linked sub-QTLs affecting fruit weight and shape. Several candidate genes that control organ growth located within the QTL region were identified and are currently being analyzed for sequence variation and expression pattern.

#### **W980: Solanaceae**

##### **Meta-Analysis of GWAS Provides New Insights into Genetic Control of Tomato Flavor**

Jiantao Zhao, Christopher Sauvage, Frederique Bitton and **Mathilde Causse**, INRA, Montfavet, France

Tomato flavor has changed over the course of long-term domestication and intensive breeding. To understand the genetic control of flavor related traits, we performed a meta-analysis of genome-wide association studies (GWAS) for 18 traits, using 775 tomato accessions and 2,316,117 SNPs from three GWAS panels (Sauvage et al., 2014; Bauchet et al., 2017 and Tieman et al., 2017). In a first step, we imputed SNP data to increase genome coverage for two out of three panels and ran EMMAX software for each panel separately. From each run, the genomic inflation factor and standard errors of the beta coefficients were implemented in two meta-GWAS models: the inverse variance-weighted fixed effect and the Han Eskin random-effect model that takes heterogeneity across studies into account. Then, from the meta-GWAS results, we conducted a statistical fine-mapping of the candidate loci following a heuristic linkage disequilibrium (LD) approach. We discovered 305 significant associations for the contents of sugars, acids, amino acids and flavor-related volatiles. We showed that fruit citrate and malate contents have been impacted by selection during domestication and improvement, while sugar content has undergone less stringent selection. Results suggest that it may be possible to significantly increase volatiles that positively contribute to consumer preferences while reducing unpleasant volatiles, by selection of the relevant allele combinations. Our results provide genetic insights into the influence of human selection on tomato flavor and demonstrate the benefits obtained from meta-analysis.

#### **W981: Solanaceae**

##### **Toward the First Commercialization of Gene Edited Tomato with Improved Nutrition**

**Hiroshi Ezura**<sup>1,2</sup>, Minako Sumiyoshi<sup>3</sup>, Mariko Takayama<sup>2,3</sup> and Naozumi Mimida<sup>3</sup>, (1)Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan, (2)Tsukuba Plant Innovation Research Center, University of Tsukuba, Tsukuba, Japan, (3)Tsukuba R&D Center, Sanatech Seed Co., Ltd., Tsukuba, Japan

GABA ( $\gamma$ -aminobutyric acid) is a non-proteinogenic amino acid with health-promoting functions for human. Although tomato fruits have a relatively high GABA content compared with other crops, the levels must be further increased to effectively confer the health-promoting functions such as lowering blood pressure and improving sleep. Glutamate decarboxylase (GAD) is a key enzyme in GABA biosynthesis in tomato; it has a C-terminal autoinhibitory domain that regulates enzymatic function, and

deletion of this domain increases GAD activity. The tomato genome has five GAD genes, of which two are expressed during tomato fruit development. To increase GABA content in tomato, we deleted the autoinhibitory domain of SIGAD3 using CRISPR/Cas9 technology. Introducing a stop codon before the autoinhibitory domain in the SIGAD3 increased GABA accumulation by 7 to 15 folds (Nonaka et al., 2017). We also evaluated the potential of the gene edited tomato as a breeding material for hybrid tomatoes. The hybrid lines showed high GABA accumulation in the fruits, which was sufficiently high for expecting health-promoting functions and had minimal effects on other fruit traits, suggesting that the high GABA is a dominant trait and that the gene-edited tomato would be useful as a parental line of hybrid cultivars (Lee et al., 2018). We applied this technique to a parental line of commercial cultivars, and produced a high GABA hybrid cultivar. In this presentation, we introduce these results as well as regulatory consideration and its recent progress in Japan.

## **W982: Solanaceae**

### **Accessing Epigenetic Variation within Tomato Germplasm Toward Improved Heterosis**

Sally A Mackenzie, Pennsylvania State University, University Park, PA

Sally Mackenzie, Penn State University

Plants sense their environments and can display phenotypic plasticity under altered environmental conditions. To investigate the nature of this plasticity in detail, our group developed a system to exploit changes in plant behavior that occur with suppression of the *MSH1* gene. In tomato, RNAi-suppression of *MSH1* results in epigenetic reprogramming of the plant and involves genome-wide methylome repatterning. Segregation of the RNAi transgene results in a heritable state of *msh1* memory; crossing or grafting (as rootstock) of *msh1*-modified lines produces progeny that are enhanced in growth vigor and resilience. Introduction of this system to tomato permits cross-species comparison of *MSH1* effects with *Arabidopsis* for the identification of conserved gene networks underpinning the emergent phenotypes. Data indicate that the *MSH1* system, with integrated methylome, gene expression and phenotype datasets, effects changes in circadian rhythm, auxin response, phytohormone signal transduction and mRNA spliceosome integrated gene networks. Tomato *MSH1* crossing and grafting experiments conducted in both greenhouse and field demonstrate that the epigenetic variation deriving from *MSH1* manipulation scales stably for crop production testing. Data to date indicate that the *MSH1* system enhances fruit yield in both Rutgers and elite variety genotypes, markedly outperforming heterosis outcomes in parallel control crosses. Resilience phenotypes associated with *MSH1* effects imply that epigenetic variation may also serve to stabilize yield in uneven environmental conditions. These effects, and their epigenomic underpinnings, appear to reflect phenotypic plasticity amenable to adaptation.

## **W983: Solanaceae**

### **Diseases of Potato: Breaking Down Resistance**

Dennis Halterman, USDA-ARS, Madison, WI

Microorganisms that cause plant diseases present a substantial burden to agriculture through yield losses due to plant stress, costs associated with disease control, and efforts to detect infections and limit disease epidemics. Plant breeders are interested in the identification and incorporation of simply inherited genes that confer robust resistance to diseases. These resistance genes typically encode proteins that recognize the presence of very specific pathogen molecules, termed effectors, resulting in the activation of defense responses. However, deletion or mutation of effectors can allow individuals in pathogen populations to elude recognition, cause disease, and pass this advantage on to subsequent generations. The discovery and characterization of molecular mechanisms that allow pathogens to 'break' host resistance will aid in the development of germplasm that can better resist this phenomenon. In potato we are fortunate to have a readily accessible source of germplasm that includes wild relatives of cultivated potato. Due to their coevolution with pathogens in their native environments, these wild species provide a vast resource of resistance to economically important diseases. In our laboratory we are using a combination of approaches to combat the potato late blight pathogen, *Phytophthora infestans*, including the identification of novel sources of resistance (resistance gene identification in

germplasm), characterization of existing genetic resistance (how do the resistance genes recognize the pathogen to turn on resistance), and identification of microorganisms that could help prevent late blight disease from forming (biocontrol). Our progress on using all of these approaches will be discussed.

#### **W984: Solanaceae**

##### **Uncovering the Mysteries of Haploid Induction in Potato**

**Ek Han Tan**<sup>1</sup>, Oluwafemi A. Alaba<sup>1</sup>, Kirk Amundson<sup>2</sup>, Benny Ordonez<sup>2</sup>, Isabelle M. Henry<sup>2</sup>, Monica Santayana<sup>3</sup>, Elisa Mihovilovich<sup>4</sup>, Gregory Porter<sup>5</sup>, Merideth Bonierbale<sup>3</sup> and Luca Comai<sup>6</sup>, (1)University of Maine, Orono, ME, (2)UC Davis, Davis, CA, (3)International Potato Center, Peru, (4)International Potato Center, Lima, Lima, Peru, (5)University of Maine, (6)Plant Biology and Genome Center, UC Davis, Davis, CA

Ever since the discovery of the haploid induction cross in potato (*Solanum tuberosum* L.), there has been much interest in developing this tool to accelerate potato breeding and improvement efforts. So far, the most common use of the potato haploid induction cross is to generate diploid potato lines (known as primary dihaploids) from elite, tetraploid cultivars. Consequently, the nature of the potato haploid induction cross, which is thought to proceed via uniparental genome elimination, may give rise to dihaploids that inherit haploid inducer genome. In other well-characterized haploid induction systems such as corn, barley and *Arabidopsis*, incomplete loss of the haploid inducer genome can give rise to aneuploidy when the haploid inducer chromosome is partially retained. However, because of the autotetraploid nature of cultivated potato, primary dihaploids derived from haploid induction may also be aneuploid as a result of gametic aneuploidy. Using genomic methods, we characterized a population of primary potato dihaploids to ascertain the extent of haploid inducer genome contribution. We hope to further our understanding on the mechanism behind potato haploid induction and to promote diploid potato breeding efforts.

#### **W985: Sorghum/Millet**

##### **Introduction of Sorghum and Millet Workshop**

**Yinghua Huang**, USDA ARS, Stillwater, OK

#### **W986: Sorghum/Millet**

##### **Structural Genomic Variation and the Evolution of Sugar Accumulation in Sweet Sorghum**

**Elizabeth Cooper**, University of North Carolina at Charlotte, Kannapolis, NC

The process of crop domestication by humans often consists of two stages: 1) initial domestication, where the wild, ancestral species is first cultivated, and traits that allow easier propagation are strongly selected for, and 2) diversification, where domesticated species are adapted to multiple, more specialized uses, and many different types of traits may be selected on. Selective pressure to increase sugar accumulation in certain varieties of the cereal crop *Sorghum bicolor* is an excellent example of the latter; this has resulted in pronounced phenotypic divergence between sweet and grain-type sorghums, but the genetic mechanisms underlying these differences remain poorly understood. Using a *de novo* assembled genome from a representative sweet sorghum genotype, we uncovered a handful of structural genomic changes and loss of function mutations that may be related to differences in sugar accumulation and could represent new targets for crop improvement.

#### **W987: Sorghum/Millet**

##### **A Step Forward in Predicting Phenotypic Performance in Diverse Field Environments**

**Qi Mu**, Xianran Li, Tingting Guo, Xin Li and Jianming Yu, Iowa State University, Ames, IA

Phenotypic plasticity describes that a genotype behaves differently when exposed to different environments. Phenotypic plasticity creates difficulty for plant breeders to select the best breeding lines

across locations and years. Understanding and being able to predict phenotypes across diverse environments will facilitate selecting stable or adaptive lines with given resources. Recently, we established a joint genomic regression analysis (JGRA) framework to dissect the complex flowering time plasticity observed in natural field environments by leveraging an explicit environmental index. This is a critical step forward to incorporate environmental inputs into performance prediction under phenotypic plasticity. In this study, we hypothesized that plant height plasticity can be unraveled and explained in a similar manner. The objectives were to 1) uncover the patterns of sorghum plant height plasticity in diverse environments; 2) predict performance in new environments; 3) identify and dissect the genetic determinants to explain the observed plasticity. Our results showed that varied degree of plasticity in plant height of sorghum lines could be explained, modeled, and predicted with a biologically meaningful environmental index. High prediction accuracy was achieved by using this environmental index. The effects of three height QTLs changed dynamically across environments, contributing to the observed phenotypic plasticity. In conclusion, integrating environmental data with genomic components has enhanced our understanding for phenotypic plasticity, and enabled predictive modeling for multiple agronomically important traits under diverse field environments.

### **W988: Sorghum/Millet**

#### **Large Scale Genome-Wide Association Study Reveals that Drought Induced Lodging in Grain Sorghum is associated with Plant Height and Traits Linked to Carbon Remobilisation**

**Xuemin Wang**<sup>1</sup>, Emma Mace<sup>1</sup>, Yongfu Tao<sup>2</sup>, Alan W Cruickshank<sup>3</sup>, Colleen H Hunt<sup>3</sup>, Graeme L Hammer<sup>4</sup> and David Jordan<sup>1</sup>, (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, QLD, Australia, (2)The University of Queensland, Warwick, QLD, Australia, (3)Department of Agriculture and Fisheries, Queensland Government, Warwick, QLD, Australia, (4)The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, St Lucia QLD, Australia

Sorghum is generally grown in water limited conditions and often lodges under post-anthesis drought, which reduces yield and quality. Due to its complexity, our understanding on the genetic control of lodging is very limited. We dissected its genetic architecture in grain sorghum through genome-wide association study (GWAS). GWAS was conducted on 2308 unique hybrids grown in 17 Australian sorghum trials over 3 years. The GWAS detected 213 QTL, the majority of which showed a significant association with leaf senescence and plant height (72% and 71% respectively). Only 16 lodging QTL were associated with neither leaf senescence nor plant height. The high incidence of multi-trait association for the lodging QTL indicates that lodging in grain sorghum is associated with plant height and traits linked to carbon remobilisation. This result supported the selection for stay-green (delayed leaf senescence) to reduce lodging susceptibility, rather than selection for short stature and lodging resistance *per se*, which likely reduces yield. Additionally, our data suggested a protective effect of stay-green on weakening the association between lodging susceptibility and plant height. Our study also showed that lodging resistance might be improved by selection for stem composition but was unlikely to be improved by selection for classical resistance to stalk rots.

### **W989: Sorghum/Millet**

#### **Genomics-Enabled Dissection of Enhanced Water and Nitrogen Use Efficiency in Sorghum**

**Andrea L. Eveland**, Donald Danforth Plant Science Center, Saint Louis, MO

### **W990: Sorghum/Millet**

#### **Genetic Dissection of Drought Tolerance in a Sorghum Backcross Nested Association Mapping Population**

**Hongxu Dong**, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

With a worldwide water crisis looming, a primary goal of sorghum breeding is to improve drought tolerance. To address this challenge in sorghum, we developed a backcross nested association mapping (BCNAM) population using 12 diverse founder lines crossed with an Ethiopian elite cultivar Teshale. The sorghum BCNAM population was trialed under two natural drought environments and one normal growing environment in Ethiopia. We characterized 1178 BC<sub>1</sub>F<sub>4</sub> lines with 4395 single nucleotide polymorphisms (SNPs) and conducted joint linkage (JL) and genome-wide association analyses (GWAS) for nine adaptive traits. Phenotype was less heritable under drought conditions than under normal growing condition. A total of 177 JL quantitative trait loci (QTL) and 267 GWAS hits were detected across the three environments for the nine traits. 159 JL QTLs (89%) in this study had correspondence with QTLs for the same traits from 72 previous studies. Associations detected in JL and GWAS clustered within known stay-green QTL regions and largely overlapped with the differentially expressed genes between stay-green and senescent sorghum genotypes from previous transcriptomic study. Fixation index ( $F_{ST}$ ) between a subset of 58 lines that failed to survive drought stress (*i.e.* drought-sensitive) and the remaining 1120 plants (*i.e.* drought-tolerant) showed elevated peaks within stay-green QTL regions, where most associations were detected in this study. Flowering regulator such as *Ma6* and drought resistant gene such as *P5CS2* were in proximity to these associations. Interestingly, using the model-selected SNPs that associated with nine traits across three environments, phenotypic prediction accuracies for grain yield were equivalent to genome-wide SNPs and were significantly better than an equivalent number of random SNPs, indicating that these drought-related traits are predictive of sorghum grain yield. These findings validate the BCNAM resource in trait mapping in sorghum and demonstrate the value of NAM design for dissection of adaptive traits.

## **W991: Soybean Genomics**

### **Genomic of Floral Transition in Soybean**

**Prem L Bhalla**, The University of Melbourne, Melbourne, Australia

Floral transition is a critical developmental switch that can impact vital agronomic traits such as yield. Understanding the floral transition is key in ensuring future food security for the growing population. Soybean is an important legume crop, providing a significant source of worldwide oilseed production. However, soybean is a photoperiod sensitive plant whose floral transition is triggered by exposure to short-day conditions. Our knowledge of crucial molecular regulators and networks of flower initiation in legumes, including soybean, is limited. We used integrated genomics and experimental approach to addresses this gap in our knowledge. Our comparative genomics (Jung et al *PLoS ONE* 7(6): e38250) and RNA-Seq analysis (Wong et al *PLoS ONE* 8(6): e65319. doi:10.1371/journal.pone.0065319) of shoot apical meristem and leaf undergoing floral transition revealed major reprogramming events in leaves and the shoot apical meristem. The data also showed an extensive reprogramming of genes associated with the epigenetic chromatin modifications and RNAi gene silencing in the shoot apical meristem during the floral transition. Further, soybean is a paleopolyploid legume with multiple gene copies. Our study also revealed the functional divergence of flowering genes in soybean (Liew et al *Scientific Reports* 7: 10605; Arya et al *Scientific Reports* 8:12569).

## **W992: Soybean Genomics**

### **Ionome as an Intermediate Phenotype of Soybean and its possible use for Genomic Prediction**

**Yoshihiro Ohmori**<sup>1</sup>, Ryokei Tanaka<sup>1</sup>, Kiyoshi Yamazaki<sup>1</sup>, Yusuke Toda<sup>1</sup>, Yuji Yamasaki<sup>2</sup>, Hirokazu Takahashi<sup>3</sup>, Hideki Takanashi<sup>1</sup>, Hiromi Kajiya-Kanegae<sup>4</sup>, Mai Tsuda<sup>5</sup>, Hisashi Tsujimoto<sup>2</sup>, Akito Kaga<sup>6</sup>, Mikio Nakazono<sup>3</sup>, Hiroyoshi Iwata<sup>1</sup> and Toru Fujiwara<sup>1</sup>, (1)The University of Tokyo, Tokyo, Japan, (2)Arid Land Research Center, Tottori University, Tottori, Japan, (3)Nagoya university, Nagoya, Japan, (4)Research Center for Agricultural Information Technology, NARO, Tsukuba, Japan, (5)University of Tsukuba, Tsukuba, Japan, (6)National Institute of Agrobiological Sciences, Tsukuba, Japan

The ionome is defined as the mineral composition of an organism. Plant ionome is determined by both genetic and environmental factors and plant ionome affects plant growth. Therefore, ionome can be considered as parameters reflecting an internal state of plants and also as determinants of growth. As a

part of a CREST project to develop a novel genomic selection system for improvement of drought stress tolerance, we determined the ionome of leaves from nearly 200 genotypes of soybean germplasm grown in fields of Arid Land Research Center of Tottori University under different levels of water supply. Analysis of the data using a random forest algorithm revealed that the leaf ionome of field grown soybean predicts their growth, such as plant height and fresh weight with correlation coefficients in the range of 0.30 to 0.34. This suggests that the leaf ionome represents plant status that allows prediction of plant growth. In other words, ionome is an intermediate phenotype that has a potential to be used for improvement of the genomic selection system. In this presentation, we will introduce our project and show recent results obtained by ionomics of field grown soybean. We will also mention a prospect of ionomic information for genomic prediction.

### **W993: Soybean Genomics**

#### **Characterization of Soybean STAY-Green Genes in Susceptibility to Foliar Chlorosis of Sudden Death Syndrome**

**Hao-Xun Chang**, National Taiwan University, Taipei, Taiwan

### **W994: Soybean Genomics**

#### **A Map of Genetic Variation from 781 Soybean Genomes**

**Soon-Chun Jeong**<sup>1</sup>, Myung-Shin Kim<sup>2</sup>, Roberto Lozano<sup>3</sup>, Ji Hong Kim<sup>1</sup>, Dong Nyuk Bae<sup>1</sup>, Man Soo Choi<sup>4</sup>, Namhee Jeong<sup>4</sup>, Soo-Kwon Park<sup>5</sup>, Michael A. Gore<sup>6</sup> and Jung-Kyung Moon<sup>7</sup>, (1)Korea Research Institute of Bioscience and Biotechnology, Chungbuk, South Korea, (2)Department of Plant Science, Plant Genomics and Breeding Institute and Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul, Korea, Republic of (South), (3)Cornell University, Ithaca, NY, (4)National Institute of Crop Science, Jeonbuk, South Korea, (5)National Institute of Crop Science, Rural Development Administration, Jeonju, Korea, Republic of (South), (6)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (7)National Institute of Crop Science, Rural Development Administration, Jeonju-si, Korea, Republic of (South)

Soybean is an economically and environmentally important major crop worldwide. It is a predominant plant protein and oil source of both food and feed and has capacity to fix atmospheric nitrogen by intimate symbioses with microorganisms. Here we present a fine genome-wide variation map in 781 accessions including 418 domesticated (*Glycine max*) and 345 wild (*Glycine soja*) soybeans and 18 of their natural hybrids. We identified 31 million single nucleotide polymorphisms and 5.7 million small indels that contribute to within- and between-population variation. We describe a comprehensive characterization of the geographic and functional differentiation of rare and common genetic variants with insights into the domestication history of soybean and detection of domestication-selective sweeps. We show that this resource enables us to increase marker density of existing data sets for improving the resolution of association studies.

### **W995: Soybean Genomics**

#### **Global Gene Coexpression Networks Give Insight into the Evolution of Nodulation in Non-Legumes and Legumes**

**Xuelu Wang**, Huazhong Agricultural University, Wuhan, China

Legumes can form nodules by intimate symbioses with rhizobia to fix atmospheric nitrogen. Symbiotic nitrogen fixation (SNF) provides about 40-60 million tons of nitrogen for agricultural systems each year in an ecofriendly manner. What happens for SNF during the evolution of legumes is always an outstanding question in evolutionary developmental biology. Soybean (*Glycine max*) is the most important crops for proteins and dietary oil and the fourth largest crop in production in the world. In addition, soybean belongs to ureide-forming legume and has a high efficiency of translocation of fixed N as ureide. We used weighted coexpression network analysis (WGCNA) to mine a nodule-related module (NRM) in

soybean. Comparative genomic analysis of 78 green plant species revealed that NRM genes are recruited from different evolutionary nodes along with gene duplication events. A set of core coexpressed genes within legumes may play vital roles in regulating nodule environments essential for nitrogen fixation, including oxygen concentrations, sulfur transport, and iron homeostasis (such as *GmCHY*). We revealed that ancient orthologs and duplication events before the origin of legumes were preadapted for symbiosis. Conserved coreregulated genes found within legumes paved the way for nodule formation and nitrogen fixation. These findings provide significant insights into the evolution of nodulation and indicate promising candidates for identifying other key components of legume nodulation and nitrogen fixation. We are currently working on the hub genes in the NRM to investigate their role in soybean nodule development.

## W996: Soybean Genomics

### Lipidomic and Transcriptomic Profiling of Developing Nodules Reveal the Essential Roles of Active Glycolysis and Fatty Acid and Membrane Lipid Biosynthesis in Soybean Nodulation

Gaoyang Zhang<sup>1</sup>, Beibei Chen<sup>2</sup>, Muhammad Z. Ahmad<sup>3</sup>, Sehrish Manan<sup>3</sup>, **Jian Zhao**<sup>4</sup> and Gaoyang Zhang, (1)Huazhong Agricultural University, Wuhan, China, (2)Huazhong Agricultural University, Wuhan, China, (3)Anhui Agricultural University, China, (4)Anhui Agricultural University, Hefei, China

Symbiotic rhizobia-legume interactions are energy-demanding processes, and the carbon supply from host cells that is critically required for nodulation and nitrogen fixation is not fully understood. The investigated lipidomic and carbohydrate profiles of the transcriptome of developing nodules revealed highly activated glycolysis, fatty acid (FA), 2-monoacylglycerol (2-MAG), and membrane lipid biosynthesis and transport during nodule development. RNA-Seq profiling of metabolic genes in roots and developing nodules highlighted particularly upregulated pathways involved in biosynthesis and transport of FAs, membrane lipids, and 2-MAG in rhizobia-soybean symbioses via the GRAS-WRI-FatM-GPAT-STR pathway, similar to that in legume-arbuscular mycorrhizal fungi symbiosis. The essential roles of the metabolic pathway during soybean nodulation were further supported by analysis of transgenic hairy roots overexpressing soybean *GmWRI1b*-OE or-KD and *GmLEC2a*-OE. *GmLEC2a*-OE hairy roots produced fewer nodules, in contrast to *GmWRI1b*-OE hairy roots. *GmLEC2a*-OE hairy roots displayed different or even opposite expression patterns of the genes involved in glycolysis and the synthesis of fatty acids, 2-MAG, TAG, and membrane lipids compared to *GmWRI1b*-OE hairy roots. Glycolysis, fatty acid and membrane lipid biosynthesis were repressed in *GmLEC2a*-OE but increased in *GmWRI1b*-OE hairy roots, which may account for the reduced nodulation in *GmLEC2a*-OE hairy roots but increased nodulation in *GmWRI1b*-OE hairy roots. *GmWRI1b*-KD hairy roots with reduced FA, membrane lipids (PC and MGDG), and TAG/MAG, also had reduced nodule numbers. These data show that active fatty acid, MAG and membrane lipid biosynthesis are essentially required but that TAG biosynthesis is not essential for nodulation and rhizobia-soybean symbioses. These data shed more light on root lipid metabolism as a prerequisite for soybean nodulation, laying foundations for future detailed investigations of soybean nodulation.

#### References:

Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., Wang, X., and Zhao, J. (2019) Multiple *GmWRI1s* are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (*Glycine max*). *Plant Biotechnol J*. doi: 10.1111/pbi.13183.

Ahmad, M.Z., Rehman, N., Yu, S., Zhou, Y., Haq, B., Li, P., Wang, J., Zeng, Z., and Zhao, J. (2019) *GmMAX2-D14* and *-KAI* interactions-mediated SLs and KAR signaling pathway play essential roles in soybean nodulation and nodule evolution in legume. *Plant J*. doi.org/10.1111/tpj.14545

Zhang G, Bahn SC, Wang G, Zhang Y, Chen B, Zhang Y, Wang X, **Zhao J**. (2019) *PLDα1*-knockdown soybean seeds display higher unsaturated glycerolipid contents and seed vigor in high temperature and humidity environments. *Biotechnology for Biofuels*. 12:9. doi: 10.1186/s13068-018-1340-4.

Manan, S., Ahmad, M.Z., Zhang, G., Chen, B., Haq, B.U., Yang, J. and Zhao, J. (2017a) LEAFY COTYLEDON 2 regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development. *Front. Plant Sci.* **8**, 1604.

### **W997: Soybean Genomics**

#### **Integrative Approach using RNA-Seq and Whole-Genome Resequencing Data to Detect Causal Variants Underlying Cool Temperature Stable High Protein in Soybean**

Tri D. Vuong<sup>1</sup>, Juexin Wang<sup>2</sup>, Weiwei Wang<sup>2</sup>, Gunvant Patil<sup>3</sup>, Babu Valliyodan<sup>4</sup>, Dong Xu<sup>2</sup>, Trupti Joshi<sup>5</sup>, Pengyin Chen<sup>6</sup>, Rouf Mian<sup>7</sup> and Henry T. Nguyen<sup>1</sup>, (1)University of Missouri, Columbia, MO, (2)Department of Electrical Engineering and Computer Sciences, University of Missouri, Columbia, MO, (3)University of Minnesota, St Paul, MN, (4)Lincoln University, Jefferson City, MO, (5)Institute for Data Science and Informatics, University of Missouri, Columbia, MO, (6)University of Missouri, Portageville, MO, (7)USDA-ARS, NC State University, Raleigh, NC

Cool temperature has been shown to negatively affect seed protein content in soybeans grown in the northern areas of the United States. In a previous study, several exotic soybean germplasm with stable protein content were identified when evaluated in the northern states. However, the genetic basis controlling this agronomically important trait remains unknown. The objectives of our study were to assess the expression of genes involved in response to cool temperature stress during seed developmental stages using RNA-Seq transcriptome profiles and to leverage whole-genome resequencing (WGRS) data to detect causal variants conferring stability of protein content. Six soybean genotypes were selected and grown under controlled temperature conditions in growth chambers. Early (S2) and late (S6) seed developmental stages were collected for RNA sequencing. In this study, an integrative approach was initiated by identifying high confident differentially expressed genes (DEGs) in cool temperature stable samples, followed by screening single nucleotide polymorphisms (SNPs) located in these DEGs from the WGRS data. Among a total of 764 SNPs, 212 were successfully detected to be associated with five candidate genes. Most of these were located in the 5kb upstream region (promoter) of these genes. Our analysis revealed that these causal variants altered the binding motifs in regulatory regions, suggesting that a regulatory promoter motifs may play vital roles in cool temperature stability in these soybean germplasm. This study presented a new approach to investigate soybean biological mechanism by an integration of RNA and DNA sequence data

### **W998: Soybean Genomics**

#### **Genetic Variation Underlying Seed Oil Content in Soybean**

Yan Li, National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China

Soybean is recognized as the largest oilseed crop that contributes to 61% of the world oilseed production and 28% of the total global vegetable oil consumption in 2018. The seed oil content has been increased during soybean improvement from landraces to released cultivars, and some genetic loci controlling oil content are likely subjected to selection. However, the genes underlying the selection are largely unknown.

Here, a population of 382 cultivated soybean accessions, including 187 landraces and 195 released cultivars, was used to screen the loci associated with seed oil content and likely subjected to selection. We identified 47 putative improvement-selective SNPs, and 25 of them overlap with the previously reported QTL for seed oil content. The 864 genes within the LD decay distance of these 25 SNPs were examined for their expression levels in different soybean tissues using RNA-seq data, and five genes were highly expressed in soybean seeds. Only one gene, *Glyma.15g049200 (GmSWEET39)*, showed much higher expression level in the seeds of high-seed-oil soybean variety than the low-seed-oil variety. Further analyses showed that the relative expression level of *GmSWEET39* was significantly correlated with soybean seed oil content. The sequence polymorphism in the promoter and coding region of *GmSWEET39* was significantly associated with oil content. The allelic effects of *GmSWEET39* on total oil content were confirmed in transgenic Arabidopsis, transgenic soybean hairy roots and soybean

recombinant inbred lines. The frequencies of its superior alleles increased from wild soybean to soybean landraces, and are much higher in released cultivars. These findings suggest that the sequence variation in *GmSWEET39* affects its relative expression and oil content in soybean seeds, and *GmSWEET39* has been selected to increase seed oil content during soybean domestication and improvement.

In addition to *GmSWEET39*, the 864 genes in the 25 loci subjected to selective sweeps and locate within the previously mapped seed-oil QTL regions were further analyzed. The genes which were predicted to be involved in lipid metabolism related pathways were identified, and their roles in fatty acid accumulation should be investigated in future researches. Furthermore, among the SNPs which showed significant association with seed oil content by GWAS, many of them have not been selected during soybean improvement. The superior alleles and their germplasm carriers identified in this study would be valuable resources for the genetic improvement of seed oil content in soybean breeding program.

### **W999: Statistical Genomics**

#### **The Limits of Recombination in Cultivated Barley**

**Peter L. Morrell**, Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN

To be uploaded!

### **W1000: Statistical Genomics**

#### **A Statistics, Machine Learning and Optimization Combined Approach Toward Identifying Hub Genes from Multi-Source Gene Expression Data**

**Hairong Wei**, Michigan Technological University, Houghton, MI

The rapid accumulation of vast amount of gene expression data from multiple tissues or environmental conditions in public data repositories makes it imperative to develop novel methods that can jointly model multi-source data sets for identifying common and unique hub genes. We here report that we have attempted to combine statistics, machine learning, and optimization methods to achieve this. We have developed a new graphic Gaussian model for jointly reconstructing multiple gene regulatory networks (JRmGRN), which highlight both common and hub genes in the networks, using gene expression data from multiple tissues or conditions. Under the framework of Gaussian graphical model, JRmGRN method constructs the GRNs through maximizing a penalized log likelihood function. We formulated it as a convex optimization problem, and then solved it with an alternating direction method of multipliers (ADMM) algorithm. The performance of JRmGRN was first evaluated with synthetic data and the results showed that JRmGRN outperformed several other methods for reconstruction of GRNs. We also applied our method to real *Arabidopsis* RNA-seq data from two light regime conditions and maize data from different tissues in comparison with other methods, and both common hub genes and some conditions/tissue-specific hub genes were identified with higher accuracy and precision.

### **W1001: Statistical Genomics**

#### **Plant Genotype-Phenotype (G2P) Association Discovery via Integrative Genome-Scale Biological Network & Genome-Wide Association Analysis**

**Patrick X. Zhao**, Noble Research Institute, Ardmore, OK

Understanding the mechanisms of genotype and phenotype (G2P) associations has been an essential, yet a challenging task in modern biology. The challenge lies in the high-dimensional gene variables and the complexity of gene regulations and interactions that collectively define particular phenotypes, also called traits. To facilitate the deciphering of how biological processes, pathways, and complex traits in plants are regulated at systems-level, enabling the discovery of G2P associations, we developed a series of tools and bioinformatics systems for analyzing genome-scale biological networks and genome-wide associations in plants. In this talk, I will present several recently develop tools, including (1) PEPIS:

a pipeline for estimating epistatic effects in quantitative trait locus mapping and genome-wide association studies; (2) two-dimensional (2D) association and integrative omics analysis in rice provides systems biology view in trait analysis; and (3) mining functional modules in heterogeneous biological, and genetic networks.

#### **W1002: Statistical Genomics**

##### **Deshrinking Ridge Regression for Linkage Analyses and Association Studies**

**Meiyue Wang**, University of California, Riverside, Riverside, CA

Quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS) are still the primary steps towards gene discovery. With the ever-growing number of genetic markers, more efficient algorithms for genetic mapping are still necessary. The urgency is more obvious in the big data era when QTL mapping is to be conducted simultaneously for thousand traits, e.g., transcriptomic and metabolomic traits. Efficient mixed model association (EMMA) and genome-wide efficient mixed model association (GEMMA) are the widely used methods for GWAS. An algorithm with high computational efficiency is badly needed. It is interesting to note that the test statistics of the ordinary ridge regression (ORR) have the same patterns across the genome as those obtained from the EMMA method. However, ORR has never been used for GWAS due to its severe shrinkage on the estimated effects. Here we introduce a degree of freedom for marker effect obtained from ORR and use it to deshrink both the estimated effect estimate and the standard error so that the Wald test of ORR is brought back to the same level as the Wald test of EMMA. The new method is called deshrinking ridge regression (DRR). By evaluating the methods under three different model sizes (small, medium and large), we demonstrate that DRR is efficient for all model sizes while EMMA only works for medium and large models. Furthermore, DRR detect all markers in a simultaneous manner instead of scanning one marker at a time. As a result, the computing time of DRR is about  $m$  (number of genetics variants) times faster than that of EMMA.

#### **W1003: Statistical Genomics**

##### **A Tutorial on the Beavis Effect for QTL Mapping and Genome-Wide Association Studies**

**Shizhong Xu**, Dept. of Botany & Plant Sciences, University of California, Riverside, CA

In comparative genomics, the estimated variance of a trait explained by a detected quantitative trait locus (QTL) is often biased upwards due to a significance test associated with the detected QTL. The bias is escalated for small samples and high thresholds of the test statistics used for QTL declaration. This phenomenon is called the Beavis effect because it was discovered by Bill Beavis, an American quantitative geneticist and plant breeder, in 1994 just a few years after the QTL interval mapping technology was developed. Ever since the Beavis effect paper was published, numerous manuscripts in QTL mapping with small samples may have been rejected by reviewers and journal editors using the Beavis effect as justification for the rejection, making Bill Beavis the most "hated person" in the QTL mapping community. The reviewers and editors may have abused the Beavis effect. It is a phenomenon in comparative genomics, only occurring if you compare your QTL mapping results with other studies of the same traits. If you do not compare, there is no Beavis effect. In this tutorial, I will review the theory behind the Beavis effect, describe a method to correct for the bias and extend the theory and technology to QTL detected from association populations under the mixed linear model (MLM) genome-wide association study (GWAS) framework.

#### **W1004: Strawberry Genomics**

##### **Resources for Strawberry Genomics, Genetics and Breeding Research in GDR, Genome Database for Rosaceae**

**Sook Jung**<sup>1</sup>, Taein Lee<sup>2</sup>, Chun-Huai Cheng<sup>2</sup>, Ping Zheng<sup>1</sup>, Katheryn Buble<sup>3</sup>, Jodi L. Humann<sup>1</sup>, Jing Yu<sup>1</sup>, Heidi Hough<sup>1</sup>, James Crabb<sup>4</sup> and Dorrie Main<sup>4</sup>, (1)Washington State University, Pullman, WA,

(2)Washington State University, Pullman, Pullman, WA, (3)Washington State University, Pullman, WA, (4)Department of Horticulture, Washington State University, Pullman, WA

The Genome Database for Rosaceae (GDR, <https://www.rosaceae.org>) strives to provide a one-stop online resource that facilitates basic, translational and applied research for many rosaceous crops including strawberry. Integrated data include curated genome sequences, genes, transcripts, genetic maps, markers, SNP array data, QTL, traits, germplasm, and publications, made accessible to browse, query and download through easy-to-use web interfaces and tools. Whole genome assembly data, along with functional annotation done by GDR team, are available for the cultivated octoploid strawberry and four genome assemblies for the diploid woodland strawberry through search pages, BLAST, and JBrowse. A new tool, MegaSearch, allows more powerful and flexible searches for all data types in GDR, allowing users to build datasets using various categories and customize data fields to view and download. Conserved syntenic regions across rosaceous genomes, identified using MCScanX, are accessible through synteny viewer, allowing data transfer across species. Also available is Breeding Information Management System (BIMS), a secure and comprehensive management system for breeders. In this presentation, we will highlight these new features and future development as well as provide an overview of existing resources in GDR for strawberry researchers.

#### **W1005: Strawberry Genomics**

##### **Strawberry Genomics: Whole Genome Assembly and Structure Analysis in Japanese Strawberry**

**Sachiko Isobe**<sup>1</sup>, Kenta Shirasawa<sup>1</sup>, Claire Chung<sup>2</sup>, Delbert Boncan<sup>2</sup>, Takuya Wada<sup>3</sup>, Hideki Hirakawa<sup>4</sup>, Eiji Yamamoto<sup>1</sup>, Fumi Maeda<sup>5</sup> and TingFung Chan<sup>6</sup>, (1)Kazusa DNA Research Institute, Kisarazu, Japan, (2)The Chinese University of Hong Kong, Hong Kong, (3)Fukuoka Agriculture and Forestry Research Center, Chikushino, Fukuoka, Japan, (4)Kazusa DNA Research Institute, Kisarazu, Japan, (5)Chiba Prefectural Agriculture and Forestry Research Center, Chiba, Japan, (6)The Chinese University of Hong Kong, Shatin, Hong Kong

Strawberry (*Fragaria × ananassa*) is allo-octoploid (2n=8X=56) and allogamous species. Whole genome assembly in a Japanese strawberry variety, 'Reikou' was performed by Denovo MAGIC 3.0 with paired ends (PE), mate pair (MP) and 10X genomics libraries. The total length of the assembled sequences was 1,406 Mb, consisted of 32,715 sequences and N50 of 3.9 Mb. Meanwhile, a high density SNP linkage map was constructed with a S1 mapping population of 'Reikou' by mapping SNPs on the IStraw 90K Axiom® Array and previously published SSR markers. The 'Reikou' linkage map was consisted of 11,608 loci spanning 2,827.4 cM of 31 linkage groups. A total of 62 (31 × 2 haplotypes) pseudomolecules were developed based on the linkage map in a total length of 1,125 Mb. By comparison with the linkage map, 40 miss-assembly was found in the scaffolds. Optical mapping was furthermore performed by Bionano Saphyr, and a total of 307 conflicts were identified between the pseudomolecule sequences and the Bionano genome contigs. Whole genome resequencing was performed for 35 strawberry varieties and three wild species. Kmer-frequency analysis revealed that recent strawberry varieties bred in Japan tends to be less heterozygosity.

This study was partly supported by the Science and technology research promotion program for agriculture, forestry, fisheries and food industry (27003A) and the NARO Bio-oriented Technology Research Advancement Institution (30020B) of the Ministry of Agriculture, Forestry, and Fisheries of Japan.

#### **W1006: Strawberry Genomics**

##### **Genomic Prediction of Hybrid Performance in Strawberry**

**Mitchell J. Feldmann**, Michael A. Hardigan, Randi A. Famula, Cindy M. López Ramirez, Glenn S. Cole and Steven J. Knapp, Department of Plant Sciences, University of California, Davis, Davis, CA

Genetic gains for yield have been dramatic in strawberry, an outcrossing allo-octoploid, and have played a pivotal role in the expansion of production over the last half century. Despite evidence for strong

directional selection for high yielding cultivars, genetic improvement continues to an effective method to advance crop yields in this highly heterozygous species hypothesized to harbor significant genetic load. Surprisingly, this hypothesis has not been tested, and the importance of heterosis has not been investigated in strawberry. Here, in a 14♀16♂ factorial population, we show that the genetics of yield components are primarily additive, and that high parent heterosis is essentially non-existent in this genetically diverse strawberry population, despite being hypothesized to drive genetic gains for yield in strawberry. Broad-sense heritability ranged from 0.64 to 0.87, whereas narrow-sense heritability ranged from 0.54 to 0.73 for yield components. High parent heterosis was non-significant for 95% of the hybrid progeny tested. We found that the yields of individual hybrids could be accurately predicted using genomic BLUP with additive effects alone and accuracies were not improved by incorporating dominance or epistatic effects. General combining ability was substantially more important than specific combining ability, which did not increase the accuracy of hybrid prediction. Our study shows that yield is highly heritable, well predicted by GCA, and an ideal target for genomic selection, especially since multiple-harvest yield phenotyping costs are exorbitant in strawberry.

**W1007: Strawberry Genomics**

**Genomic and Genetic Resources in Wild Strawberry and its Application Toward Understanding Mechanisms of Fruit and Stolon Development**

**Zhongchi Liu**, Dept. of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD

*Fragaria vesca*, the wild strawberry, is emerging as a model for the commercial garden strawberry to investigate gene function. The wild strawberry has a small and diploid genome and exhibits many of the same characteristics as the commercial strawberry. Further, abundant natural variations and chemical mutagen-induced mutants provide rich genetic materials for novel gene identification. Extensive transcriptome data and co-expression network analyses lay the foundation for addressing important biological questions in sexual and asexual reproduction. In addition, consensus co-expression network effectively predicts gene-gene interaction and function. We demonstrate that both forward and reverse genetic approaches including CRISPR genome editing, mapping by sequencing, and transient gene expression can be successfully utilized to link genes to important traits in the wild strawberry.

**W1008: Strawberry Genomics**

**Genomic Approaches to Improve Disease Resistance and Fruit Quality in Octoploid Strawberry**

**Seonghee Lee**<sup>1</sup>, Youngjae Oh<sup>1</sup>, Saket Chandra<sup>1</sup>, Ronald Tapia<sup>1</sup>, Natalia Salinas<sup>1</sup>, Jin-Hee Kim<sup>1</sup>, Yi-Tien Lu<sup>1</sup>, Sadiqshya Sharma<sup>1</sup>, Cheol-Min Yoo<sup>1</sup>, Sujeet Verma<sup>1</sup>, Jason Zurn<sup>2</sup>, Nahla Bassil<sup>2</sup> and Vance M. Whitaker<sup>1</sup>, (1)University of Florida, Wimauma, FL, (2)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR

Strawberry is an important commodity as well as a model plant species in the Rosaceae family. In the last years, genomics enabled breeding approaches have successfully applied for strawberry breeding to develop new cultivars. Throughout the genomics-assisted breeding, we were able to identify a number of important traits for fruit quality and disease resistance in cultivated strawberry. The genome of octoploid strawberry is highly complex and heterogeneous, and this greatly limits the molecular-assisted breeding in octoploid strawberry. However, recent advances in strawberry genomics and availability of octoploid reference genome sequences now allow us to identify candidate genes and develop subgenome (or gene) specific DNA markers for target QTL. The use of DNA markers can effectively introduce target characteristics into elite breeding materials via marker-assisted selection (MAS). In this presentation, it will be discussed what DNA tests and other genomic tools are available for cultivated strawberry, and how integrated molecular genomic approaches can be accomplished to improve fruit quality and disease resistance in strawberry.

**W1009: Strawberry Genomics**

## **Transposable Elements Shape the Transcriptional Landscape in Woodland Strawberry**

**Elizabeth Alger**, Michigan State University, East Lansing, MI

Transposable elements (TE) are an important major driving force in shaping the evolution of eukaryotic genomes. The majority of observed genome size variation across land plants is due to the proliferation of various TE families. Novel TE insertions may have deleterious effects on the host genome, and genomes have evolved various mechanisms, including RNA-directed DNA methylation, to suppress the activity of TEs. The epigenetic silencing of TEs by DNA methylation has also been shown to result in decreased expression of neighboring genes. Thus, the genome must balance the 'trade-off' of silencing TEs with negatively impacting the expression of its genes. Here, we examined the 'trade-off' hypothesis by examining a natural diversity panel of woodland strawberry (*Fragaria vesca*). We uncovered that genome sizes of diploid *F. vesca* from populations distributed across the northern hemisphere range between 184 to 408 million bases (Mb) with a median genome size of 229 Mb. This is among the greatest genome size variation (more than two-fold) documented in a single species that is not due to polyploidization. Furthermore, our analyses revealed that gene expression is not only negatively correlated with increased DNA methylation levels of nearby TEs, but that this has the potential to predict parental gene expression levels in intraspecific hybrids. Lastly, functional gene enrichment analyses suggest TEs may have played an important role in the adaptive radiation of this species and that unregulated TEs contributed to 'genome bloat' observed in some populations.

### **W1010: Strawberry Genomics**

#### **Identification of New Actors Influencing the Balance between Flowering and Runnering in Strawberry**

**Amélia Gaston**<sup>1</sup>, Marie Alonso<sup>1</sup>, Aline Potier<sup>1</sup>, Marc Labadie<sup>1</sup>, Frédéric Delmas<sup>2</sup>, Silvia Sabbadini<sup>3</sup>, Bruno Mezzetti<sup>3</sup>, Christophe Rothan<sup>1</sup>, Michel Hernould<sup>2</sup> and Béatrice Denoyes<sup>1</sup>, (1)INRA, Villenave d'Ornon, France, (2)University of Bordeaux, Villenave d'Ornon, France, (3)Università Politecnica delle Marche, Ancona, Italy

Plants can reproduce sexually or asexually. In strawberry, both reproduction modes take place jointly and both present agronomical interests: flowering success impacts fruit yield and asexual reproduction enables cultivar propagation via stolon production. Flowering and stolon production, are physiologically and genetically linked but these traits are in competition. In fact, compared to once-flowering genotypes that flower once a year and are good stolon producers, natural continuous-flowering mutants present an extended period of floral initiation and do not produce stolons or scarcely. These data suggest the interconnection of the regulatory networks involved in the balance between sexual and asexual reproductions. Understanding the gene network regulating this balance is of great importance since modification of this balance will affect plant architecture and so the fruit and daughter-plant yield.

Our objective was to identify new key actors involved in the gene network regulating the balance between sexual and asexual reproduction. We focused on the diploid *F. vesca* strawberry model but, because physiological processes and genome organization are well conserved and because *F. vesca* is the dominant subgenome in cultivated octoploid strawberry, insights on key actors identified in the diploid model is easily transferred to the cultivated model. Based on fine mapping and candidate gene approaches, we identified new key actors involved in the gene network regulating the balance between flowering and runnering. These new actors were characterized using functional characterization and plant architecture approaches. Our findings offer potential breeding targets to modulate flowering and runnering responses in cultivated strawberry.

### **W1011: Strawberry Genomics**

#### **Dissection of Key Genes Controlling Important Agricultural Traits by using the Diploid Strawberry *Fragaria vesca***

**Chunying Kang**, Huazhong Agricultural University, Wuhan, China

The wild diploid strawberry *Fragaria vesca* is an excellent model system for the Rosaceae family and the fruit development study. My research interest is to dissect the molecular mechanisms regulating flower and fruit development in strawberry. To this end, we have built a comprehensive transcriptome dataset generated from floral and fruit tissues, identified long non-coding RNAs and alternatively spliced genes in *F. vesca*. Making use of these data, we polished the gene annotation of the *F. vesca* genome. With these genomic and transcriptomic resources, our lab has been using EMS mutagenesis to make a mutant population of *F. vesca*, screening mutants on important agricultural traits, cloning the causative mutations, and investigating gene functions. Through this strategy, we have successfully identified more than 10 genes regulating flower development and fruit quality. One example is RAP that is responsible for the foliage and fruit coloration in strawberry. RAP encodes a glutathione S-transferase (GST) gene that mediates anthocyanin transportation. Among all the homologs in strawberry, RAP is most abundantly expressed in the ripening fruit. Transient expression assay demonstrated that RAP is the principal transporter of anthocyanins among the paralogs. Moreover, stable over-expression of RAP driven by the 35S constitutive promoter in rap not only restores anthocyanin accumulation in leaf petiole, but also results in strong coloration in fruit receptacle starting from early developmental stages independent of FveMYB10. In addition, knock-out of RAP by CRISPR/Cas9 resulted in no leaf petiole coloration in cultivated strawberry, being a promising tool for fruit color breeding. In summary, all the toolkits are available to identify new genes in *F. vesca*, genetically manipulate the homologous genes in cultivated strawberry, and finally create new varieties potentially used for breeding.

#### **W1012: Sugar Beet Workshop**

##### **Getting to the Root of Sugar Beet: Generalizable Gene Discovery *via* Agron-Omics**

**Mitch McGrath**<sup>1</sup>, Paul Galewski<sup>2</sup>, Andy Funk<sup>2</sup> and Kevin Dorn<sup>3</sup>, (1)USDA-ARS Sugarbeet & Bean Research Unit, East Lansing, MI, (2)Michigan State University, East Lansing, MI, (3)USDA-ARS, Fort Collins, CO

Out with the old! In with the new! Genomics is revolutionizing biology and is gradually coming to a beet near you. Where do we stand, where do we go, how do we get there are good questions to ask. Here we will take a brief look at the state of the art in beets, but perhaps more importantly, the potential for beet genome resources to generate insights on old questions and bring new questions to the fore. An old question ... why does beet produce luxurious amounts of sucrose? This foundational sugar beet trait is still mysterious, yet obviously responsive to breeding and selection, and therefore with a genetic basis. No doubt a signature exists in the sugar beet genome. And of the supernumary cambia? They look like tree rings, but form over the course of a single season, and not in a cylinder, but as a cone shaped tap root. Of the new ... where and how has genetic variation responsible for agronomic traits arose? Where might they be found or sought today? Clues are to be found in the genome of course, and some insights are available, many more remain. Much beet breeding targets disease resistance. Might genomics help in seeking new alleles/ genes/ pathways to not only address, but actually solve long standing disease issues faced routinely by growers? We all know better understanding will lead to better outcomes, and a complete description of the heritable processes in beet is still to emerge. Hopefully, we'll not have to select solely by intuition much longer, and via identifying every nucleotide in the beet pan-genome, identify numerous targets of opportunity for beet breeding grower sustainability.

#### **W1013: Sugar Beet Workshop**

##### **Potential of Novel Sequencing Technologies in Advancing Sugar Beet Genomics**

**Subhashini Srinivasan**, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Plant breeding and crop improvement are emerging as the biggest beneficiary of next-generation sequencing, demanding sequencing of hundreds of thousands of plants for discovery of signatures associated with varying traits and diseases. At present, the only model for scalable sequencing is reduced genome representation using technologies such as Genotyping-by-Sequencing and microarrays. The success of these technologies has only emphasized the need to look into cost-effective

models for the whole-genome sequencing of thousands of plants. Here, we will present the potential and future of whole-genome sequencing opportunities to benefit plant breeders.

**W1014: Sugar Beet Workshop**

**Cercospora Leaf Spot (CLS) affects Leaf Microbiome of Sea Beets**

**Chiara Broccanello**, University of Padova, Legnaro (Padova), Italy

**W1015: Sugar Beet Workshop**

**Building a Genomic Resource in *Beta maritima***

**Christopher M. Richards**, USDA ARS National Laboratory for Genetic Resources Preservation, Fort Collins, CO

**W1016: Sugar Beet Workshop**

**A Genome-Wide Association Study of the Sugar Beet Pathogen *Cercospora beticola* Identifies Novel Fungicide Resistance Mutations**

**Rebecca E Spanner**, North Dakota State University, Fargo, ND

*Cercospora* leaf spot (CLS) is caused by the fungus *Cercospora beticola* and is the most destructive foliar disease of sugar beet worldwide. The sterol demethylation-inhibiting (DMI) fungicides are one of the most important tools for managing CLS. DMI fungicides bind to and inhibit the cytochrome P450 enzyme CYP51 required for synthesizing ergosterol which provides integrity to the fungal cell membrane. Quantitative resistance to DMI fungicides has emerged in *C. beticola* populations due to their repeated and widespread use. In previous studies, isolates with higher EC50 values overexpressed *CbCYP51* compared to DMI-sensitive strains. However, no causal mutation has been found linked to this expression change. In order to identify mutations responsible for DMI resistance in *C. beticola*, a genome-wide association study was carried out. Illumina paired-end whole genome re-sequencing was performed for 194 unique *C. beticola* strains sampled from different fields in the Red River Valley sugar beet growing region in 2016 and 2017. Their sensitivity to the DMI tetraconazole was phenotyped as EC50 values calculated via agar plate growth. Genome-wide association identified a significant locus on Chromosome 8 in close proximity to *CbCYP51*. Two mutations at this locus are our top candidates underlying this resistance. CRISPR-Cas9 genome editing is being developed for *C. beticola* to elucidate which mutations are contributing to DMI resistance.

**W1017: Sugar Beet Workshop**

**BeetRES-MaBS: Mapping by Sequencing for Economically Relevant Traits in Sugar Beet**

**Elena Orsini**, Strube Research GmbH Co. KG, Söllingen, Germany and Elashry Abdelnaser, Strube Research GmbH & Co. KG, Daniela Holtgräwe, Boas Pucker, Bernd Weisshaar Chair for Genetics and Genomics of Plants, Center for Biotechnology, Bielefeld

Modern agriculture of sugar beet requires the breeding of multiple resistant traits in single-seeded cultivars. Improvement of resistance to the major diseases can be achieved via marker assisted selection applied in breeding programs. The BeetRES-MaBS project funded by the BMBF (Germany) focus on mapping candidate genomic regions for (1) tolerance toward the beet cyst nematodes (BCN) *Heterodera schachtii* and (2) resistance to the leaf pathogen *Cercospora beticola* (CB).

A biparental mapping population of 407 lines was obtained by crossing a monogerm DH line resistant to CB and a multigerm pollinator line tolerant to BCN.

The parental lines were deep sequenced and the reads were assembled, aligned against the reference genome and variants detection was performed. More than 600.000 SNPs polymorphic between the

parental lines and heterozygous in the F<sub>1</sub> individual were identified. A sparse marker panel was developed as KASP for genotyping the F<sub>2</sub> lines for QTL mapping.

For nematode tolerance, one major QTL was detected on chrom 5 that explained 25% of the phenotypic variance. This QTL was further validated in an independent breeding panel of sugar beet breeding material. QTL mapping for cercospora resistance revealed a major QTL on chromosome 4 detected in three of the four locations tested and explaining up to 30% of the genetic variance.

DNA of F<sub>2</sub> individuals at the extreme of the phenotypic distribution for the traits under consideration is currently combined into pools and sequenced with Illumina technology. Allele frequencies for variants will be determined for both pools. Deviations in allele frequencies between the pools should identify the genomic loci responsible for the quantitative traits. In order to pinpoint to the causal genes, the candidate regions will then be further analyzed with regard to e.g. functional gene annotation and the impact of variant positions in the diverse alleles.

#### **W1018: Sugar Cane (ICSB)**

##### **Leveraging Multiple Sequencing Technologies to Generate a Haplotype Specific Assembly of Sugarcane R570**

**Adam Healey**<sup>1</sup>, John Lovell<sup>2</sup>, Olivier Garsmeur<sup>3</sup>, Jerry Jenkins<sup>2</sup>, Jane Grimwood<sup>2</sup>, Karen Aitken<sup>4</sup>, Robert J. Henry<sup>5</sup>, Angélique D'Hont<sup>3</sup>, Jeremy Schmutz<sup>2</sup> and HudsonAlpha-Genome Sequencing Center, (1)HudsonAlpha Institute For Biotechnology, Huntsville, AL, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (3)CIRAD, UMR AGAP, Montpellier, France, (4)CSIRO Agriculture and Food, St Lucia, Australia, (5)University of Queensland/QAAFI, Brisbane, QLD, Australia

While sugarcane is one of the world's most important economic grasses for its sugar production and biofuel potential, tools and resources to understand its genetics are lacking. This is owed to the complexity of its genome which is highly polyploid, aneuploidy and heterozygous. Additionally, modern sugarcane cultivars are the result of interspecific hybridization and repeated backcrossing between domesticated *S. officinarum* and wild *S. spontaneum* parents. Cultivar R570 is best characterized sugarcane genome to date with the release of the BAC clone single tiling path, but this assembly represents a gene-rich and collapsed view of each of R570's homeologous chromosomes. To generate a haplotype specific assembly of R570, we devised a strategy that combines two *de novo* assemblies of R570 (Illumina; Pacbio), 96 selfed offspring (15X cov), single chromosome libraries and HiC to sequence and separate each homeologous chromosome. Using Illumina libraries, we generated a 5 Gb *de novo* genome assembly, using it to extract 55 Million unique 80bp genetic markers. Genotyping these markers in 96 selfed offspring isolated 1.9 million simplex (single dose) markers that were projected onto the 7.4 Gb PacBio assembly to generate a genetic map and anchor contigs onto separate linkage groups. Contigs that cannot be anchored by simplex markers will be ordered and oriented using HiC contact maps and single chromosome libraries. This strategy of combining multiple sequencing technologies will generate a more complete assembly for one of the most complex genomes to date in the Plant Kingdom.

#### **W1019: Sugar Cane (ICSB)**

##### **Sequencing the Transcriptome of R570 to Explore the Complexity of the Sugarcane Genome**

**Adhini Sudhindra Kumar Pazhany**<sup>1</sup>, Virginie Perlo<sup>2</sup>, Frikkie Botha<sup>3</sup>, Agnelo Furtado<sup>3</sup>, Angela O'Keeffe<sup>4</sup>, Ardy Kharabian Masouleh<sup>4</sup>, Robert Henry<sup>5</sup>, Karen Aitken<sup>6</sup>, Angélique d'Hont<sup>7</sup>, Adam Healey<sup>8</sup>, Jane Grimwood<sup>9</sup>, Kerrie Barry<sup>10</sup> and Jeremy Schmutz<sup>9</sup>, (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia, (2)University of Queensland, Brisbane, Australia, (3)University of Queensland/QAAFI, Brisbane, QLD, Australia, (4)QAAFI (Queensland Alliance for Agriculture and Food Innovation), The University of Queensland, Brisbane, QLD, Australia, (5)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia, (6)CSIRO Agriculture and Food, St Lucia, Australia,

(7)CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), Montpellier, France, (8)HudsonAlpha Institute For Biotechnology, Huntsville, AL, (9)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (10)Department of Energy Joint Genome Institute, Walnut Creek, CA

Sugarcane is a crop of unequivocal importance which can meet requirements for food, feed fiber and fuel. This crop, with a large wealth of genetic resources and a rich breeding history, has only a very nascent genomic history with a monoploid genome sequence. Numerous efforts are underway to unravel the mysteries of this complex crop with advances in sequencing technologies. Efforts to capture the functional part of the genome using transcriptomic studies have long been a priority. R570, a French cultivar has been the most suitable material for many sugarcane genomics studies. In an attempt to utilize the revolutionary technology of long read sequencing, we have carried out isoform sequencing of this representative cultivar with a Pac Bio sequel I system. Recently we have compared sequel II data of the same cultivar from various vegetative and reproductive tissues. A greater sequencing depth may benefit the sugarcane fraternity with a more complete transcriptome and accurate gene annotation. The present study aimed at comparing sequel I and sequel II data for R570 to harness the information contained in the large transcriptome resources and to find novel sources of variation in the light of these recent advances. The results may support future transcriptomic studies in sugarcane to make informed decisions on depth of sequencing and help to unravel the complexities of this transcriptome.

#### **W1020: Sugar Cane (ICSB)**

##### **Characterisation of Metabolic Regulation of Carbon Partitioning in the Sugarcane Culm through different Stages of Development using Transcriptome and Metabolome Data**

**Virginie Perlo**<sup>1</sup>, Frederik Botha<sup>2</sup>, Agnelo Furtado<sup>3</sup> and Robert J. Henry<sup>3</sup>, (1)University of Queensland, Brisbane, Australia, (2)QAAFI Queensland Alliance for Agriculture and Food Innovation - UQ University, St Lucia, QLD, Australia, (3)University of Queensland/QAAFI, Brisbane, QLD, Australia

Sugarcane has a high potential to be used to generate environmentally friendly by-products for use in chemical, pharmaceutical, medical, cosmetic and food industries. A crucial challenge for a long-term economic viability is to optimise the crop for production of a biomass composition that will ensure maximum economic benefit. Transcriptome data analysis provides a relevant explanation of phenotypic variances and gives a more accurate prediction of phenotypes than genomic information. This study of genetic variation in gene expression and correlations with metabolic data and phenotype relied on high-throughput methodology, measurement and analysis of 360 samples, 24 commercial sugarcane cultivars with different phenotypic characteristics at 5 different development stages with 3 replicates.

A multi-omic approach, with an integrated transcriptomics and metabolomics analysis may reveal details of biological mechanisms and pathways. A global view of transcriptional regulation and the identification of differentially expressed genes (DEGs) and metabolites may improve the feasibility of tailoring or engineering targeted biosynthetic pathways to improve the production of bio-products from sugarcane. We are using a profiling analysis workflow (pipeline) to generate empirical correlations between gene expression, metabolites, phenotypic traits and pathway analysis.

#### **W1021: Sugar Cane (ICSB)**

##### **Investigation of the UDP-Glucose Metabolism in Sugarcane (*Saccharum* spp. hybrids)**

**Patrick J Mason**<sup>1</sup>, Frikkie Botha<sup>1</sup>, Nam Hoang<sup>1</sup>, Annelie Marquardt<sup>1</sup>, Gabriella Papa<sup>2</sup>, Jenny C. Mortimer<sup>2</sup>, Agnelo Furtado<sup>1</sup>, Blake Simmons<sup>2</sup> and Robert J. Henry<sup>1</sup>, (1)University of Queensland/QAAFI, Brisbane, QLD, Australia, (2)Joint Bioenergy Institute, San Francisco, CA

The synthesis and degradation of UDP-glucose is central to the amount of carbon (C) moving into the major pools within sugarcane, specifically sucrose, cellulose and hemicellulose. Knowledge regarding the differences in UDP-glucose metabolism and in turn C partitioning throughout the major organs within the sugarcane plant is still limited. The major organs of sugarcane roots, leaves and intermodal tissues

utilise carbon in very different ways. Leaf tissue is a source tissue, meaning it will be a net exporter of carbon throughout its lifecycle, as it the primary photosynthetic tissue. Conversely, root and internodal tissue are net importers of carbon in the form of sucrose, by which a significant proportion of C is cleaved into UDP-glucose to drive cell wall expansion, hydrolysed into glucose driving respiration, or retained as sucrose for storage. An additional degree of difference exists between the major sink organs; internodes and roots, whereby, following cessation of cell wall deposition, sucrose is stored in internodal parenchyma tissue. This study displays the clear differences in UDP-glucose metabolism between the major sugarcane organs, leading to compositional difference in the insoluble and soluble biomass fractions. Also, via RNA-seq analysis, the transcriptional mechanisms behind these differences were also defined. The identification of genes associated these changes in UDP-glucose metabolism may have utility in altering carbon partitioning in sugarcane to better suit a diversity of applications outside of traditional sugar production.

#### **W1022: Sugar Cane (ICSB)**

##### **Genome-Wide Alternative Splicing Landscapes Modulated by Biotrophic Sugarcane Smut Pathogen**

**Renesh Bedre**<sup>1</sup>, Sonia Irigoyen<sup>1</sup>, Patricia D C Schaker<sup>2</sup>, Claudia B. Monteiro-Vitorello<sup>3</sup>, Jorge A Da Silva<sup>1</sup> and Kranthi K. Mandadi<sup>4</sup>, (1)Texas A&M AgriLife Research, Weslaco, TX, (2)Universidade Tecnológica Federal do Paraná, Toledo, Brazil, (3)University of São Paulo, Piracicaba, Brazil, (4)Texas A&M AgriLife Research and Extension, Weslaco, TX

Alternative splicing (AS) of pre-mRNA generates transcriptome and proteome diversity during growth, development, and stress responses in eukaryotes. In sugarcane (*Saccharum* spp.), genome-scale studies of AS are lacking, mainly due to the absence of a high-quality sequenced reference genome, sugarcane's large, complex genome, and aneuploidy and polyploidy of sugarcane cultivars. To identify AS landscape in sugarcane, we performed the isoform-level transcriptome (Illumina RNA-seq) analysis in sugarcane during infection with the smut fungus (*Sporisorium scitamineum*) using a hybrid approach by integrating *Sorghum bicolor* reference based and Trinity de novo mapping tools. Differential expression analysis detected 16,039 and 15,379 transcripts ( $\geq 2$  FPKM) at 5 and 200 days after infection, respectively. Overall, isoform level expression analysis suggested that approximately 5,000 (14%) sugarcane genes undergo AS and differential transcript analysis identified 896 AS events modulated at different stages of smut infection. Analysis of AS landscape revealed that alternative donor (AD) was predominant event followed by intron retention (IR) among the four major AS events. Gene family and gene ontology functional enrichment analysis of the differentially spliced genes revealed overrepresentation of functional categories related to the cell wall, defense, and redox homeostasis pathways. AS generated transcripts that have shown protein-level changes by gain and loss of functional domains. To our knowledge so far, this is the first study of AS in sugarcane and, demonstrates novel insight and bioinformatics approach to explore the AS landscape of sugarcane during smut disease interactions.

#### **W1023: Sugar Cane (ICSB)**

##### **Variation in Biomass Composition and Enzymatic Hydrolysis Efficiency in Sugarcane**

**Katrina J.M. Hodgson-Kratky**<sup>1</sup>, Gabriella Papa<sup>2</sup>, Virginie Perlo<sup>3</sup>, Agnelo Furtado<sup>4</sup>, Hemant Choudhary<sup>5</sup>, Frikkie Botha<sup>4</sup>, Blake Simmons<sup>6</sup> and Robert Henry<sup>7</sup>, (1)Queensland Alliance for Agriculture and Food Innovation, University of Queensland, BRISBANE, QLD, Australia, (2)University of Milan, Milan, Italy, (3)University of Queensland, Brisbane, Australia, (4)University of Queensland/QAAFI, Brisbane, QLD, Australia, (5)Joint Bioenergy Institute, Emeryville, CA, (6)Joint Bioenergy Institute, San Francisco, CA, (7)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia

Meeting future demands for renewable fuels and other bioproducts is dependent on the use of lignocellulosic biomass feedstocks from highly productive crops such as sugarcane. However,

the rigid and highly cross-linked cell wall impedes the efficient breakdown of biomass into fermentable sugars. Developing sugarcane varieties with altered biomass composition is vital for improving the efficiency of enzymatically hydrolyzing cellulose into glucose. With the aim of identifying the major biomass components that influence cell wall recalcitrance, enzymatic hydrolysis efficiency was measured in pretreated leaf and culm tissues of sugarcane genotypes that varied in fiber composition. The strongest negative associations with hydrolysis efficiency were observed for acid-insoluble lignin content, syringyl to guaiacyl (S/G) ratio and xylan content while acid-soluble lignin had a strong positive influence indicating that these traits should be the key focus for genetic modification and breeding. To accelerate the development of new sugarcane varieties, the genes controlling S/G ratio were investigated by sequencing the transcriptome of genotypes that had contrasting values for the trait. Differential expression between these genotypes was observed in genes encoding enzymes in the phenylpropanoid biosynthesis pathway, dirigent proteins and a number of secondary cell wall transcription factors which may be important targets for modifying S/G ratio. Furthermore, several sequence variants associated with S/G ratio were identified which may be promising markers for use in marker-assisted selection to advance the development of sugarcane varieties with improved enzymatic hydrolysis efficiency.

#### **W1024: Sugar Cane (ICSB)**

##### **Exploiting *Erianthus* Diversity to Enhance Sugarcane Cultivars**

**Nathalie Piperidis**, SUGAR RESEARCH AUSTRALIA, MACKAY, QLD, Australia, Chris Tom, SRA, te Kowai, QLD, Australia, Olivier Garsmeur, CIRAD, UMR AGAP, Montpellier, France, Karen Aitken, CSIRO Agriculture and Food, St Lucia, Australia, Felicity Atkin, SRA, Gordonvale, QLD, Australia and George Piperidis, Sugar Research Australia, Te Kowai, Australia

Introgression of *Erianthus arundinaceus* into the SRA sugarcane-breeding program has been a goal for researchers for many years. The *Erianthus* genome was finally accessible to sugarcane breeders with the identification in 2005 of the first *Saccharum/Erianthus* fertile hybrids, developed in China. Today, *Saccharum/Erianthus* BC3 and BC4 clones are available in Australia, and *Erianthus*-sugarcane hybrids have been characterised by cytogenetics and investigated for their potential resistance against pachymetra root rot, sugarcane smut and nematodes. Some clones have shown potential as new sources of resistance for incorporation into the SRA breeding program. These hybrids were created from *Erianthus* clones indigenous to China and their reaction to the above diseases is unknown in Australian conditions. In Meringa we also have access to many *Erianthus* clones of Indonesian origin. Some of these *Erianthus* clones have previously shown immunity to pachymetra root rot. In the late 1990s, these Indonesian *Erianthus* clones were used in crossing but no fertile hybrids were ever produced due to an incompatibility between the *Saccharum* and the *Erianthus* genomes. We revisited this untapped source of resistance by utilising the fertile *Erianthus* hybrids derived from China to cross with the Indonesian *Erianthus* of known resistance to pachymetra root rot. Here we report on the early stage results of introgressing Indonesian *Erianthus* into the SRA breeding program.

#### **W1025: Sweetpotato Genomics**

##### **Toward a Hexaploid Sweetpotato Pan-Genome**

**Zhangjun Fei**, Boyce Thompson Institute, Cornell University, Ithaca, NY

Sweet potato, *Ipomoea batatas* ( $2n=6x=90$ ), is among the most important food crops in the world and an extremely important food crop for subsistence farmers in sub-Saharan Africa. Despite its significant importance, currently knowledge of the genetic, molecular, and physiological basis of key agronomic traits in sweet potato is very limited, which has substantially constrained the improvement of sweet potato varieties across the world. The polyploidy (hexaploid) and highly heterozygous nature of the

sweet potato genome makes genetic and functional analyses extremely challenging. To facilitate research and breeding in this important crop, we assembled the genomes of *I. trifida* and *I. triloba*, two diploid wild relatives of sweet potato, and proved that these high-quality reference-grade genomes can serve as robust references for hexaploid sweet potato. Nonetheless, recent advances in sequencing and computational technologies have substantially increased our ability to assemble complex genomes. Phased hexaploid genome assemblies will undoubtedly better facilitate genome-enabled sweet potato breeding. Furthermore, Sweet potato varieties show wide phenotypic variations and display varying degrees of biotic/abiotic stress tolerances, implying that some important genes and alleles might be present only in certain genotypes. To maximize the capture of these genes and alleles, we are establishing a sweet potato pan-genome initiative with the main goal being to build a hexaploid sweet potato pan-genome by de novo assembling genomes of several representative sweet potato cultivars at the 'reference' level.

#### **W1026: Sweetpotato Genomics**

##### **Characterization of Sweetpotato Inheritance using Ultradense Multilocus Genetic Map**

**Marcelo Mollinari**<sup>1</sup>, Bode Adebawole Olukolu<sup>2</sup>, Guilherme Da Silva Pereira<sup>1</sup>, Awais Khan<sup>3</sup>, Dorcus C. Gemenet<sup>4</sup>, G. Craig Yencho<sup>5</sup> and Zhao-Bang Zeng<sup>1</sup>, (1)North Carolina State University, Raleigh, NC, (2)University of Tennessee, Knoxville, TN, (3)Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, NY, (4)International Potato Center (CIP), Lima, Peru, (5)North Carolina State University (NCSU), Raleigh, NC

The cultivated sweetpotato (*Ipomoea batatas* (L.) Lam.,  $2n = 6x = 90$ ) is an important staple food crop with an annual production of 112.84 million tons. Despite its undeniable social and economic importance, genetic studies in sweetpotato significantly lag behind major diploid crops due to its complex polyploid genome. To fully characterize the inheritance pattern in sweetpotato, we built an ultra-dense multilocus integrated genetic map of a full-sib population derived from a cross between the cultivars 'Beauregard' and 'Tanzania' (BT population) using our newly implemented software, MAPpoly. The resulting genetic linkage map consisted of 30,684 SNPs distributed in 15 homology groups with a total length of 2708.3 cM (11.3 SNPs/cM). We observed 96.5% collinearity between *I. batatas* and its diploid relative *I. trifida*. Using the genotypic probabilities computed across all linkage groups, we inferred the complete hexaploid haplotypes for all individuals in the offspring. We also observed that 73.3% of the meiotic configurations in the parents were resolved in bivalents, 15.7% presented multivalent signatures, and 11.0% were inconclusive. Moreover, the BT population presented vastly hexasomic inheritance mechanisms in all linkage groups, except for linkage group 2, which presented low levels of preferential pairing in parent Tanzania. We propose that the hexasomic-bivalent inheritance promotes stability to the allelic transmission in sweetpotato.

#### **W1027: Sweetpotato Genomics**

##### **Identification of QTL for Storage Root $\alpha$ - and $\beta$ -Amylase Activity in Sweetpotato at Harvest and during Post-Harvest Storage**

Victor A. Amankwaah<sup>1,2</sup>, Guilherme da Silva Pereira<sup>1</sup>, Marcelo Mollinari<sup>1</sup>, Bode A. Olukolu<sup>3</sup>, Sharon Williamson<sup>1</sup>, Xiaofei Zhang<sup>1</sup>, Van-Den Truong<sup>1</sup>, Edward Carey<sup>4</sup>, Zhao-Bang Zeng<sup>1</sup> and **G. Craig Yencho**<sup>5</sup>, (1)North Carolina State University, Raleigh, NC, (2)CSIR-Crops Research Institute, Kumasi, Ghana, (3)University of Tennessee, Knoxville, TN, (4)International Potato Center (CIP), Kumasi, Ghana, (5)North Carolina State University (NCSU), Raleigh, NC

The activity of  $\alpha$ - and  $\beta$ -amylase in sweetpotato storage roots is one of the major factors determining sweetpotato sweetness, texture and end-user preference. In the present study, the  $\alpha$ - and  $\beta$ -amylase activity of 248 genotypes in a bi-parental population derived from the cross of Tanzania x Beauregard (TB) was measured at harvest and during post-harvest storage replicated in two years. A high density genetic map with 14,813 markers was developed using MAPpoly. The linear association between 2016 and 2017 storage root amylase activity was moderate ( $r = 0.42-0.65$   $p < 0.001$ ) across uncured, cured and stored treatments. We identified four QTL each for  $\alpha$ - and  $\beta$ -amylases activity explaining 61.72% and

72.42% of the total phenotypic variation of each trait, respectively. QTL on LG 3 at a peak position of 29.82 cM for  $\alpha$ -amylase activity explained the highest percentage (25.09 %), followed by the QTL peak on LG 9 (21.28%) of the phenotypic variation. The QTL on LG 13 at peak position of 79.68 cM on LG 13 exclusively explained 54.36% of the phenotypic variation for  $\beta$ -amylase. Identified QTL on LG 9 for  $\alpha$ -amylase activity and QTL on LG 13 for  $\beta$ -amylase co-located with  $\alpha$ -amylase like and  $\beta$ -amylase annotated genes.

#### **W1028: Sweetpotato Genomics**

##### **Genetic Diversity and Population Structure of the USDA Sweetpotato (*Ipomoea batatas*) Germplasm Collection**

**Phillip A. Wadl**, USDA, ARS, U.S. Vegetable Laboratory, Charleston, SC, Bode Adebowale Olukolu, University of Tennessee, Knoxville, TN, Sandra E. Branham, Clemson University, Charleston, SC, Robert L. Jarret, USDA, ARS, Plant Genetic Resources Conservation Unit, D. Michael Jackson, USDA, ARS, U.S. Vegetable Laboratory and G. Craig Yencho, North Carolina State University (NCSU), Raleigh, NC

Sweetpotato, *Ipomoea batatas*, plays a critical role in food security and is the third most important root crop worldwide following potatoes and cassava. Sweetpotato is an important crop in the United States (US) and is valued at over \$700 million dollars annually. The sweetpotato germplasm collection of the US is maintained by the USDA, ARS, Plant Genetic Resources Conservation Unit and provides the genetic basis for sweetpotato crop improvement. In our study, population structure and genetic diversity of 417 sweetpotato accessions originating from 8 broad geographical regions (Africa, Australia, Caribbean, Central America, Far East, North America, Pacific Islands, and South America) were determined using over 30,000 single nucleotide polymorphisms (SNPs) using a genotyping-by-sequencing (GBS) protocol optimized for highly heterozygous and polyploid species. Bayesian clustering analyses (STRUCTURE) grouped the accessions into four genetic groups (Central American, North American, South American, other regions) and indicated a high degree of mixed ancestry. A neighbor-joining cladogram, principal components analysis, and a genetic distance matrix of the accessions supported the population structure analysis. Pairwise  $F_{ST}$  values between broad geographical regions based on the origin of accessions ranged from 0.017 (Far East – Pacific Islands) to 0.110 (Australia – South America) and supported the clustering of accessions based on genetic distance. The markers developed for use with this collection of accessions provide an important genomic resource for the sweetpotato community and contribute to our understanding of the genetic diversity present within the US sweetpotato collection and the species.

#### **W1029: Sweetpotato Genomics**

##### **Breeding 3.0 in Sweetpotato: Advances Towards Genome-Enabled Breeding in Sub-Saharan Africa**

**Dorcus C. Gemenet**<sup>1</sup>, Robert Mwangi<sup>2</sup>, Jolien Swanckaert<sup>2</sup>, Reuben T. Ssali<sup>3</sup>, Benard Yada<sup>4</sup>, Wolfgang Grüneberg<sup>5</sup>, Maria Andrade<sup>6</sup>, G. Craig Yencho<sup>7</sup> and Hugo Campos<sup>5</sup>, (1)International Potato Center (CIP), Nairobi, Kenya, (2)International Potato Center (CIP), Kampala, Uganda, (3)International Potato Center (CIP), Kumasi, Ghana, (4)National Agricultural Research Organization (NARO), Kampala, Uganda, (5)International Potato Center (CIP), Lima, Peru, (6)International Potato Center (CIP), Maputo, Mozambique, (7)North Carolina State University (NCSU), Raleigh, NC

Sweetpotato, [*Ipomoea batatas* (L.) Lam.], is a crop of increasing importance towards food and nutritional security especially in sub-Saharan Africa (SSA), due to its environmental resilience and ability to accumulate high levels of  $\beta$ -carotene, a precursor of vitamin A. As a complex, auto-hexaploid crop, genome-enabled breeding in sweetpotato remains a challenge for most breeding programs. Recently, foundational tools such as reference genomes, bioinformatic pipelines, and statistical genetic pipelines have been developed and continue to be improved to better represent this genome complexity. In this seminar, we discuss how such foundational tools are being adopted into specific breeding programs through trait discovery, quality assurance and control, and optimization for genomic selection. We also

discuss how multidisciplinary teams are working to ensure that adoption of such new tools is carried out within the realm of breeding program optimization and resource allocation, towards the overall purpose of increasing adoption and dissemination of novel sweetpotato varieties.

#### **W1030: Swine**

##### **The Porcine Muscle Thanatotranscriptome**

**Dan Nonneman**, Aaron M. Dickey and Andy King, USDA, ARS, USMARC, Clay Center, NE

While most gene expression studies for identification of meat quality candidate genes involve tissues collected immediately after harvest, several studies have shown that many genes are upregulated after death (thanatotranscriptome). It is generally accepted that anaerobic glycolysis is the primary postmortem metabolic pathway in the conversion of muscle to meat. However recent evidence shows that enough residual muscle oxygenation remains in postmortem muscle for hours to support mitochondrial function. This study was done to determine changes in gene expression with postmortem interval and how these genes and pathways are related to pork quality. RNAseq libraries were prepared from porcine longissimus muscle collected from five gilts ages 262-325 days, at 0, 24 and 48 hours after conventional harvest with electrical stunning and chilling. An average of 58.5 million paired-end reads were collected from each library, mapped to Sscrofa11.1 and differential gene expression determined using DESeq2. Compared to 0 hour samples, 4 and 1943 more highly expressed genes, and 132 and 2280 lower expressed genes were found at 24 and 48 hours, respectively, with log2 fold changes ranging from -7.15 to 2.55. The most overrepresented pathways included ribosomal protein, protein translation, oxidative phosphorylation and cytochrome-C oxidase activity. ELISA for 3 proteins (HSPA6, CCL21 and EPB42) with gene expression fold changes of 5.1, 2.94 and 4.4 at 48 hours showed protein content changes of 88.01%, 42.19% and -37.17% at 48 hours. These results imply that gene expression and protein translation continues to occur in postmortem muscle and could impact meat quality.

#### **W1031: Swine**

##### **From GWAS Peak to Causal Mutation; Utilizing p(ig)CADD Scores to Prioritize Sequence Variation**

**Martijn F.L. Derks**<sup>1</sup>, Christian Gross<sup>2</sup>, Marcos S. Lopes<sup>3</sup>, Marcel J.T. Reinders<sup>4</sup>, Mirte Bosse<sup>2</sup>, Arne B Gjuvsland<sup>5</sup>, Hendrik-Jan Megens<sup>2</sup>, Dick de Ridder<sup>2</sup> and Martien A.M. Groenen<sup>6</sup>, (1)Wageningen University & Research, Animal Breeding and Genomics, Wageningen, Netherlands, (2)Wageningen University, Wageningen, Netherlands, (3)Topigs Norsvin, Curitiba, Brazil, (4)Delft Bioinformatics Lab, Delft University of Technology, Delft, Netherlands, (5)Topigs Norsvin, Hamar, Norway, (6)Wageningen University & Research, Wageningen, Netherlands

The genotype-phenotype link is a major research topic in life sciences, but remains highly complex to disentangle. Part of the complexity arises from the polygenicity of phenotypes, in which many (interacting) genes contribute to the observed phenotype. Genome wide association studies have been instrumental to associate genomic markers to important phenotypes. However, despite the vast increase of molecular data (e.g. whole genome sequences), pinpointing the causal variant underlying a phenotype of interest is still a major challenge, especially due to high levels of linkage disequilibrium.

In this study we present a method to prioritize genomic variation underlying traits of interest from genome wide association studies in pigs. First, we select all sequence variants associated with the trait. Subsequently, we prioritize variation by utilizing and integrating predicted variant impact scores, gene expression data, epigenetic marks for promotor and enhancer identification, and associated phenotypes in other (well-studied) mammalian species. The power of the method heavily relies on variant impact scores, for which we used pCADD, a tool which can assign scores to any variant in the genome including those in non-coding regions. Using our methodology, we are able to either pinpoint the likely causal mutation or substantially narrow down the list of potential causal candidates from any association result. We demonstrate the efficacy of the tool by reporting known

and novel causal variants, of which many affect (non-coding) regulatory sequences associated with important phenotypes in pigs.

This study provides a framework to pinpoint likely causal variation and genes underlying important phenotypes in pigs. Hence, the tool accelerates the discovery of new causal variants that could be directly implemented to improve selection. Finally, we report several common pathways and molecular mechanisms involved in analogous phenotypes between human and pig, proving the suitability of pig as a model to study human (metabolic) disease.

## **W1032: Swine**

### **Phenomics of Pig Behavior and Estimation of Social Effects in Pigs**

**Juan P. Steibel**<sup>1</sup>, Belcy K. Angarita Barajas<sup>2</sup>, Janice M. Siegford<sup>1</sup>, Kaitlin Wurtz<sup>3</sup>, Catherine W. Ernst<sup>1</sup>, Rodolfo Cantet<sup>4</sup> and Simon Turner<sup>5</sup>, (1)Department of Animal Science, Michigan State University, East Lansing, MI, (2)Michigan State University, East Lansing, MI, (3)MSU, East Lansing, MI, (4)UBA, CABA, Argentina, (5)Scotland Rural University College, Easter Bush, Midlothian, United Kingdom

Behavioral phenotyping is a time-consuming endeavor whose application is usually restricted to small experimental settings. However, implementation of large scale behavioral phenomics is important for selection and for precision livestock management. We present the use of image analyses in conjunction with feeding records to collect feeding behavior data. And we also illustrate the use of behavioral records in improving genomic prediction models. First, we show how to reliably identify multiple occupancy of single feeder spaces (a proxy for competition). Deep learning was used to classify images of feeder occupancy with accuracy between 92% and 100%. Some multiple occupancy events were also detected through automatic feeding records when the double occupancy of the feeder lasted for longer than 25% of the total time a feeding event of one of the two pigs. Second, to illustrate the potential of incorporating social interaction data into models of genetic effects, we used behavioral observations from manually decoded video. We quantified reciprocal fights and single sided attacks and we used the duration of those interactions to parameterize social genetic effects matrices to model skin lesion counts in almost 800 growing-finishing pigs grouped in 59 pens. We show that the proposed approach recovered 40% to 80% more genetic variance compared to a model without social genetic effects, while traditional social genetics effects assuming uniform interactions between all group mates were not able to recover any additional variance. Automatic behavioral phenotyping will unlock the use of better social genetic effects modeling as well as precision livestock application.

## **W1033: Swine**

### **Updating the Annotation of the SLA Complex from the Genome Assembly Sscrofa 11.1**

Mathieu Charles<sup>1</sup>, Benjamin D. Rosen<sup>2</sup>, Hamid Beiki<sup>3</sup>, Christopher K. Tuggle<sup>4</sup>, Hiruni R Wijesena<sup>5</sup>, Daniel Ciobanu<sup>5</sup>, Joan K Lunney<sup>6</sup> and **Claire Rogel-Gaillard**<sup>1</sup>, (1)GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)Department of Animal Science, Iowa State University, Ames, IA, (4)Iowa State University, Ames, IA, (5)University of Nebraska - Lincoln, Lincoln, NE, (6)Animal Parasitic Diseases Laboratory, BARC, NEA, ARS, USDA, Beltsville, MD

The pig major histocompatibility complex (MHC), or swine leukocyte antigen (SLA) complex, maps to a large genomic region on chromosome 7. It includes the highly polymorphic series of class I and class II histocompatibility genes involved in peptide presentation and self-recognition, which vary in number and in alleles according to haplotypes. In order to ensure swine health, it is essential to characterize the complexity of the SLA complex and to preserve its diversity. Our aim was thus to update the annotation of the SLA complex from the Sscrofa11.1 assembly, as an opportunity to enrich our current knowledge on SLA and annotate a whole new haplotype. We found that the SLA complex (2.7Mb from *MOG* to *RING1*) is well-assembled, even in the highly duplicated regions that comprise the class I and II gene series. However, the Ensembl or NCBI-based automated annotations of these duplicated genes was not accurate and even misleading. We have reannotated the whole SLA genomic region using IsoSeq and

RNAseq data from tissues of the Duroc female used for the Sscrofa11.1 assembly. In total, we have refined the annotation of 27 SLA genes (12 class I genes, 15 class II genes) and more than 120 other genes. For class I and II genes, we have corrected the gene structures and names, as well as characterized their allele sequences and splicing variants. We also mapped SNPs, present on various available SNP chips, to report their distribution and will discuss their use in representation of SLA diversity.

**W1034: Swine**

### **High Throughput Phenotyping and Integrative Analysis of Feeding Behavior and Feed Efficiency Traits in Pigs**

**Wen Huang**, Michigan State University, East Lansing, MI

In this talk, I will present our work on the phenotypic and genetic characterizations of feeding behavior and feed efficiency traits in a dataset containing more than 5,000 Duroc pigs and 3.7 million feeding records. These characterizations were made possible by high throughput phenotyping using an automatic pig feeding system. We performed extensive data filtering and imputation before summarizing feeding records into phenotypes, including behavioral traits such as feeding frequency, duration, rhythm and feed efficiency traits such as feed conversion rate, residual feed intake. In addition, 50K SNP Chip genotypes were collected and imputed to whole genome sequences. Feeding behavior traits including the rhythm exhibited seasonality, suggesting an association between feeding and environmental factors such as light/dark cycle and temperature. By leveraging imputed whole genome sequences, we mapped genes for feeding behavior and feed efficiency traits with unprecedented resolution. Our study demonstrated the power of high throughput phenotyping and integrating multiple sources of large-scale data and may provide further insights into the genetics and biology of feeding and growth in pigs.

**W1035: Swine**

### **An Integrative GWAS and RNA-Seq Study to Identify SNPs and Transcripts Related to Sperm Quality Traits in Pigs**

**Marta Godia**<sup>1</sup>, Antonio Reverter<sup>2</sup>, Rayner González-Prendes<sup>3</sup>, Yuliaxis Ramayo-Caldas<sup>4</sup>, Anna Castello<sup>1</sup>, Joan E. Rodríguez-Gil<sup>5</sup>, Armand Sanchez<sup>1</sup> and Alex Clop<sup>6</sup>, (1)Animal Genomics Group, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Bellaterra (Barcelona), Spain, (2)CSIRO Agriculture and Food, Australia, (3)Wageningen University & Research, Netherlands, (4)IRTA, Caldes de Montbui, Spain, (5)Unit of Animal Reproduction, Department of Animal Medicine and Surgery, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain, (6)Centre for Research in Agricultural Genomics (CRAG), Cerdanyola del Valles, Spain

For the last decades, boars have been selected for their genetic merit on carcass and meat quality traits. However, breeders and researchers are now paying attention to additional phenotypes including sperm quality. The molecular processes affecting sperm quality remain largely unexplored. Genetic pressure in animal breeding is sparking the interest to select for boars with high sperm quality to maximize ejaculate doses and fertility rates. We identified candidate genes, pathways and DNA variants associated to sperm quality in swine by analyzing 25 sperm-related phenotypes with a systems biology approach combining GWAS with 288 boars and genotypes from the Axiom porcine high-density genotyping array and RNA-seq (total and small) from 40 of these pigs. With the GWAS, we identified 12 regions associated to head and neck abnormalities, abnormal acrosomes and motility. Candidate genes included *CHD2*, *KATNAL2* or *SLC14A2*. By RNA-seq, we detected 6,128 significant correlations between sperm traits and gene abundances. To build a robust gene network, only the pair-wise interactions present in both the SNP co-association and the RNA co-abundance networks were kept. The network also included genes which RNA abundances correlated with more than 4 traits. The final network contained genes involved in gamete generation and development, meiotic cell cycle, DNA repair or embryo implantation. A selection of 74 SNPs from the network, GWAS and eGWAS lead hits were used to build a SNP panel that explained between 5 to 36% of the phenotypic variance of these sperm quality traits.

**W1036: Swine**

**Quantitative Genetic Analysis of the Blood Transcriptome of Young Healthy Pigs to Improve Disease Resilience**

**Kyu-Sang Lim**<sup>1</sup>, Jian Cheng<sup>1</sup>, Christopher K. Tuggle<sup>1</sup>, Michael Dyck<sup>2</sup>, Frederic Fortin<sup>3</sup>, John Harding<sup>4</sup>, Graham S. Plastow<sup>2</sup>, Jack C. M. Dekkers<sup>1</sup> and PigGen Canada, (1)Iowa State University, Ames, IA, (2)Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (3)Centre de développement du porc du Québec inc., Québec, QC, Canada, (4)Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada

The complexity of gene expression is determined by not only the environment but also genetics. Here, we estimated the heritability of gene expression in the blood of young healthy piglets (~27 days of age) and its genetic correlation with measures of resilience after exposure to a natural polymicrobial disease challenge. Weaned barrows (n=3,205, Yorkshire x Landrace, in 50 batches) from healthy multiplier farms were evaluated for disease resilience in an experimental facility consisting of a high-health quarantine nursery and a challenge nursery and finisher. All pigs were genotyped with the 650k porcine genotyping array. Gene expression in blood samples collected in the quarantine nursery (n=903) was quantified by 3'mRNA sequencing. Average daily gain (ADG) over each stage for pigs that survived was evaluated as a resilience phenotype (qNurADG, n=3,138; cNurADG, n=2,784; cFinADG, n=2,341). Heritability estimates for qNurADG, cNurADG and cFinADG were 0.32 ( $\pm 0.04$ ), 0.25 ( $\pm 0.04$ ), 0.29 ( $\pm 0.05$ ), respectively. Among 15,872 evaluated genes, the expression of 292 and 1,487 genes had a high ( $h^2 > 0.4$ ) and moderate ( $0.4 > h^2 \geq 0.2$ ) estimates of heritability, respectively. The top 5,006 heritable genes were used to estimate genetic correlations with ADG. The numbers of genes that showed significant ( $p < 0.05$ ) genetic correlations with qNurADG, cNurADG, and cFinADG were 288, 199, and 191, respectively. These results provide new insight into the heritability of the porcine blood transcriptome, implicating its possible use in young healthy pigs as early predictors to improve disease resilience. Funding from USDA-NIFA, Genome Canada, Genome Alberta, and PigGen Canada.

**W1037: Swine**

**Combined Analysis Reveal Association of Changes in Gene Expression and H3K27ac Chromatin Modification at Regulatory Regions in Porcine Alveolar Macrophage in Response to LPS and Poly(I:C)**

**Juber L. Herrera-Uribe**, Department of Animal Science, Iowa State University, Ames, IA

Regulation of transcription is associated with changes in chromatin structure by histone modifications (HMs) associated with chromatin accessibility for transcription factors. Alveolar macrophages (AM) play an important role in host defense given that they are the most plastic cells of the immune system, giving them the ability to adapt and to provide an effective immune response against pathogenic microorganisms. By combining RNA sequencing (RNAseq) with chromatin immunoprecipitation and sequencing (ChIP-seq) for four HMs (H3K4me3, H3K4me1, H3K27ac and H3K27me3), we established the chromatin state map of AM, and investigated the potential regulatory effect of these chromatin modifications on RNA changes in AM stimulated with lipopolysaccharide (LPS) and Poly(I:C) at 2h and 6h. The integrative analysis suggests that the differential gene expression between non-stimulated and stimulated AM is significantly associated with changes in H3K27ac at active regulatory regions in the genome. Although globally changes to chromatin states were minor after stimulations at 2h and/or 6h, we found chromatin state changes for selected differentially expressed genes involved in TLR4, TLR3 and RIG-I signaling pathways. This could suggest that regulatory elements (i.e. active promoters) are already active/poised for immediate inflammatory response in porcine AM. In summary, our data reported here provides the first chromatin state map of AM in response to bacterial and viral mimics, contributing to the Functional Annotation of Animal Genomes (FAANG) project. Furthermore, this work demonstrates the role of HMs, especially H3K27ac, in macrophage response to LPS and Poly(I:C).

**W1038: Swine**

### **Jorgensen Award: Anthrax Toxin Receptor 1-Knockout Pigs are Protected from Senecavirus a Infection**

**Paula R. Chen**, Division of Animal Sciences, University of Missouri, Columbia, MO

Senecavirus A (SVA) has been the cause of numerous cases of vesicular disease in swine across the world in recent years. Studies investigating the oncolytic properties of SVA in humans revealed anthrax toxin receptor 1 (ANTXR1) as its probable receptor. The objective of the current study was to determine if ANTXR1 functioned as the receptor for SVA in pigs by employing the CRISPR/Cas9 system to edit exon 1 and create a premature stop codon. Two founder *ANTXR1*-knockout pigs and two age-matched wild-type pigs were challenged with SVA. Serum, fecal swabs, and nasal swabs were collected throughout the duration of the study. Presence of viral nucleic acid was determined by PCR, and SVA antibody responses were assessed. *ANTXR1*-knockout pigs had a distinct phenotype, including frontal bossing and wide, short statures, which is characteristic of GAPO syndrome in humans. The knockout pigs did not develop vesicular lesions while the wild-type pigs had coronary band lesions after SVA infection. Moreover, SVA nucleic acid was not detected in either *ANTXR1*-knockout pig, but virus was present in fecal and nasal swabs of one knockout pig. The same pig demonstrated evidence for production of SVA-specific antibodies; however, both knockout pigs did not exhibit virus neutralizing activity. Because founder pigs created by microinjection of the CRISPR/Cas9 system can have mosaic genotypes, a study on F1s is warranted. Overall, knocking out *ANTXR1* appears to confer protection against SVA infection in pigs, and modulation of this region may be needed to correct the phenotype associated with the edit.

**W1039: Swine**

### **Jorgensen Award: Protein Levels in Blood of Young Healthy Pigs as Indicators of Disease Resilience**

**Yulu Chen**, Iowa State University, Ames, IA

**W1040: Swine**

### **Iowa State University Swine Workshop Station Report**

**Christopher K. Tuggle**, Animal Science, Iowa State University, Ames, IA

This talk will summarize the swine research work in the Tuggle lab in 2019. The talk will include research in two areas; (1) the epigenetic analysis of gene expression in tissues and cells of the pig as part of the Functional Annotation of Animal Genomes Consortium, and (2) research on characterizing a novel mutation in swine that causes Severe Combined Immune Deficiency. For (1), I will update on the newly funded FAANG pig Genome project and discuss relevance to swine genomics. In (2), I will provide a summary of recent results including creation and analysis of a new SCID mutant line and demonstration that the SCID pig is an excellent model for xenograft testing.

**W1041: Swine**

### **Whole-Genome DNA Methylation Analyses in Pig Fetal and Immune Tissues**

**Ryan J. Corbett**, Genetics & Genome Sciences Graduate Program, Michigan State University, East Lansing, MI and Catherine W. Ernst, Department of Animal Science, Michigan State University, East Lansing, MI

DNA methylation is an epigenetic modification occurring almost exclusively at CpG dinucleotides in mammalian genomes, and is known to regulate gene expression through the alteration of transcription factor binding sites and chromatin conformation. DNA methylation plays important roles in development, genomic imprinting, and stress response; however assessment of genome-wide methylation beyond healthy adult pig tissues is currently lacking. Our efforts are currently focused on using whole-genome bisulfite sequencing (WGBS) to characterize patterns of DNA methylation in pig fetal tissues of economic relevance, and in a variety of immune cell types at both adult and weaning stages. In a pilot study in

*longissimus dorsi* muscle, we observed global hypomethylation during myogenic progression that was associated with changes in expression of both DNA methylation enzymes and differentially methylated genes. Our current efforts are expanding upon developmental profiling of DNA methylation in additional fetal tissues (brain, liver, placenta) and will identify regions of allele-specific methylation. We have also generated WGBS and RNA-sequencing data in piglet peripheral blood mononuclear cells (PBMCs) before and after weaning to assess effects of a naturally occurring stressor on gene regulation. To further increase annotation in the immune system, we are generating WGBS data in nine immune cell types from healthy adult male pigs, and in alveolar macrophages stimulated with lipopolysaccharide and Poly-IC. This work will contribute to FAANG objectives by increasing the annotation of steady-state DNA methylation patterns in pig tissues, and will identify novel sites of gene regulation during fetal development and immune response.

**W1042: Swine**

**NCSU Station Report**

**Christian Maltecca**, NC State University, Raleigh, NC

**W1043: Swine**

**Genomic Analysis of Weaning to Estrus Interval from Crossbred Sows Spanning 4 Parities**

**Gary A. Rohrer**, USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, Clay Center, NE, Amanda J. Cross, DNA Genetics, Columbus, NE and Lea A. Rempel, USDA, ARS, USMARC

Weaning to estrus interval (WEI) is an important phenotype in swine production as it is a critical component to defining non-productive days in the sow herd. Therefore, a genomic analysis of weaning to estrus interval was conducted in a Landrace-Yorkshire composite population where all animals were genotyped using a 60k Illumina Beadchip. Sows had the opportunity to have up to 4 parities and the time between day of weaning and the first estrus detected post-weaning was defined as WEI. WEI was analyzed for each parity separately fitting fixed effects of sire breed, birth month, farrowing group and covariates for number born alive and number weaned. Genomic heritability was low for first parity sows (0.018) but was between 0.044-0.083 for later parity sows. Number of genomic regions explaining > 2.0% of genomic variation were 3 (parity 3), 4 (parities 1 and 2) and 7 (parity 4). There were no regions associated with more than one parity. In conclusion, heritability of WEI increases with parity indicating that phenotypic variation in first parity sows is largely controlled by non-genetic factors. Furthermore, GWAS analysis indicates that each parity appears to be controlled by a unique set of genes, although this finding may be due to limited number of animals with later parity data. Selection to reduce non-productive days should focus on WEI in later parities to achieve greater progress.

USDA is an equal opportunity provider and employer.

**W1044: Swine**

**Distinguishing Fetal and Placental Immune Responses to Congenital Infection with Porcine Respiratory and Reproductive Syndrome Virus with Nanostring Arrays**

**Joan Lunney**, Animal Parasitic Diseases Laboratory, BARC, NEA, ARS, USDA, Beltsville, MD

Porcine reproductive and respiratory syndrome virus (PRRSV) infections cause major reproductive losses, with an estimate of over \$300 million annual losses in the U. S. alone. Joint studies of ARS scientists at Beltsville, Maryland, with scientists at the University of Saskatchewan, have probed responses to PRRSV infection in third-trimester pregnant gilts, and assessed maternal and fetal factors that could be predictive of PRRS severity and resilience in fetal pigs. The expression of immune-related genes in fetuses with no, low or high viral load at 5 to 12 days post maternal PRRSV infection was investigated. Differential expression (DE) of genes was evaluated using a 230 gene NanoString array (designed on biomarkers previously predicted to alter PRRS resistance and susceptibility). Based on log viral load PLC and fetal thymus (THY) samples were assigned to 3 experimental groups: ND (none

detected), LOW and HI viral load. The resulting data were normalized, and a univariate analysis conducted using a Generalized Least Squares (GLS) model with false discovery rate (FDR) correction. In the PLC a total of 52 genes were found to be significantly upregulated between ND and HI. In the THY a total of 197 and 84 genes were found to be differentially expressed between ND-LOW and ND-HI, respectively. 35 genes were found to be commonly DE for PLC and THY, with response to type 1 interferons (IFNs) including upregulation of STAT1-3 and IFN response genes (IFIT1-3, IFIH1, IRF1&5 and GBP). Efforts are continuing to assess the impact of viral load in PLC and thymuses to distinguish the effect of viral infection and cross placental transmission on fetal survival and local immune responses. These studies have affirmed the diversity of fetal pig anti-PRRSV response within each litter and have set the stage for more detailed analyses now underway to probe for key markers of fetal pig PRRS resilience.

**W1045: Swine**

**University of Nebraska - Lincoln Station Report**

**Daniel C. Ciobanu**, University of Nebraska - Lincoln, Lincoln, NE

**W1046: Systems Biology and Machine Learning**

**Elucidating and Re-Designing Regulatory Networks Underlying Plant-Pathogen Interaction**

**Katherine Denby**, University of York, York, United Kingdom

Plant responses to biotic stress involve large-scale transcriptional reprogramming. We are elucidating the gene regulatory networks underlying these transcriptional responses to pathogen infection using a combination of experimental and computational/mathematical tools. We generated high-resolution time series expression data from Arabidopsis leaves following infection with bacterial and fungal pathogens. These time series data sets have enabled us to identify transient changes in gene expression and resolve the chronology of plant defence responses. We have generated transcriptional network models predicting regulatory relationships between differentially expressed transcription factors and identified key regulators of the Arabidopsis defence response from our networks. Crucially many of these key regulators were not previously known to affect susceptibility to plant pathogens. We have applied this time series-based network analysis gene discovery strategy to lettuce to predict genes conferring disease resistance against two fungal pathogens, *Botrytis cinerea* and *Sclerotinia sclerotiorum*, and speed up the breeding of these traits.

**W1047: Systems Biology and Machine Learning**

**Comparing Time Series Transcriptome Data between Plants using a Network Module Finding Algorithm**

**Jiyoung Lee**, Virginia Tech, Blacksburg, VA

**Background**

Comparative transcriptome analysis is the comparison of expression patterns between homologous genes in different species. Since most molecular mechanistic studies in plants have been performed in model species, including Arabidopsis and rice, comparative transcriptome analysis is particularly important for functional annotation of genes in diverse plant species. Many biological processes, such as embryo development, are highly conserved between different plant species. The challenge is to establish one-to-one mapping of the developmental stages between two species.

**Results**

In this manuscript, we solve this problem by converting the gene expression patterns into co-expression networks and then apply network module finding algorithms to the cross-species co-expression network.

We describe how such analyses are carried out using bash scripts for preliminary data processing followed by using the R programming language for module finding with a simulated annealing method. We also provide instructions on how to visualize the resulting co-expression networks across species.

## Conclusions

We provide a comprehensive pipeline from installing software and downloading raw transcriptome data to predicting homologous genes and finding orthologous co-expression networks. From the example provided, we demonstrate the application of our method to reveal functional conservation and divergence of genes in two plant species.

### **W1048: Systems Biology and Machine Learning**

#### **Multiscale Modeling of Lignin Biosynthesis in *P. trichocarpa***

**Cranos Williams<sup>1</sup>**, Megan L. Matthews<sup>2</sup>, Jack P. Wang<sup>1</sup>, Ronald Sederoff<sup>1</sup> and Vincent L. Chiang<sup>1</sup>, (1)North Carolina State University, Raleigh, NC, (2)University of Illinois, Urbana Champaign

Understanding the mechanisms behind lignin formation is an important research area with significant implications on the bioenergy and biomaterial industries. Computational models are indispensable tools for understanding this complex process. Models of the monolignol pathway in *Populus trichocarpa* and other plants have been previously developed and used to explore metabolic regulation and how transgenic modifications important bioenergy traits. However, it is still unclear how these modifications propagate through the different biological layers, resulting in changes to lignin and other wood properties. We developed a multi-scale model spanning the transcript, protein, metabolic, and phenotypic layers of monolignol biosynthesis in *P. trichocarpa*. The three main components of this multiscale model are (1) a transcript-protein model, (2) a kinetic-based metabolic model, and (3) random forest models relating the steady-state metabolic fluxes to 25 lignin and wood physical traits. For the transcript-protein model, we used a sparse maximum likelihood approach to capture cross-influence behavior between the monolignol specific transcripts and proteins under a series of systematic transgenic knockdowns. Using *in-silico* simulations and root-mean square error, we show that including these cross-influences improve our ability to estimate the transcript and protein abundances when individual and families of monolignol genes are perturbed. The estimated proteins drive a dynamic metabolic model, and we use the resulting steady-state fluxes and machine learning approaches to predict 25 lignin and wood traits. We show that using a random forest approach results in lower errors and better  $R^2$  values than multiple linear regression. Further, including the cross-influences between transcripts and proteins results in smaller predictive error for 23 of the 25 traits when emulating the transgenic knockdowns. Our approach provides an *in-silico* multi-scale model that can be used to explore the predicted behavior of novel combinatorial monolignol gene knockdowns.

### **W1049: Systems Biology and Machine Learning**

#### **Transcriptome-Based Prediction of Complex Traits in Maize**

**Shin-Han Shiu**, Michigan State University, East Lansing, MI

The ability to predict traits from genome-wide sequence information (i.e. genomic prediction), has improved our understanding of the genetic basis of complex traits and transformed breeding practices. Transcriptome data may also be useful for genomic prediction. However, it remains unclear how well transcript levels can predict traits, particularly when traits are scored at different development stages. Using maize genetic markers and transcript levels from seedlings to predict mature plant traits, we found transcript and genetic marker models have similar performance. When the transcripts and genetic markers with the greatest weights (i.e. the most important) in those models were used in one joint model, performance increased. Furthermore, genetic markers important for predictions were not close to or identified as regulatory variants for important transcripts. These findings demonstrate that transcript levels are useful for predicting traits and that their predictive power is not simply due to genetic variation in the transcribed genomic regions. Finally, genetic marker models identified only one of 14 benchmark

flowering time genes, while transcript models identified five. Highlighting that, in addition to being useful for genomic prediction, transcriptome data can provide a link between traits and variation that cannot be readily captured at the sequence level.

**W1050: Systems Biology and Machine Learning**

**Resolution of Alternative Maize Genome Annotations by Machine Learning**

Laura de Boer<sup>1</sup>, Zhouxin Shen<sup>1</sup>, Nathan M. Springer<sup>2</sup> and Steven Briggs<sup>1</sup>, (1)University of California San Diego, La Jolla, CA, (2)University of Minnesota, St. Paul, MN

Annotated genomes each contain a proposed set of protein-coding genes. Most current genome annotations are missing some true genes and contain some false genes; true or false means able to express proteins or not. These errors in genome annotation can interfere with interpretations of genetic, genomic, transcriptomic, or proteomic data. Maize researchers are fortunate to have four alternative sets of gene models that annotate the genome of reference inbred B73 (MaizeGDB v2, MaizeGDB v4, NCBI, UniProt). We found that a machine learning method can accurately distinguish silent genes from expressible genes using DNA methylation or histone marks (PMID: 31420517). We subsequently used this method to classify the gene models in the four alternative sets and combined them into a non-redundant set of expressible genes. All genes not included and for which we had observed proteins were added to create a gene set called v4-Pro. Comparison of the four alternatives to v4-Pro using a large, independent proteomics data set showed that v4-Pro contains the largest number of true genes and the smallest number of false genes.

**W1051: Systems Biology and Machine Learning**

**Exascale Biology: From Genome to Climate with a Few Steps Along the Way**

Daniel Jacobson, Oak Ridge National Laboratory, Oak Ridge, TN

**W1052: Systems Biology and Machine Learning**

**DeepTE: A Computational Method for *de novo* Classification of Transposons with Convolutional Neural Network**

Haidong Yan, School of Plant and Environmental Sciences, Virginia Tech, USA, Blacksburg, VA, Song Li, Virginia Tech, Blacksburg, VA and Aureliano Bombarely, Department of Bioscience University of Milan, Milan, VA, Italy

Transposable elements (TEs) classification is an essential step to decode their roles in genome evolution. With a large number of genomes from non-model species becoming available, accurate and efficient TE classification has emerged as a new challenge in genomic sequence analysis.

We developed a novel tool, DeepTE, which classifies unknown TEs using convolutional neural network. DeepTE transferred sequences into input vectors based on of k-mer counts. A tree structured classification process was used where eight models were trained to classify TEs into super families and individual TE order. DeepTE also detected domains inside TEs to correct false classification. An additional model was trained to distinguish between non-TEs and TEs in plants. Given exclusive TEs of different species types, DeepTE classified seven orders, and 15, 24, and 16 super families in plants, metazoans, and fungi, respectively.

DeepTE outperformed other existing tools for TE classification in our benchmarking experiments. This tool successfully leverages convolutional neural network for TE classification, assisting to precisely identify and annotate TEs in newly sequenced eukaryotic genomes.

**W1053: Systems Biology and Ontologies**

## **Computational Identification of Target-Binding Sites in Plant Immune Genes**

**Daniil Prigozhin**, Lawrence Berkeley National Laboratory, Berkeley, CA, Janina Tamborski, University of California Berkeley, Berkeley, CA and Ksenia Krasileva, Department of Plant and Microbial Biology UC Berkeley, Berkeley, CA

Plant NLR immune receptors act as intracellular monitors that respond to pathogen attack by either directly interacting with pathogen molecules or indirectly recognizing their presence via pathogen-induced molecular changes to native plant proteins. Recent advances in targeted resequencing of NLRs enable the genome-wide analyses of sources of novel recognition specificities in plant immune genes. We used over 11,000 NLRs from the recently published dataset spanning over 60 accessions of the model plant *A. thaliana* to identify receptor subfamilies that show the highest sequence diversity; these accounted for roughly a quarter of the Arabidopsis NLRs. The identified highly variable NLRs included the known autoimmune NLR loci, highlighting the costs associated with derivation of new specificities. By analyzing the protein sequence alignments of the highly variable NLRs using information entropy, we identified the residues most likely to define recognition specificity. As expected, these were concentrated in the leucine-rich repeat domain. Protein structure modeling indicated that a subset of these formed continuous surfaces that contained exposed hydrophobic residues, a characteristic feature of protein-protein binding sites. In summary, our work identifies the sources of new specificity within plant immune genes and provides a foundation for engineering of new recognition specificities.

### **W1054: Systems Biology and Ontologies**

#### **Predicting Phenotype from Multi-Scale Genomic and Environment Data using Neural Networks and Knowledge Graphs: An Introduction to the NSF GenoPhenoEnvo Project**

**Anne Thessen**<sup>1</sup>, Michael Behrisch<sup>2</sup>, Emily Cain<sup>3</sup>, Remco Chang<sup>4</sup>, Bryan Heidorn<sup>3</sup>, Pankaj Jaiswal<sup>5</sup>, David LeBauer<sup>3</sup>, Ab Mosca<sup>4</sup>, Monica C. Munoz-Torres<sup>6</sup>, Arun Ross<sup>7</sup> and Tyson Swetnam<sup>3</sup>, (1)Oregon State University, Corvallis, OR, (2)Utrecht University, Netherlands, (3)University of Arizona, (4)Tufts University, (5)Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, OR, (6)Lawrence Berkeley National Laboratory, Berkeley, CA, (7)Michigan State University

To mitigate the effects of climate change on public health and conservation, we need to better understand the dynamic interplay between biological processes and environmental effects. Machine learning (ML) methods in general, and Deep Learning (DL) methods in particular, are a potential way forward because they are able to cope with the nonlinearity of natural systems. However, there are several barriers that exist, including the opaque nature of the algorithm output and the absence of ML-ready data. We propose to develop a machine learning framework capable of predicting phenotypes based on multi-scale data about genes and environments. A critical part of this framework is a visualization system to contextualize the results of an ML model, that is, to examine model decisions, connect decisions to input samples, and test alternative decisions. Further, we will develop data transformation methods that map the heterogeneous input data, ranging from simple vectors to complex images, into formats that are consumable by the ML techniques. The central hypothesis of this research is that deep learning algorithms and biological knowledge graphs will predict phenotypes more accurately across more taxa and more ecosystems than do current numerical and traditional statistical modeling methods. Our long term goal is to develop predictive analytics for organismal response to environmental perturbations using innovative data science approaches. This pilot project on predicting emergent properties of complex systems and multidimensional interactions is funded by the NSF (Award # 1940330).

### **W1055: Systems Biology and Ontologies**

#### **Genes and Gene-like Sequences in Maize, Sorghum, and Arabidopsis.**

**James C Schnable**, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE

The number of gene models identified based on homology or molecular evidence in plant genomes has grown over time. However, in many cases reverse genetics efforts have been unsuccessful at linking these gene models to organismal functions. Validated gene models with known loss of function phenotypes in maize and arabidopsis show a consistent structural, molecular, and evolutionary signature relative to the overall population of annotated gene models. Genes linked to phenotypes by conventional quantitative genetics show similar, but weaker signatures, while multi-trait-multi-marker methods identify larger populations of gene models with many of the same features which characterize genes with known loss of function phenotypes. Machine learning algorithms trained on structural, molecular, and evolutionary features can distinguish both classical mutants and genes identified via quantitative genetic analyses from the overall population of gene models. This creates the potential to reinvigorate reverse genetics as a tool for genetic investigation by prioritizing those gene sequences most likely to produce observable phenotypes when disrupted.

**W1056: Systems Biology and Ontologies**

**Are All Plastids Created Equal?**

Ryan W. Christian, Washington State University, Pullman, WA, Eric H. Roalson, School of Biological Sciences, Washington State University, Pullman, WA and **Amit Dhingra**, Molecular Plant Sciences, Washington State University, Pullman, WA

Plastids are the defining organelle for a plant cell and are critical for a myriad metabolic function. We are interested in characterizing the function of plastids in non-model crops, especially as it pertains to chromoplasts. An ultrastructural analysis approach was used to characterize fruit plastids in the epidermal and collenchymal cell layers at eleven developmental time points in three genotypes of apple (*Malus × domestica* Borkh.). This study enabled the identification of discrete ontology during which specific functions are most likely being performed by the plastids as indicated by accumulation of plastoglobuli, starch granules and other sub-organelle structures. Another overarching question is how plastid proteomes vary temporally, spatially, and taxonomically both within and across multiple species. A bioinformatics workflow was developed and evaluated on the predicted proteomes of 15 sequenced plant genomes. Between 628 – 828 protein families were found to have conserved plastid targeting across Angiosperm genera with each genus also reporting species-specific plastid targeted proteins. In plants, the plastid and its various morphotypes import a large and varied number of nuclear-encoded proteins to orchestrate vital biochemical reactions in a spatiotemporal context. Recent comparative genomics analysis and high-throughput shotgun proteomics data indicate that there are a large number of plastid-targeted proteins that are either semi-conserved or non-conserved across different lineages. This implies that homologs are differentially targeted across different species, which is feasible only if proteins have gained or lost plastid targeting peptides during evolution. A broad, multi-genome analysis of 15 phylogenetically diverse genera and in-depth analyses of pangenomes from Arabidopsis and Brachypodium were performed to address the question of how proteins acquire or lose plastid targeting peptides.

**W1057: Systems Biology and Ontologies**

**Can We Use Machine Learning to Predict Circadian Genes in Wheat Using DNA Sequence?**

**Laura-Jayne Gardiner**, IBM Research, Warrington, United Kingdom

We focus on the biological process of circadian regulation that has been found to underpin many agronomic traits in wheat, a key crop of global importance. Genes involved in the circadian clock show rhythmic expression patterns of approximately 24 hours that can be defined using parameters such as period, phase and amplitude. We use a 48-hour time course transcriptomics dataset generated by the Earlham Institute to identify 30,065 high confidence genes that are likely to be circadian in wheat.

We demonstrate the use of machine learning approaches for classification of the time-series expression profiles of genes into one of five classes [morning-circadian, day-circadian, evening-circadian, night-circadian or non-circadian] with an average accuracy of 85%. Furthermore, our accuracy is maintained (80%) using only 12 of the 24 timepoints available, where other commonly used tools showed an accuracy of only 64.5%. Now we report our exploration of the possibility of assigning genes into our five classes based on DNA sequence using features such as enriched putative regulatory DNA elements or motifs, SNPs or epigenetic marks.

This methodology can be applied to a wide range of genomics problems to reduce the time, cost and effort to identify patterns using predictive models. This work was supported by the STFC Hartree Centre's Innovation Return on Research programme, funded by the UK Department for Business, Energy & Industrial Strategy and is part of an ongoing collaboration between IBM Research, UK, and the Earlham Institute. IBM brings cutting-edge computational science alongside applicable technologies to support UK research.

**W1058: Systems Biology and Ontologies**

**Novel Maize Ear Phenotyping Methods Enable Large-Scale Functional Validation of Transcriptome Data**

**Cedar Warman**, Oregon State University, Corvallis, OR

Abundant transcriptome data has supported unprecedented description and analysis of organismal RNA dynamics. However, linking such data to organismal functions and phenotypes has proved a challenging problem. Here, we demonstrate a novel maize ear scanning system that addresses one aspect of this challenge. Our scanning system creates a 2D projection of the ear, which can then be processed using an internally-developed computer vision and machine learning pipeline to identify kernel locations and corresponding markers. Marker-linked mutations can thus be easily tracked across generations, with the large number of kernels assessed providing statistical sensitivity. As a test case, we measured the fitness cost of dozens of exon insertion mutations in highly expressed pollen genes and identified several genes with male-specific transmission defects. Regression analysis showed a positive correlation of a gene's transcript level with the corresponding mutation's likelihood for association with non-Mendelian segregation, presumably a consequence of reduced fitness via loss of gene function. Thus, our approach enables large-scale assessment of male-specific fitness phenotypes and the testing of these phenotypes for quantitative relationships with transcriptomes, demonstrating the potential for functional validation of large gene sets.

**W1059: Systems Biology and Ontologies**

**Gene Structure and Expression of the Rice S-Domain Kinase (SD-RLK) Gene Family**

**Sushma Naithani**, Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR and **Daemon Dikeman**, Oregon State University, Corvallis, OR

Receptors-like kinases (RLKs) are major players in perceiving and transducing extracellular signals into appropriate cellular responses and have been associated with nearly every aspect of plant growth and development, plant reproduction, and how a plant responds to pathogens and abiotic stress conditions present in its environment. The RLKs are encoded by members of one of the largest gene families, the RLK gene family (~600 in *Arabidopsis* and 1,429 in *rice*). The RLKs differ greatly in their extracellular ligand-binding domain and are divided into several sub-families. The the S-Domain RLKs (SD-RLKs) that share extracellular domain structure with Brassica S-locus receptor kinase represent the second largest RLKs subfamily. We present a detailed analysis of the gene structure and expression of SD-RLK gene family members from rice (*O. sativa japonica*).

**W1060: Systems Genomics**

**Maize Hybrids over-Express Chloroplast Digenomic Protein Complexes**

Laura de Boer<sup>1</sup>, Devon Birdseye<sup>2</sup>, Zhouxin Shen<sup>1</sup>, **Nathan M. Springer**<sup>3</sup> and Steven Briggs<sup>1</sup>,  
(1)University of California San Diego, La Jolla, CA, (2)University Of California San Diego, La Jolla,  
CA, (3)University of Minnesota, St. Paul, MN

In green leaves of maize, nearly all proteins encoded by the plastid genome are subunits of digenomic protein complexes that include subunits from the nuclear genome (e.g., the plastid ribosome, PSI, PSII, cyt b6/f, and ATP synthase). We have found that each member of these digenomic protein complexes is expressed above mid-parent levels in hybrids. In contrast, most proteins in the chloroplast and in the leaf as a whole are expressed at mid-parent levels. The digenomic proteome patterns may account for the greater photosynthetic capacity of hybrids. The patterns were not reflected in companion RNAseq data suggesting that the molecular phenotype of hybrids arises from post-transcriptional mechanisms. Multiple hybrids have been examined and in each case one parent expressed all of the digenomic proteins, both nuclear and plastid-encoded, above the levels of the other parent, indicating that proteome dominance is associated with the molecular phenotype of hybrids.

#### **W1061: Systems Genomics**

##### **Network-Based Feature Selection Associates Gene Modules in Co-Expression Network to Phenotypic Data**

**Chi Zhang**, University of Nebraska - Lincoln, Lincoln, NE

The association of gene co-expression network and phenotypic data has potential to identify causal genes, but there is no standard method to associate stress phenotype with gene co-expression networks. A novel method for the integration of the gene co-expression network and stress phenotype data was developed to conduct a systems analysis to link genotype to phenotype. We applied LASSO method to the gene co-expression network of rice with salt stress to discover key gene interactions for salt-tolerance related phenotypes. Submodules in gene modules being identified from the co-expression network were selected by the LASSO regression, which establishes a linear relationship between gene expression profiles and physiological responses, i.e. sodium/potassium condenses under salt stress. Genes in these modules have functions related to ion transport, osmotic adjustment, and oxidative tolerance. Our method identified a set of submodules in a rice gene co-expression network constructed with rice transcriptome data set under salt stress. We argued that these genes in submodules are biologically meaningful and useful for studies on rice salt tolerance. The method can be applied to other studies to efficiently and reliably integrate co-expression networks and phenotypic data.

#### **W1062: Systems Genomics**

##### **Empirical Comparison of Tropical Maize Hybrids Selected through Genomic and Phenotypic Selections**

**Yoseph Beyene Aydagn**<sup>1</sup>, Manje Gowda<sup>2</sup>, Michael Olsen<sup>3</sup>, Kelly Robbins<sup>4</sup>, Paulino Perez-Rodriguez<sup>5</sup>, Gregorio Alvarado<sup>6</sup>, Kate A. Dreher<sup>7</sup>, Star Yanxin Gao<sup>4</sup>, Stephen Mugo<sup>3</sup>, Prasanna Boddupalli<sup>1</sup> and Jose Crossa<sup>8</sup>, (1)CIMMYT, NAIROBI, Kenya, (2)CIMMYT (International Maize and Wheat Improvement Center), Nairobi, Nairobi, Kenya, (3)CIMMYT, Nairobi, Kenya, (4)Cornell University, Ithaca, NY, (5)Colegio de Postgraduado, Texcoco, Mexico, (6)International Maize and Wheat Improvement Center, Texcoco, Edo. de Mexico, Mexico, (7)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (8)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, EM, Mexico

Genomic selection predicts the genomic estimated breeding values (GEBVs) of individuals not previously phenotyped. The main objectives of this study were to (1) empirically compare the performance of tropical maize hybrids selected through phenotypic selection (PS) and genomic selection (GS) under well-watered (WW) and managed drought stress (WS) conditions in Kenya, and (2) compare the cost-benefit analysis of GS and PS. We used two experimental maize data sets (stage I and stage II yield trials). The stage I data set consisted of 1492 DH lines genotyped with rAmpSeq SNPs. A subset of these lines (855) crossed with a single-cross tester and the resulting hybrids were evaluated under WW

and WS conditions for grain yield and other agronomic traits, while the remaining 637 lines were predicted using the 855 lines as a training set. The second data set consists of 348 DH lines from the first data set of which 172 lines selected based on GEBVs, and 176 lines based on phenotypic performance. The 348 DH lines were crossed with three testers and the resulting 1042 hybrids and checks were evaluated across locations. The top 15% hybrids advanced through GS and PS gave 21-23% higher grain yield under WW and 51-52% more grain yield under WS than the mean of the checks. The GS reduced the cost by 32% over the PS. We concluded that the use of GS for yield in maize can produce selection candidates with similar performance as those generated from conventional PS, but at a lower cost.

**Key words:** phenotypic selection; genomic selection; genetic gain; maize; well-watered and water stress environments.

### **W1063: Systems Genomics**

#### **Long-Read RNA Sequencing Reveals Transcriptome Complexity in *Sesamum indicum* L.**

**Libin Wei**, Hongmei Miao and Haiyang Zhang, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, China, Zhengzhou, China

As an important oilseed crop, sesame (*Sesamum indicum* L.,  $2n = 26$ ), is known as the 'Queen of the oil seeds' for its high oil content and quality. In this study, the full length transcriptome of 8 samples including root, stem, leaf, bud and seeds with various development stages were sequenced using the long-read single-molecule Oxford Nanopore MinION sequencing technology (ONT RNA-seq) at the first time. In total, 104.44 million clean reads (138.32 Gb) were obtained with an average of 1,324 bp per read. A total of 7,034 novel genes and 86,776 novel transcripts were obtained after comparing with the sesame reference genome. Additionally, we identified 77,646 novel open reading frames (ORFs), 1,921 long non-coding RNAs (lncRNAs), and 8,724 transcription factors (TFs). Moreover, a total of 33 fusion transcripts, 93,300 alternative polyadenylation (APA) and 50,249 alternative splicing (AS) events were detected. Overall, our results not only offer a comprehensive view of sesame transcriptome, but also reveal the transcriptome complexity in sesame.

### **W1064: Systems Genomics**

#### **Exploring Alternative Genome-Wide Approaches to Analyze SNP Data for Complex Traits in Beef Cattle**

**Camila Urbano Braz**<sup>1</sup>, Henrique Nunes de Oliveira<sup>2</sup>, Troy N. Rowan<sup>1</sup>, Robert D. Schnabel<sup>3</sup> and Jared E. Decker<sup>1</sup>, (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Sao Paulo State University, Jaboticabal, Brazil, (3)Division of Animal Sciences, Informatics Institute, Columbia, MO

Exploring genomic information is crucial to increase the odds of detecting biomarkers underlying phenotypic variation. We present two genome-wide association (GWA) studies using different approaches to search for additive and non-additive loci influencing complex traits in beef cattle. The first study aimed to compare single-SNP with haplotype-based GWA analyses using 413,355 SNPs for meat tenderness (WBSF) in 3161 Nelore bulls. The results showed that GWA analyses using overlapping sliding window haplotypes provide substantially more power to detect QTL than does single-SNP analysis. Analyses performed with smaller haplotype windows (three and five SNPs) detected higher proportions of QTL. Thirty-seven candidate genes that participate in myogenesis, neurogenesis, lipid and fatty acid metabolism, and skeletal muscle structure or composition processes were identified influencing WBSF. In the second study, we searched for QTL enriched for genotype-by-environment interaction (GxE) for birth, weaning and yearling weights (BW, WW, and YW, respectively) using ~850,000 SNPs in ~13,500 Simmental cattle by conducting: direct GxE GWA analyses using continuous environmental variables (temperature, precipitation, and elevation), and in combination (U.S. ecoregions); and variance-heterogeneity GWA (vGWA) analyses, indicative of interactions, using residuals adjusted for additive, dominance, and epistatic relationships. GxE contributed to 10%, 4%, and 3% of the phenotypic variance of BW, WW, and YW, respectively. Genes were related to response to stimulus, nitrogen compound,

gene expression, development and metabolic processes. Twenty-two vQTL (difference in variance between genotypes) were detected. Some vQTL were enriched with GxE effects while one vQTL was also a QTL (difference in the mean between genotypes). These studies reveal the importance to investigate alternative approaches using genomic information to identify loci contributing to genetic control of complex traits.

**W1065: Systems Genomics**

**The Genes Controlling a Quantitative Trait are Multiple-Fold more Likely to Form a Co-Expression Network in Plants**

Meiping Zhang<sup>1</sup>, Yun-Hua Liu<sup>1</sup>, Wenwei Xu<sup>2</sup>, C. Wayne Smith<sup>1</sup>, Seth Murray<sup>1</sup> and **Hongbin Zhang<sup>1</sup>**, (1)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, (2)Texas A&M AgriLife Research, Lubbock, TX

Most traits or biological processes of agricultural importance, such as crop yield, crop quality and plant responses to biotic and abiotic stresses, are quantitative traits or complex traits and controlled by numerous genes that individually may have small effects on the trait. However, it remains unclear about the molecular mechanisms underlying polygenic traits or biological processes. Here, we report whether and how the genes controlling a polygenic trait or biological process are related to shape the trait or biological process performance and whether their relationship, if any, is related with its phenotype. The genes controlling three polygenic traits or biological process randomly selected from ginseng, cotton and maize were analyzed. We found that 75% or more of the genes controlling each of these traits or process were spliced into multiple transcripts and significantly enriched; nevertheless, only one to four, but not all, of their transcripts were correlated with its phenotype. The genes controlling a polygenic trait or biological process were multiple-fold more likely to form a co-expression network than other genes in an organ. The network varied substantially among genotypes and was associated with their phenotypes. These findings indicate that the genes controlling a polygenic trait or a biological process are more likely pleiotropic and functionally interacted, and provide knowledge necessary to develop advanced technologies for efficiently manipulating polygenic traits and for genome-wide identifying the candidate genes controlling polygenic traits in a species.

**W1066: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

**Delivering Bioinformatics Lectures and Hands-on Training using State-of-the-Art Solutions**

Subhashini Srinivasan, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India and **Saurabh Babanrao Whadgar**, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

Advances in streaming technology and cloud computing is enabling teachers to deliver LIVE lectures from the comfort of their homes to eager students anywhere in the world. These lectures can now be synchronized with hands-on training on AWS cloud. All a student may need is a laptop with access to internet and a keen interest in mastering bioinformatics. In this talk we will demonstrate the use of twitch streaming technology to deliver concepts/algorithms used in bioinformatics and allow access to UNIX servers on AWS cloud to receive hands-on training in areas including genome assembly, gene expression analysis, variant calling and metagenome analysis.

**W1067: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

**Galaxy as an Educational Tool and Community Resource for Undergraduate Training**

**Mohammad Heydarian**, Galaxy Project - Johns Hopkins University, Baltimore, MD

**W1068: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

## **Enhancement of Experiments in Artificial Selection: QTL Analysis of Loci that Condition Expression of Variable Traits in Rapid-Cycling *Brassica rapa***

**Scott Woody**, UW-Madison, Madison, WI

Experiments in artificial selection are a staple of the biology laboratory curriculum, but students are only rarely provided insight into the nature of the loci and associated allelic variants that might account for the response to selection they imposed on parental populations. To overcome that deficiency, we conducted QTL analysis of phenotypically variable traits in *Brassica rapa* that are readily quantified and so suitable as targets for selection. Specifically, we used two *B. rapa* RIL populations (R500 x IMB211 and R500 x FPsc) to identify loci where allelic variation conditions expression of seed coat color (SCC), anthocyanin pigments in hypocotyls (*AN*), trichome abundance on leaf margins (*TC*), flowering time (*DTF*) and height to first flower (*HtFF*). Our consideration of QTL peak data was aided considerably by the availability of *de novo* sequence assemblies of both FPsc and the common maternal parent R500. As expected, QTL peaks were evident on in both RIL populations, consistent with prior evidence showing that insertion of a *HELITRON* transposable element in a *B. rapa* orthologue of Arabidopsis *TRANSPARENT TESTA8* (*TT8*) largely accounts for the bright yellow seeds of R500 and the brown seeds of IMB211 and FPsc parental varieties. Similarly, both RIL populations yielded *AN* QTL peaks near an ortholog of Arabidopsis *PRODUCTION OF ANTHOCYANIN PIGMENTS2* (*PAP2*). However, the genetic architecture underlying expression of *TC*, *DTF*, and *HtFF* traits were surprisingly divergent. The results of all QTL analyses will be described, as well as the potential of these new data to enhance student learning outcomes.

## **W1069: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

### **Integration of Research and Teaching in Plant Genetics and Genomics: Two Decades Experience**

**Khalid Meksem**, Southern Illinois University at Carbondale, Carbondale, IL

## **W1070: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

### **Large Scale CRISPR Projects for First-Year Students**

**Burnette Jim**, University of California, Riverside, CA; UC-Riverside, Riverside, CA

Biology 20: The Dynamic Genome Course is a quarter-long authentic research experience offered to first-year undergraduates at the University of California, Riverside. The course introduces students to the Central Dogma and genome organization. They also master experimental techniques such as PCR, agarose gel electrophoresis, basic bioinformatics, bacterial transformation, and notebook keeping. Around week four of the quarter, the students transition to the research project which is led by a UCR professor. They professor works with the course instructors to devise protocols and teaching material that fit in the meeting schedule. Currently eight sections per quarter of 24 students are offered. This large scale allows professors to develop screens for collections of CRISPR induced alleles or design and make guide RNAs constructs to target large numbers of genes. Twice the class has screened loci in transformed plants for induced mutations. The students germinated the seeds in the first week and then later extracted DNA, amplified the loci with provided primers, and analyzed the final results. Sequencing of amplicons is done *en masse* using a PacBio strategy that avoids heterozygosity issues with Sanger sequencing and will be described in the Sunday morning session. In a carefully guided approach, students generated guide RNA constructs to over 70 genes and transformed them into Arabidopsis, with gene targeted by at two students. The screens successfully identified new alleles and transformed plants were recovered for every targeted locus. All three projects were funded as greater impacts for NSF grants.

## **W1071: The Analysis and Role of the Microbiome**

### **Impact of Host Genetics on the Rumen Microbiome**

**Leluo Guan**, University of Alberta, Edmonton, AB, Canada, Edmonton, AB, Canada

**W1072: The Analysis and Role of the Microbiome**

**Microbiome Enrichment in Controlled Environment Vegetable Production**

Michael Lynch<sup>1</sup>, Elena Zaikova<sup>1</sup>, Aashi Saraf<sup>2</sup>, Jiujun Cheng<sup>1</sup>, John Heil<sup>1</sup> and **Trevor C. Charles**<sup>3</sup>,  
(1)Metagenom Bio, Canada, (2)University of Waterloo, Canada, (3)University of Waterloo, Waterloo, ON, Canada

Controlled environment systems, such as hydroponic greenhouses and vertical farms, offer unprecedented opportunities for ensuring favourable microbiome composition for plant growth, product yield and quality, and resilience against disease-causing microbes. Microbiomic and metagenomic surveys of the recirculating nutrient delivery systems of commercial vegetable greenhouses demonstrate clear, crop-specific effects on microbial community structure. The conditions of the rhizosphere are expected to have microbial community enrichment effects. The ability to degrade 1-aminocyclopropane-1-carboxylic acid (ACC), the immediate precursor to ethylene, is associated with many rhizospheric and endophytic bacteria that have plant beneficial effects. The key enzyme is ACC deaminase, which catalyzes the conversion of ACC to ammonia and  $\alpha$ -ketobutyrate. As a result, ACC can serve as nitrogen and carbon source, and the resulting reduction of ACC levels reduces stress ethylene in the plants. Of interest is the role that ACC plays in shaping the phytobiome, and how this in turn may influence crop health and productivity. As an initial step towards understanding the effects of ACC metabolism, we used 16S rRNA gene sequence analysis and shotgun metagenomics to investigate the community dynamics of soil and hydroponic nutrient solution enrichment cultures with ACC as nitrogen source, compared to ammonia. We found that the community became much more constrained on ACC, consistent with ACC metabolism being more of a specialized trait. The ACC-enriched cultures were able to promote plant growth. Metagenome-assembled genomes (MAGs) and genomes of pure culture isolates confirmed the presence of *acdS*, encoding ACC deaminase, and provided insight into the nature and diversity of ACC metabolizing strains in recirculating nutrient delivery systems. These enrichment experiments lay the groundwork to guide strategies for microbiome optimization in operating hydroponic systems.

**W1073: The Analysis and Role of the Microbiome**

**Long Reads for Non-Model Metagenomes: From Structural Variation to Pangenomes in Colonization and Infection**

**Olga Francino**, Autonomous University of Barcelona, Barcelona, Spain

**W1074: The Analysis and Role of the Microbiome**

**Metagenomic Profiling of Thousands of Samples to Aid Selection of Environmentally Friendly Ruminants**

Melanie K. Hess<sup>1</sup>, Larissa Zetouni<sup>1</sup>, Juliana Budel<sup>1</sup>, Tracey Van Stijn<sup>1</sup>, Hannah Henry<sup>1</sup>, Michelle Kirk<sup>2</sup>, Rudiger Brauning<sup>1</sup>, Alan McCulloch<sup>1</sup>, **Andrew S Hess**<sup>1</sup>, Graham Wood<sup>1</sup>, Sharon M. Hickey<sup>3</sup>, Arjan Jonker<sup>2</sup>, Patricia L Johnson<sup>1</sup>, Mark Camara<sup>4</sup>, Peter Janssen<sup>2</sup>, John C. McEwan<sup>5</sup> and Suzanne J. Rowe<sup>1</sup>,  
(1)AgResearch Limited, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)AgResearch, Grasslands Research Centre, Palmerston North, New Zealand, (3)AgResearch, Ruakura Research Centre, Hamilton, New Zealand, (4)DairyNZ Limited, Hamilton, New Zealand, (5)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand

Ruminant livestock are a major source of protein for humans and play an important role in feeding the world. Methane is an undesirable by-product of feed breakdown in the rumen, a process driven by ruminal microbes. Reducing ruminant methane emissions is a critical step in reducing the environmental impact of agriculture. Individual animals have stable and heritable differences in methane emissions, and these are linked to differences in their rumen microbiota. Selection of ruminants based on rumen microbial profiles has the potential to reduce methane emissions, however, for selection to occur successfully, microbial profiles on thousands of

animals will be needed. We have developed a low-cost, high-throughput sequencing and bioinformatic approach for generating microbial profiles based on restriction-enzyme reduced representation sequencing, followed by a reference-based or reference-free bioinformatic pipeline for generating microbial profiles. Our reference-based pipeline uses BLAST to assign reads against the Hungate 1000 Collection of rumen microbial genome assemblies, and the reference-free pipeline counts the abundance of a set of common 65bp sequences in each sample. We generated microbial profiles on over 4,500 sheep and cattle rumen samples using this approach, with samples on some individuals taken at different ages and on different diets. We have evaluated the impact of different host factors (species, genetics, age) and environmental factors (cohort, diet) on these rumen microbial profiles. These results are being used to develop an approach for incorporating rumen microbial profiles into prediction equations for selection purposes in a practical, agricultural setting.

**W1075: The Analysis and Role of the Microbiome**

**Metagenomic Sequencing of Seven Hypervariable Regions with Ion GeneStudio S5 in Sea Beet**

**Piergiorgio Stevanato**, DAFNAE, Università degli Studi di Padova, Legnaro (Padova), Italy and Chiara Broccanello, University of Padova, Legnaro (Padova), Italy

Metagenomics sequencing rely on the analysis of 16S ribosomal RNA sequences. Some of these regions are highly conserved and used for taxa identification, while the more variable regions serve to identify genus or species. The choice of 16S RNA region can significantly affect the accuracy in the estimate of taxonomic diversity. In this study, we present results coming from Ion GeneStudio S5 technology using 16S Ion Metagenomics Kit to amplify seven hypervariable region, V2, V3, V4, V6, V7, V8 and V9, of bacterial 16S rRNA in sea beet (*Beta vulgaris* spp. *maritima*). We analysed leaf samples of sea beets seedlings growing in 12 sites of Adriatic coast. The on-line pipeline Ion Reporter provides simple tools to manage, analyse and track samples. We highlighted the key features of Ion GeneStudio S5 sequencing technology together with Ion Reporter software for metagenomic 16S sequencing and data analysis. These information could be fundamental in guiding the choice towards a specific technology for data sequencing and analysis.

**W1076: The Analysis and Role of the Microbiome**

**Searching for Links between Gut Microbiota Collected before Vaccination and Variabilities of Vaccine Response in Pigs**

**Marion Borey**, GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

Understanding and predicting why some animals respond better to vaccination than others is a main concern to strengthen vaccination efficiency. Our aim was to study whether the gut microbiota before vaccination presents composition patterns associated with individual variabilities of vaccine responses in pigs. Ninety-eight Large White piglets were vaccinated against the influenza A virus (IAV) at weaning at 28 days of age (D28) with a booster three weeks later. Stools were collected before the vaccination at D28, and were further processed to perform 16SRNA gene sequencing (Illumina MiSeq) and assess microbial taxonomic composition. The piglets' humoral response was evaluated by ELISA of seric IAV-specific IgGs and by hemagglutination inhibition assays (HAI) at D49, D56, D63, and D146 to identify extreme animals with either high or low responses to vaccination. Piglets with a richer microbiota had higher levels of HAI at D63 ( $p < 0.05$ ) and had a tendency towards more IAV-specific IgGs. Extreme high and low responders for IAV-specific IgGs at D63 had also a dissimilar microbiota ( $p < 0.01$ ) and displayed differentially abundant operational taxonomic units (OTUs); bacteria from the Paludibacteraceae family and *Prevotella* genera were more abundant in high responders, while bacteria from *Helicobacter* and *Escherichia-Shigella* genera were more abundant in low responders ( $FDR < 0.05$ ). Thus, our results show that the faecal microbiota before vaccination could be further investigated to identify biomarkers predictive of vaccine response levels and analyse the underlying biology.

**W1077: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**20 Years of Interagency Coordination and Cooperation on Plant Genomics**

**Sharlene Weatherwax**, Office of Biological and Environmental Research / Department of Energy, Washington, DC

The National Plant Genome Initiative was initiated in 1999. This talk will provide a historical perspective on the NPGI, highlighting accomplishments from 20 years of interagency coordination and cooperation.

**W1078: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**Frontiers in Genomics**

**Donal T. Manahan**, National Science Foundation, Alexandria, VA

Dramatic advances in genome science have occurred with support from multiple agencies. This talk will focus on plant genomics and the impact those advances are having on the growing field of functional genomics, now being applied across a wide range of organisms.

**W1079: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**TBD**

**Daniel Jacobson**, Oak Ridge National Laboratory, Oak Ridge, TN

**W1080: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**New Technologies for Predictive Plant Phenomics**

**Patrick S. Schnable**, Department of Agronomy, Iowa State University, Ames, IA

To overcome some of the myriad challenges facing sustainable crop production we are seeking to develop statistical models that will predict crop performance in diverse agronomic environments. Crop phenotypes such as yield and drought tolerance are controlled by genotype, environment (considered broadly) and their interaction (GxE). As a consequence of the next generation sequencing revolution genotyping data are now available for a wide diversity of accessions in each of the major crops. The necessary volumes of phenotypic data, however, remain limiting and our understanding of molecular basis of GxE is minimal. To address this limitation, we are collaborating with engineers to construct new sensors and robots to automatically collect large volumes of phenotypic data. New sensors and high-throughput, high-resolution, *field-based* phenotyping systems will be described. Finally, an administrative structure that fosters transdisciplinary collaborations will be described.

**W1081: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**Challenges and Opportunities in Genomics and Precision Breeding**

**Wayne A. Parrott**, University of Georgia, Athens, GA

Diversity – in the form of changes at the DNA level—helps the study of gene function and a requirement for plant improvement. For most of history, the only diversity available was what occurred naturally. The advent of mutagenesis and other technologies increased the amount of diversity that was available. However, the ability to alter DNA sequences in specified locations has fired up the human imagination. All of the sudden, we can create allelic diversity and eliminate duplicate genes or even entire gene families. With further refinements, precision breeding technology will replace backcrossing, lead to site-

specific DNA insertion, and to numerous other uses. Nevertheless, several technological developments are still necessary for this vision to materialize. Tissue culture is still needed for efficient editing, but tissue culture is still genotype dependent, and is too complicated for most research labs. The more complex the genome, the more difficult it is to identify proper target sites—thus different types of Cas enzymes are necessary for maximum target flexibility. Efficient knock-in production remains elusive. Technological challenges aside, the scientific community must also address excessive regulations and tentative public support if these technologies are to reach their full support.

**W1082: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**Accelerating Resilient Crop Development**

**Julia Bailey-Serres**, Center for Plant Cell Biology, UC Riverside, Riverside, CA

Crop yields may be insufficient to nourish mankind within 30 years. In the US and elsewhere, crop production is limited each year by extremes in water and temperature, soil degradation, pests and pathogens. The challenge of increasing yields in a changing climate is accompanied by the necessity to reduce the environmental impacts associated with pre- and post-harvest production of food. Plant scientists are accelerating crop improvement by leveraging naturally genetic variation in resilience, promoting beneficial plant-microbial interactions, and implementing transformative engineering and synthetic biology technologies (1). Crop improvement can be achieved more rapidly and confidently when there is knowledge of the underlying genetic mechanisms and understanding of associated trade-offs between resilience and productivity. The timely identification and mobilization of genetic improvements necessary to ensure future harvests will require greater collaboration, increased transparency and sharing of data, further coupling of genetics and management strategies, far greater investment in research and development, as well as broad education of producers and consumers about the crucial challenges.

(1) Bailey-Serres J, Parker JE, Ainsworth EA, Oldroyd GED, Schroeder JI. (2019) Challenges and progress in the application of genetic strategies to enhance crop yields. *Nature* 575:109-118.

**W1083: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

**Redesigning Fruit Crops for the Future: Understanding Genome to Phenome Plasticity**

**Christopher Dardick**, USDA-ARS Appalachian Fruit Research Station, Kearneysville, WV

Sustainably feeding the growing world population over the next century will require new innovations in agriculture that will dramatically increase productivity while minimizing environmental impacts. Current cropping systems have been largely developed around fixed aspects of plant genetics that cannot be readily changed through traditional breeding. For example, high yielding orchard systems are designed to accommodate the large size, long juvenility period, and woody characteristics of fruit trees but require substantial amounts of labor, land space, and chemical inputs. From an evolutionary and genomic perspective these 'tree' traits have repeatedly appeared in diverse plant lineages. The degree to which the phenomic plasticity of plant species is limited by genome content and/or structure is currently unknown. The ability to re-engineer crops with novel traits such as small architectures, shortened juvenility, lack of dormancy, and continual flowering/fruitletting habits could enable more sustainable fruit production systems along with the expansion of production periods and ranges. Accomplishing this will require an in-depth understanding of genomic-phenomic relationships and their underlying genetic pathways in order to identify appropriate combinations of gene targets that preserve desirable fruit qualities while dramatically altering development and growth habit traits. Though challenging, such efforts stand to usher in a new wave of domesticated crops that will enable higher yields and quality yet in a cheaper and more sustainable fashion.

**W1084: The National Plant Genome Initiative: Catalyzing Transdisciplinary Frontiers in Genomics for Breakthroughs, Innovation, and Community Development**

TBD

Robert Bertram, US Agency for International Development, Washington, DC

**W1085: Transposable Elements**

**Rapid, Heat-Induced Transgenerational Reactivation of a Silenced Transposable Element in Maize**

Wei Guo, Purdue University, West Lafayette, IN

Transposons make up a substantial portion of most plant genomes. Due to their mutagenic potential, most of them are silenced. Although we know a lot about the means by which transposable elements are silenced, little is known about how their silencing status is maintained. Once reactivated, what accounts for a somatically or a transgenerationally transmitted reactivation state? Here, we use a minimal *Mutator* line that includes a naturally occurring variant of the *MuDR* transposon that can heritably trigger epigenetic silencing of that transposon. *MuDR* carries two genes *mudrA* and *mudrB*. We demonstrated that *Mediator of Paramutation 1 (MOP1)*, a putative RNA-dependent RNA polymerase-encoding gene, is required for the maintenance of *mudrA* silencing. However, silenced *mudrA* is only progressively reactivated after multiple generations in a *mop1* mutant background. In contrast, *mudrB* never becomes reactivated. We find that all DNA methylation is lost at *mudrA* in *mop1* mutants in the first generation. Despite this, *mudrA* remains transcriptionally silenced in this generation. Remarkably, we find that this reactivation can be dramatically accelerated in seedlings carrying a silenced *MuDR* element in a *mop1* mutant background after a brief exposure to high temperature. In contrast to previous observations, in heat stressed plants, both *mudrA* and *mudrB* are reactivated. This active state is maintained throughout the life of the plant after the initial trigger has disappeared. Remarkably, this activity is transmitted to the next generation and is not associated with DNA methylation. This is intriguing because *mudrA* and *mudrB*, which are associated with two mutually exclusive histone marks, are both reactivated upon heat exposure, suggesting that heat stress might integrate two distinct epigenetic pathways to wake up a silenced transposon. Taken together, this project will give us an opportunity to better understand the maintenance of transposon silencing, and the relationship between epigenetic silencing and stress response.

**W1086: Transposable Elements**

**Transposable Element Dynamics in *Sitophilus oryzae***

Cristina Vieira<sup>1</sup>, Clement Goubert<sup>2</sup>, Mariana Galvao-Ferrarini<sup>3</sup>, Agnes Vallier<sup>3</sup>, Nicolas Parisot<sup>3</sup>, Ana Teresa Vasconcelos<sup>4</sup>, Rita Rebollo<sup>5</sup> and Abdelaziz Heddi<sup>5</sup>, (1)UMR 5558, LBBE, Université Lyon 1, University Lyon, Villeurbanne, France, (2)Cornell University, Ithaca, NY 14850, (3)UMR0203, Biologie Fonctionnelle, Insectes et Interactions (BF2i), Institut National des Sciences Appliquées de Lyon (INSA-Lyon), Institut National de la Recherche Agronomique (INRA), University of Lyon (Univ Lyon), France, (4)National Laboratory of Scientific Computation, Petropolis, Rio de Janeiro, Brazil, (5)UMR0203, Biologie Fonctionnelle, Insectes et Interactions (BF2i), Institut National des Sciences Appliquées de Lyon (INSA-Lyon), Institut National de la Recherche Agronomique (INRA), University of Lyon (Univ Lyon), F-69621 Villeurbanne, France

*Sitophilus* spp. cereal weevils are crop pests causing an estimated worldwide loss of hundreds of millions of dollars. These insects thrive on a nutritionally poor and unbalanced diet (cereal grains), and highly rely on intracellular symbiotic bacteria (endosymbionts) that supply them with components lacking in their diets. In order to pinpoint potential molecular targets for pest-management we have sequenced *Sitophilus oryzae*'s genome, the rice weevil. Here we report the abundance of repetitive sequences (67%), including transposable elements. While all insect TE superfamilies were observed, *S. oryzae*'s genome is enriched in DNA elements, especially MITE copies. Analysis of gut transcriptome from twelve developmental timepoints show abundant TE expression (10% of whole transcriptome), mainly from MITE copies but also LINE families. *S. oryzae*'s genome harbors therefore potential active TE copies.

We are currently inhibiting developmentally regulated TE families in order to assess the impact of TEs in *S. oryzae*.

**W1087: Transposable Elements**

**Reprogramming of *Citrus* Retrotransposons during their Early Speciation**

**Carles Borredá**, Estela Pérez-Román, Victoria Ibanez, Javier Terol and Manuel Talon, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias, IVIA, Moncada, Valencia, Spain

Speciation of the genus *Citrus* from a common ancestor has recently been established to begin approximately 8 Mya during the late Miocene, a period of major climatic alterations. In here, we report the changes in activity of *Citrus* LTR retrotransposons during the process of diversification that gave rise to the current citrus species. To reach this goal, we analyzed four pure species that diverged early during citrus speciation, three recent admixtures derived from those species and an outgroup of the *Citrus* clade. More than thirty thousand retrotransposons were grouped in 10 lineages. Estimations of LTR insertion times revealed that retrotransposon activity followed a species-specific pattern of change that could be ascribed to one of three different models. In some genomes, the expected pattern of gradual transposon accumulation was suddenly arrested during the radiation of the ancestor that gave birth to the current *Citrus* species. The individualized analyses of retrotransposon lineages showed that in each and every species studied, not all lineages follow the general pattern of the species itself. For instance, in most of the genomes activity of SIRE elements reached its highest level just before *Citrus* speciation while Retrofit activity has been steadily growing. Based on these observations we propose that *Citrus* retrotransposons might respond to those stressful conditions driving speciation as a part of the genetic response involved in adaptation. This proposal implies that the evolving conditions of each species interacts with the internal regulatory mechanisms of the genome controlling the proliferation of mobile elements.

**W1088: Transposable Elements**

**Transposable Element Contributions to Dynamic Maize Genomes and Transcriptomes**

**Sarah N. Anderson**, Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA

Transposable elements (TEs) comprise the majority of the maize genome and have the potential to contribute substantially to structural and expression variation among genotypes. Our understanding of the consequences of variable TE insertions at a genome-wide scale has been limited by computational challenges associated with assembling and analyzing repetitive sequences. We developed an approach to define shared and variable TE insertions across four maize whole genome assemblies, identifying approximately 400,000 variable transposable elements representing a combination of recent TE movement, deletions, and haplotype diversity. Importantly, the precise insertion site can be defined for thousands of variable TEs, including integrations into genomic regions devoid of SNPs. The consequences of variable TE insertions extend beyond variation in genome structure and contribute to dynamic variation in the epigenome and transcriptome. The dynamics of TE transcription in maize was assessed by quantifying per-family expression levels in >800 RNA-seq libraries representing a range of tissues across development, genotypes, hybrids, and abiotic stress conditions. We show that while a relatively small proportion of TE families are transcribed, expression is highly dynamic with most families exhibiting tissue-specific expression. TE transcript abundance is also variable across different stress treatments and in different tissues following the same stress. Furthermore, by assessing recombinant inbred line and hybrid transcriptomes, complex patterns of TE transcript abundance across genotypes with known TE variation emerged. Ongoing analysis of shared and variable TE insertions in multiple maize genomes will be critical for defining the role of TEs in creating genetic, epigenetic and gene expression variation in large crop genomes.

**W1089: Transposable Elements**

## **Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline**

**Shujun Ou**, Department of Ecology, Evolution, and Organismal Biology (EEOB), Iowa State University, Ames, IA

### **Background**

Sequencing technology and assembly algorithms have matured to the point that high-quality *de novo* assembly is possible for large, repetitive genomes. Current assemblies traverse transposable elements (TEs) and provide an opportunity for comprehensive annotation of TEs. Numerous methods exist for annotation of each class of TEs, but their relative performances have not been systematically compared. Moreover, a comprehensive pipeline is needed to produce a non-redundant library of TEs for species lacking this resource to generate whole-genome TE annotations.

### **Results**

We benchmark existing programs based on a carefully curated library of rice TEs. We evaluate the performance of methods annotating long terminal repeat (LTR) retrotransposons, terminal inverted repeat (TIR) transposons, short TIR transposons known as miniature inverted transposable elements (MITEs), and Helitrons. Performance metrics include sensitivity, specificity, accuracy, precision, FDR, and F1. Using the most robust programs, we create a comprehensive pipeline called Extensive *de-novo* TE Annotator (EDTA) that produces a filtered non-redundant TE library for annotation of structurally intact and fragmented elements. EDTA also deconvolutes nested TE insertions frequently found in highly repetitive genomic regions. Using other model species with curated TE libraries (maize and *Drosophila*), EDTA is shown to be robust across both plant and animal species.

### **Conclusions**

The benchmarking results and pipeline developed here will greatly facilitate TE annotation in eukaryotic genomes. These annotations will promote a much more in-depth understanding of the diversity and evolution of TEs at both intra- and inter-species levels. EDTA is open-source and freely available: <https://github.com/oushujun/EDTA>.

## **W1090: Transposable Elements**

### **Long-Read Transcriptome Annotation Enables High Resolution Transposable Element Bioinformatics**

**R. Keith Slotkin**, Donald Danforth Plant Science Center, St. Louis, MO

The genome feature annotations of transposable elements (TEs) typically consist of nothing more than position on the chromosome and family classification. In contrast, gene annotations contain information regarding transcriptional start sites, direction of transcription, location of introns, alternative splicing patterns and polyadenylation site(s). These additional features enable higher-resolution bioinformatics for genes compared to TEs. We have used Nanopore full-length cDNA sequencing in TE-activated genotypes to upgrade the TE annotation in the reference plant *Arabidopsis thaliana*. Our analyses have identified which individual elements are able to generate a mature mRNA, resulting in a 'gene-like' transcriptional annotation of these TEs. Our new transcript-based TE annotation avoids loss of information or bioinformatic artefacts by improving the placement of multi-mapping short-read sequences (such as Illumina RNA-seq).

## **W1091: Tripal Database Network and Initiatives**

### **Tripal, an Open-Source Toolkit for Construction of Online, Searchable, Extendable and FAIR Biological Databases.**

**Stephen P. Ficklin**<sup>1</sup>, Lacey-Anne Sanderson<sup>2</sup>, Bradford Condon<sup>3</sup>, Shawna Spoor<sup>4</sup>, Chun-Huai Cheng<sup>4</sup>, Abdullah Almsaeed<sup>5</sup>, Ming Chen<sup>5</sup>, Anthony Bretaudeau<sup>6</sup>, Helena Rasche<sup>7</sup>, Dorrie Main<sup>8</sup>, Jill L. Wegrzyn<sup>9</sup>, Kirstin Bett<sup>10</sup> and Margaret Staton<sup>11</sup>, (1)Dept of Horticulture, Washington State University, Pullman, WA, (2)Dept of Plant Sciences/University of Saskatchewan, Canada, (3)University of Tennessee, Knoxville, TN, (4)Washington State University, Pullman, WA, (5)University of Tennessee, Knoxville, TN, (6)Plant Health and Environment Division of INRA, Paris, France, (7)Department of Computer Science, University of Freiburg, Germany, (8)Department of Horticulture, Washington State University, Pullman, WA, (9)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (10)University of Saskatchewan, Saskatoon, SK, Canada, (11)University of Tennessee, Institute of Agriculture, Knoxville, TN

Tripal is an open-source, freely-available toolkit for the construction of online community-focused biological databases. Its purpose is to reduce the resources needed to construct such online repositories and to do so in a FAIR-data compliant manner. The base Tripal package provides web-based data loaders for common file formats, a bulk loader for non-traditional file formats, and a programmer's API for development of custom loaders. All imported data (typically genomic, genetic and breeding data) is housed in a GMOD Chado database schema and associated with controlled vocabularies. Any data type can be published as online searchable pages and is instantly made available via RESTful web services. Tripal site-owners can leverage a larger community of other site developers who use the Tripal API to create extensions that can be shared with others, reducing duplication of effort. There are currently 40 shared extensions developed by an international group of independent site developers that provide management, data import, visualization, searching and analytics for Tripal sites

#### **W1092: Tripal Database Network and Initiatives**

##### **Tripal MegaSearch: An Interactive and Customizable Tool for Query and Download of Big Data from Tripal Databases**

**Sook Jung**<sup>1</sup>, Chun-Huai Cheng<sup>2</sup>, Jodi L. Humann<sup>1</sup>, Jing Yu<sup>1</sup>, Taein Lee<sup>2</sup>, Heidi Hough<sup>1</sup> and Dorrie Main<sup>3</sup>, (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, WA, (3)Department of Horticulture, Washington State University, Pullman, WA

Tripal MegaSearch is a Tripal module for querying and downloading biological data stored in Chado. This module allows site users to select data types, restrict the dataset by applying various filters and then choose fields to view and download through a single interface. Set by site administrators, current data types can include gene, germplasm, marker, map, QTL, genotype, phenotype and expression data. When querying for genes, users can restrict the gene dataset using various filters such as name, chromosome position and functional annotation. They can then customize fields to download, such as name, organism, type, chromosome position, various functional annotations such as BLAST, KEGG, InterPro and GO term. FASTA file can also be downloaded as well as the data with fields. Site administrators can choose from two different data sources to serve data: Tripal MegaSearch materialized views (MViews) or Chado tables. If neither data source is desired, administrators may also create their own MViews and serve them through a flexible dynamic query form of Tripal MegaSearch. Tripal MegaSearch is currently implemented in several databases including the Genome Database for Rosaceae and TreeGenes. This presentation will showcase the interfaces for site users as well as those for site administrators.

#### **W1093: Tripal Database Network and Initiatives**

##### **Tripal EUtils and Tripal HQ – New Modules to Enable FAIR Metadata**

**Sean Buehler**<sup>1</sup>, Abdullah Almsaeed<sup>2</sup>, Bradford Condon<sup>3</sup>, Christopher Childers<sup>4</sup>, Stephen P. Ficklin<sup>5</sup>, Monica Poelchau<sup>4</sup>, Lacey-Anne Sanderson<sup>6</sup> and Margaret Staton<sup>7</sup>, (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Tennessee, Knoxville, TN, (3)University of Tennessee, Knoxville, TN, (4)USDA/Agricultural Research Service/National Agricultural Library, Beltsville, MD, (5)Dept of Horticulture, Washington State

University, Pullman, WA, (6)Dept of Plant Sciences/University of Saskatchewan, Saskatoon, SK, Canada, (7)University of Tennessee, Institute of Agriculture, Knoxville, TN

The Tripal biological database software (<http://tripal.info/>) is an established toolkit for community genome databases. Tripal supports the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles due to its integration of controlled vocabularies and ontologies with Tripal content. Here, we present two Tripal extension modules that support FAIR metadata acquisition. The first module, Tripal EUtils ([https://github.com/NAL-i5K/tripal\\_eutils](https://github.com/NAL-i5K/tripal_eutils)), provides a method to integrate genome assembly metadata from the National Center for Biotechnology Information (NCBI) into the Chado database schema. Many Tripal databases host genetic and genomic content which is also stored in primary archives, such as NCBI. However, there was no recognized method to map genome assembly metadata models between NCBI and Chado. We describe the basic metadata mapping between NCBI's data resources and Chado for genome assemblies, and demonstrate the Tripal EUtils module functionality. The second module, Tripal HeadQuarters ([https://github.com/statonlab/tripal\\_hq](https://github.com/statonlab/tripal_hq)), supports an additional method to retrieve FAIR metadata – from submitters. Tripal HeadQuarters allows users to submit metadata to a Tripal database using existing Tripal bundle configurations. Database admins can then review and approve the content before it is inserted into the database. Tripal HQ can also be configured to upload files. Combined, these two new modules enable Tripal databases to provide FAIR metadata from diverse sources for the communities they serve.

#### **W1094: Tripal Database Network and Initiatives**

##### **KnowPulse: Building a Tripal Ecosystem for Pulse Crop Research and How It Can Help You**

**Carolyn T. Caron**<sup>1</sup>, Lacey-Anne Sanderson<sup>2</sup>, Reynold Tan<sup>2</sup>, Yichao Shen<sup>1</sup>, Ruobin Liu<sup>1</sup> and Kirstin Bett<sup>1</sup>, (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)Dept of Plant Sciences/University of Saskatchewan, Saskatoon, SK, Canada

KnowPulse (doi:10.3389/fpls.2019.00965; <https://knowpulse.usask.ca>) is a web resource for pulse breeders and researchers, built using Tripal and a multitude of generic, open-source Tripal extension modules. Developed at the University of Saskatchewan for more than a decade, KnowPulse's focus is on diversity data and public tools for utilizing genomic, genotypic and phenotypic data for chickpea, common bean, field pea and lentil crops. To meet the needs of pulse breeders and researchers for user-driven loading, management and distribution of large variation datasets, we developed Tripal modules such as Natural Diversity Genotypes ([https://github.com/UofS-Pulse-Binfo/nd\\_genotypes](https://github.com/UofS-Pulse-Binfo/nd_genotypes)), Tripal QTL ([https://github.com/UofS-Pulse-Binfo/tripal\\_qtl](https://github.com/UofS-Pulse-Binfo/tripal_qtl)), Analyzed Phenotypes (<https://github.com/UofS-Pulse-Binfo/analyzedphenotypes>), VCF Filter ([https://github.com/UofS-Pulse-Binfo/vcf\\_filter](https://github.com/UofS-Pulse-Binfo/vcf_filter)) and many more. As long-term members of the Tripal community, we have benefited greatly from the support and open-source nature of shared development while prioritizing contributing back to the community by ensuring our modules are easy to use, well-documented and tested to work on a generic Tripal website. To enhance the Tripal developer experience, API modules including Chado Custom Search ([https://github.com/UofS-Pulse-Binfo/chado\\_custom\\_search](https://github.com/UofS-Pulse-Binfo/chado_custom_search)), Tripal Download API ([https://github.com/tripal/trpdownload\\_api](https://github.com/tripal/trpdownload_api)), Tripal d3.js API (<https://github.com/tripal/tripald3>), and Tripal Fancy Fields ([https://github.com/tripal/trpfancy\\_fields](https://github.com/tripal/trpfancy_fields)) were developed to facilitate further module development, customization and provide consistent user interfaces. Additionally, collaboration with other Tripal developers has led to seamless tool-embedding modules such as Tripal BLAST ([https://github.com/tripal/tripal\\_blast](https://github.com/tripal/tripal_blast)), CViTjs Embed (<https://github.com/UofS-Pulse-Binfo/cvitembed>), and Tripal JBrowse ([https://github.com/tripal/tripal\\_jbrowse](https://github.com/tripal/tripal_jbrowse)). All of these modules can be appreciated in action on KnowPulse, which we hope will inspire ways to incorporate them into your own Tripal ecosystem.

#### **W1095: Tripal Database Network and Initiatives**

##### **RNaseq Data in Tripal - Two Steps Forward, One Step Back**

**Margaret Elizabeth Staton**<sup>1</sup>, Joseph West<sup>2</sup>, Casey Richards<sup>3</sup>, Patrick Sisler<sup>3</sup>, Raymond Senu<sup>3</sup>, Charles Phillips<sup>1</sup>, Matthew Huff<sup>4</sup>, Abdullah Almsaeed<sup>1</sup>, Jill L. Wegrzyn<sup>5</sup>, Dorrie Main<sup>6</sup> and Stephen P. Ficklin<sup>7</sup>,

(1)University of Tennessee Institute of Agriculture, (2)University of Tennessee, Knoxville, TN, (3)University of Tennessee, Knoxville, TN, (4)University of Tennessee, Knoxville, TN, (5)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (6)Department of Horticulture, Washington State University, Pullman, WA, (7)Dept of Horticulture, Washington State University, Pullman, WA

The Hardwood Genomics Database ([hardwoodgenomics.org](http://hardwoodgenomics.org)) is a Tripal-based project that houses genomics data for tree species. A major expansion effort is currently underway to incorporate gene expression data from public repositories. Now fully compatible with Tripal version 3, the Tripal Expression Analysis module loads, stores and displays information from RNASeq-based gene expression experiments including descriptions of biological materials, analysis protocols, and gene expression levels. We recently updated this module to incorporate statistically significant gene expression changes. This data is stored and displayed as a compound controlled vocabulary entry of the significance testing value (p-value), relationship\_type (i.e. "up-regulated"), experimental factor ("drought"), and evidence code ("inferred from gene expression"). With this module fully developed, the HWG team has embarked on a mission to identify and upload the most useful public gene expression datasets. Our experience has revealed that standards and practices for RNASeq-based gene expression data are severely lacking, and those standards that do exist are not enforced by journals or funding agencies. For the majority of published papers, only raw data files are available which must be analyzed by a pipeline of software to regenerate normalized gene expression values and statistically significant genes. We are employing a team of undergraduate researchers to perform biocuration and are building a series of NextFlow pipelines to standardize this process. The process of incorporating RNASeq gene expression data into community databases via Tripal is still a work in progress, with new efforts needed in data standardization and automated data processing/ingest.

#### **W1096: Tripal Database Network and Initiatives**

##### **Tapping into the Tripal Galaxy: Association Mapping Studies on Forest Trees with CartograTree**

**Nic Herndon**<sup>1</sup>, Ronald Santos<sup>2</sup>, Risharde Ramnath<sup>2</sup>, Peter Richter<sup>3</sup>, Sean T Buehler<sup>4</sup>, Emily Grau<sup>5</sup>, Jill L. Wegrzyn<sup>3</sup>, Dorrie Main<sup>6</sup>, Stephen P. Ficklin<sup>7</sup> and Margaret Staton<sup>8</sup>, (1)East Carolina University, Greenville, NC, (2)University of Connecticut, Storrs, CT, (3)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (4)University of Tennessee, Storrs, CT, (5)Department of Plant Sciences, Davis, CA, (6)Department of Horticulture, Washington State University, Pullman, WA, (7)Dept of Horticulture, Washington State University, Pullman, WA, (8)University of Tennessee, Institute of Agriculture, Knoxville, TN

CartograTree is an open-source analytic web-based framework that allows the query and analysis of genotypic, phenotypic, and environmental data. It uses a map-based setting for association mapping, ecological genomics, and landscape genomics analyses, primarily on forest trees, to address the numerous threats they face worldwide. Although it was initially designed for forest trees, other plants can also be supported. CartograTree is developed as a Drupal module to enable the seamless integration with Tripal and to take advantage of some of the features provided by the latter. It uses genotypic and phenotypic data provided by three clade organism databases - TreeGenes, Hardwood Genomics, and Genome Database for Rosaceae - that use the Tripal v3 platform, which enables the data to be shared between them using Semantic Web technologies. It complements these data with environmental variables from WorldClim, and other public repositories, which are displayed on maps using GIS servers. CartograTree taps into the Tripal Galaxy API to i) transfer the data selected by the user to a Galaxy instance, ii) invoke a workflow to analyze these data, iii) notify the user that the workflow started, and iv) return the results from the Galaxy instance to CartograTree. In addition, CartograTree uses the Tripal Galaxy web forms to display to users the list of jobs submitted, along with their status.

#### **W1097: Tripal Database Network and Initiatives**

##### **The United States Swine Pathogen Database: Integrating Diagnostic Sequence Data of Emerging Pathogens of Swine**

**Blake T. Inderski**, Tavis K. Anderson and Kay S. Faaberg, Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA

**Objective:** Veterinary diagnostic laboratories annually derive partial nucleotide sequences of thousands of swine pathogen isolates. In addition, next generation sequencing has resulted in the rapid production of full-length genomes. Presently, sequence data are released solely to diagnostic clients, as data are associated with sensitive information. However, this information can provide information to: objectively design field-relevant vaccines; determine when and how pathogens are spreading across the landscape; and identify virus transmission hotspots.

**Methods:** In tandem with the USDA-ARS Big Data initiative, we have developed a centralized sequence database at the National Animal Disease Center. We have implemented the Tripal v3 toolkit's BLAST interface and JBrowse genome visualization modules on Drupal v7 using the Chado database schema. Search forms supporting multi-variable queries with customizable download options are available. Hosting is via Amazon Web Services for Federal Government with resource scaling and dedicated support.

**Results:** Sequences housed in the database contain at least four data fields: genomic information; date of collection; collection location (state level); and a unique identifier. Custom curation and annotation pipelines have been developed for multiple swine pathogens with capabilities of detecting the location of open reading frames, generating amino acid sequences, and identifying putative frame shifts.

**Conclusion:** The resource will provide researchers timely access to sequences discovered by veterinary diagnosticians, allowing for biological data mining and epidemiological studies. The result will be a better understanding concerning the emergence of novel viruses, how these novel isolates are disseminated in the US and abroad, and discovering new patterns of biological consequence.

**W1098:       Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

**The Bread Wheat Epigenomic Map Reveals Distinct Chromatin Architectural and Evolutionary Features of Functional Genetic Elements**

**Yi-Jing Zhang**, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Bread wheat is an allohexaploid species with a 16-Gb genome that has large intergenic regions, which presents a big challenge for pinpointing regulatory elements and further revealing the transcriptional regulatory mechanisms. Chromatin profiling to characterize the combinatorial patterns of chromatin signatures is a powerful means to detect functional elements and clarify regulatory activities in human studies. Here, through comprehensive analyses of the open chromatin, DNA methylome, seven major chromatin marks, and transcriptomic data generated for seedlings of allohexaploid wheat, we detected distinct chromatin architectural features surrounding various functional elements, including genes, promoters, enhancer-like elements, and transposons. Thousands of new genic regions and cis-regulatory elements are identified based on the combinatorial pattern of chromatin features. Roughly 1.5% of the genome encodes a subset of active regulatory elements, including promoters and enhancer-like elements, which are characterized by a high degree of chromatin openness and histone acetylation, an abundance of CpG islands, and low DNA methylation levels. A comparison across sub-genomes reveals that evolutionary selection on gene regulation is targeted at the sequence and chromatin feature levels. The divergent enrichment of cis-elements between enhancer-like sequences and promoters implies these functional elements are targeted by different transcription factors. Altogether, we present a systematic epigenomic map for the annotation of cis-regulatory elements in the bread wheat genome, which provides new insights into the connections between chromatin modifications and cis-regulatory activities in allohexaploid wheat.

All data could be visualized with a customized genome browser ([http://bioinfo.sibs.ac.cn/cs\\_epigenome](http://bioinfo.sibs.ac.cn/cs_epigenome)).

**W1099: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

**What Transposable Element Populations Tell Us about Triticeae Evolution**

**Thomas Wicker**, University of Zurich, Zurich, Switzerland

Transposable elements (TEs) are small genetic units that have the ability to make copies of themselves and spread through genomes. Many of them were so prosperous during evolution that they populated their “host” genomes with hundreds of thousands of copies. In Triticeae, a handful of TE families have been so successful that they contribute almost half of the genomic sequences. TEs evolve analogous to populations of individuals, having active periods (where many copies are produced) alternating with “silent” periods. The recently produced chromosome-scale assemblies of multiple Triticeae genomes allowed to trace the evolution of major TE families across different species and cultivars. We found that TE profiles can be used to identify ancient and recent hybridisation and/or chromosomal introgression events. For example, we could characterize an ancient hybridisation of two distant rye haplotypes whose signature is still recognizable in the TE profiles of modern-day chromosomes 4R and 6R. Furthermore, we were able to identify multiple recent introgressions from diverse genotypes into the modern wheat gene pool. We conclude that TE population analysis can provide insight into the evolution of species or groups of species, as well as help characterize the genetic background of modern crop germplasm.

**W1100: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

**Progress Report of International Triticeae genome Initiative (ITGC) 2019**

**Jizeng Jia**, Chinese Academy of Agricultural Sciences, Beijing, China

*Triticeae* is a large tribe of family *Poaceae* (*Gramineae*), including wheat, barley, rye and several forage and pasture grasses. It has long been of great economic importance to humanity. Among *Triticeae*, many species can cross with wheat to produce substitution and/or translocation lines for wheat improvement. The well-known 1BL/1RS translocation or 1B/1R substitution lines in which 1R (1RS) was transferred from rye to wheat have been widely employed in wheat breeding and made significant contributions to wheat yield all over the world. Therefore, it is extremely important and urgent to clear *Triticeae* genomes for assessing evolution and taxonomy of *Triticeae* genome, mining valuable genes, and breeding targeted grasses.

However, the *Triticeae* genome is very big and complex, and there are 25 generally recognized genera in *Triticeae*. It is difficult to complete sequencing all the *Triticeae* genomes by one or two institutions. Therefore, we initiate the “International *Triticeae* Genome Initiative (ITGC)”. More than 10 institutions join the club and each institution is responsible for sequencing at least one genome of *Triticeae* species. Eleven genera including *Aegilops*, *Agropyron*, *Elymus*, *Eremopyron*, *Haynaldia*, *Hordeum*, *Leymus*, *Psathyrostachys*, *Roegneria*, *Secale*, and *Triticum* are selected for doing de novo sequencing and assembling by using PacBio/Nanopore approach. All the selected genera have been crossed with wheat successfully, and addition lines, substitution lines and translocation lines generated from the crosses between wheat and above 11 genera were bred. Some lines carry agronomic important genes and have been applied in wheat breeding. Till now, twenty genomes are sequenced or on going. High quality reference genomes have been generated with contig N50 bigger than 1Mb. More details will be introduced on the works.

**W1101: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

**Exploring Barley Genetic Diversity with Chromosome-Scale Assemblies for 20 Diverse Genotypes**

**Murukarthick Jayakodi**, IPK Gatersleben, Gatersleben, Germany and The International Barley Pangenome Consortium

Genetic diversity is key to crop improvements under threats of climate change and population growth. Due to the prevalence of genomic structural variation (SV), a single reference genome is inadequate to capture the full landscape of diversity within a crop species. Multiple reference-quality genomes, a so-called pan-genome capture the full spectrum of sequence variation, potentially facilitating the discovery of casual variants for agronomical traits. Barley (*Hordeum vulgare*. L) is an important cereal crop with a long history of cultivation. High levels of natural diversity might have facilitated barley's adaptation to a wide range of agro-climatic conditions. To build a first-generation pan-genome for barley, we constructed high-quality assemblies for 20 geographically diverse barley accessions. Super-scaffolds (average N50: 27 Mb) were assembled from short-read data and arranged into chromosomal pseudomolecule by chromosome conformation capture sequencing (Hi-C) using the TRITEX pipeline. The accuracy and completeness of the assemblies were validated by collinearity with the reference genome cv. Morex, inspection of Hi-C contact matrices and alignment full-length transcripts. We catalogued genomic presence/absence variants (PAVs) and large inversions and explored the use of SV in trait mapping by conducting association scans for heritable morphological traits using genome-wide association study (GWAS) with shallow resequencing data from a panel of 200 domesticated accessions from the German *ex situ* genebank. The barley pan-genome provides access to previously hidden genetic variation and offers a new basis for studying crop evolution in barley.

**W1102: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

**Detecting Gene Duplications and Copy Number Variants in Barley from Exome Sequencing Data**

**Alessandro Tondelli**<sup>1</sup>, Davide Guerra<sup>1</sup>, Agostino Fricano<sup>1</sup>, Caterina Morcia<sup>1</sup>, Gianluca Bretani<sup>2</sup>, Laura Rossini<sup>2</sup> and Luigi Cattivelli<sup>1</sup>, (1)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (2)Università degli Studi di Milano – DiSAA, Milano, Italy

Genomic structural variants, such as gene and segmental duplications, copy number variants (CNVs), and presence absence variants (PAVs) are widespread in several animal and plant genomes and contribute to shape the genetic diversity of the species. In this work we have exploited the large dataset generated in the frame of the FP7-Wheatbarley project (Wheat and barley legacy for breeding improvement) to study the extent of structural variation in barley, both at the genome-wide level and at single locus level. Using exome sequencing and read count data, we detected >16,000 deletions and duplications that affect gene content in a panel of 397 diverse barley accessions. Specific loci, for which an association between CNVs and a phenotype is conceivable, were analysed in detail. The presence of CNVs at the locus coding for the C-repeat Binding Factors (CBFs), the major determinants of frost tolerance in the *Triticeae*, was screened across the whole panel. The observed CNVs were validated by digital PCR and associated to differences in spring barley resistance to low temperatures. Heterozygous mapping was also adopted to detect gene duplications and paralogs, and an interesting example concerning the *black lemma and pericarp* (*blp1*) locus will be shown.

**W1103: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

**Multiple Reference Genome Assemblies Identify Two Independent Sources of Stem-Solidness in Wheat**

Curtis J. Pozniak<sup>1</sup>, Kirby T. Nilsen<sup>2</sup>, Sean Walkowiak<sup>3</sup>, Teagen D. Quilichini<sup>4</sup>, Peng Gao<sup>5</sup>, Valentyna Klymiuk<sup>6</sup>, Martha Bonilla Betancourt<sup>6</sup>, Daoquan Xiang<sup>7</sup>, Penny J. Tricker<sup>8</sup>, Peter Langridge<sup>9</sup>, Ken Chalmers<sup>9</sup>, Richard Cuthbert<sup>10</sup>, Ron E. Knox<sup>10</sup>, Raju Datla<sup>11</sup> and Pierre J. Hucl<sup>1</sup>, (1)Crop Development

Centre, University of Saskatchewan, Saskatoon, SK, Canada, (2)Brandon Research and Development Centre, Agriculture and Agri-Food Canada, Brandon, MB, Canada, (3)Canadian Grain Commission, Winnipeg, MB, Canada, (4)National Research Council, Canada, (5)Global Institute for Food Security, Canada, (6)Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (7)National Research Council Canada, Saskatoon, SK, Canada, (8)School of Agriculture, Food and Wine, University of Adelaide, Adelaide, SA, Australia, (9)The University of Adelaide, Adelaide, Australia, (10)Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current, SK, Canada, (11)Global Institute for Food Security, Saskatoon, SK, Canada

The most effective way to reduce losses caused by the wheat stem sawfly (*Cephus cinctus* Norton (Hymenoptera: Cephidae) is to grow durum (*Triticum turgidum* L. var *durum*) and hexaploid wheat (*Triticum aestivum* L.) cultivars that express solid stems. As part of the 10+ Wheat Genome Project, we have generated and made available whole genome assemblies for 15 wheat genotypes ([www.10wheatgenomes.com](http://www.10wheatgenomes.com)), including CDC Landmark a solid stemmed cultivar. Using these and available genome assemblies of tetraploid wheat, we cloned the *TdDof-3B* locus on chromosome 3BL as the causal genetic factor for stem-solidness in wheat. We show that copy number variation of *TdDof-3B* drives increased gene expression and the solid-stemmed phenotype, and that transgenic complementation in two independent hollow stemmed cultivars induced stem solidness. We identified a homoeologous region responsible for stem-solidness on chromosome 3D that we hypothesized may have been derived from tall wheatgrass (*Thinopyrum ponticum*). Using the 10+ Wheat genome assemblies, we identified a 60 Mb terminal introgression from *T. ponticum* in the solid stemmed Australian cultivar Lancer within the same region on 3D. The introgression was confirmed by whole genome sequencing and comparative genomics with *T. ponticum*, cv. Orbit. Gene expression studies in eight independent genetic backgrounds annotated multiple *T. ponticum* *Dof* copies localized to the introgression. Molecular markers developed from this work are currently being assessed for selection to minimize genetic-background dependent suppression of stem-solidness to further enhance pith expression in Canadian wheat breeding programs.

**W1104: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**

**Optimising Source:Sink Balance : Association Mapping of Photosynthetic/Photoprotective Pigment Composition in Spring Wheat**

**Ryan Joynson**, Earlham Institute, Norwich, United Kingdom

One of the major challenges we face is increasing wheat yield potential (YP). A significant bottleneck to increasing YP is achieving increased biomass through optimization of radiation use efficiency (RUE) along the crop cycle. The composition and ratios of both photosynthetic and photoprotective pigments in leaves can vary greatly in wheat and are thought to have a significant effect on photosynthetic efficiency; however, the effect of this diversity on YP is still unclear. To understand the genetic basis of leaf pigment composition and its links to RUE, we conducted a GWAS using a panel of elite spring wheat genotypes including landrace and synthetically derived lines assembled at CIMMYT. The panel was evaluated for yield, biomass and RUE related traits alongside pigmentation analysis using hyperspectral reflectance. Leaf reflectance curves were used to quantify more than 30 traits relating to pigmentation, carbohydrate content and water content. The panel was genotyped using a custom hybridisation enrichment capture enabling *de novo* SNP discovery. This yielded an average of 700K SNPs per accession and revealed a high level of D genome variation in landrace and synthetically derived panel members. Marker-trait association identified multiple genomic regions related to leaf pigment levels. These included chlorophyll A and B content and breakdown along with the content photoprotective carotenoids. Allelic variation in SNPs associated with pigmentation variation in a number of traits could be attributed to the presence of exotic material in the pedigree history of the panel members. This suggests the value of integrating exotic material into global prebreeding programs as a strategy to optimise RUE in elite varieties and to alleviate genetic bottlenecks, especially in regard to the D sub genome.

**W1105: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**

## Regulation of Spikelet Number per Spike in Wheat Development

**Jorge Dubcovsky**, Chengxia Li, Huiqiong Lin, Maria Alejandra Alvarez, Juan M. Debernardi, Saarah N. Kuzay, Francine Paraiso, Kun Li, Priscilla Denise Glenn and Qiujiu Liu, University of California, Davis, CA

Wheat is an important crop for global food security and continuous increases in grain yield are required to feed a growing human population. Total grain yield is determined by the number of spikes per unit of land area, the spikelet number per spike (SNS), the number of grains per spikelet and the average grain weight. In this talk, we focus on the genes regulating SNS and their potential to increase total grain yield. The final SNS is determined at the time when the inflorescence meristem (IM) stops generating lateral spikelet meristems and transitions into a terminal spikelet. We discovered that combined mutations in the MADS-box meristem identity paralogs *VRN1* and *FUL2* are sufficient to eliminate the formation of the terminal spikelet, and result in indeterminate spike growth. Combined mutations in all three related paralogs (*VRN1*, *FUL2* and *FUL3*) result in the complete conversion of the lateral spikelets into vegetative tillers subtended by leaves in the triple mutant *vrn1ful2ful3*, indicating that these genes are essential for spikelet development and repression of the subtending bract. Individual mutations in *vrn1* and *ful2* result in higher SNS, with stronger effects in the *vrn1* mutants. However, the smaller effects of the *ful2* mutant on heading time makes this mutant more attractive for breeding applications. Overexpression of genes that up-regulate these MADS-box genes, such as *FT1* and *FT2* greatly reduce SNS, whereas mutations in these genes result in higher SNS. However, *ft1* and *ft2* mutations also result in reduced fertility or very late flowering (*ft1*), which limits their use in wheat improvement. Loss-of-function mutations in *PHOTOPERIOD 1 (PPD1)*, which upregulates *FT1* and *FT2* under long days, also results in increased SNS. By contrast, loss-of-function mutations in *ELF3*, a repressor of *PPD1*, result in reduced SNS. A hypermorphic *Elf-A<sup>m3</sup>* allele identified in *T. monococcum* increased SNS when introgressed into durum wheat and showed limited effect on heading time and a positive effect on grain yield in some field trials. Using a map-based cloning approach, we identified *WAPO1* as an additional gene affecting SNS on chromosome arm 7AL. Loss-of-function mutations in *WAPO-A1* or alleles with reduced transcript levels result in reduced number of spikelets suggesting that this gene acts as a repressor of the transition of the IM into a terminal spikelet. Mutants of the rice ortholog *APO1* also reduce the number of spikelets in the panicle through its negative regulation of rice MDAS-box genes orthologous to wheat *VRN1 FUL2 FUL3*. The *Wapo-A1b* allele (C47F polymorphism) was associated with increases in SNS and with positive effect on total grain yield. These developmental and positional cloning studies are revealing an interconnected network of genes that affect SNS by regulating the timing of the transition of the IM into a terminal spikelet. These genes provide valuable entry points to manipulate this important grain yield component in wheat.

## W1106: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification

### A Novel Class of Homoeologous Genes Targeted by a Fungal Necrotrophic Effector Triggers Disease Susceptibility in Wheat

**Zengcui Zhang**<sup>1</sup>, Katherine L.D. Running<sup>2</sup>, Sudeshi Seneviratne<sup>2</sup>, Amanda R. Peters<sup>2</sup>, Agnes Szabo-Hever<sup>1</sup>, Gongjun Shi<sup>2</sup>, Ming-Cheng Luo<sup>3</sup>, Robert S. Brueggeman<sup>2</sup>, Steven S. Xu<sup>1</sup>, Timothy L. Friesen<sup>1</sup> and Justin D. Faris<sup>1</sup>, (1)USDA-ARS, Fargo, ND, (2)North Dakota State University, Fargo, ND, (3)University of California, Davis, CA

Septoria nodorum blotch (SNB), a fungal disease caused by the necrotrophic pathogen *Parastagonospora nodorum*, is a threat to wheat production worldwide. Multiple inverse gene-for-gene interactions involving the recognition of necrotrophic effectors (NEs) by wheat sensitivity genes have been shown to play major roles in causing SNB. The previously cloned NE sensitivity genes *Tsn1* and *Snn1* are members of the NLR and wall-associated kinase classes of genes, respectively. Here, we cloned and validated by mutagenesis a pair of homoeologous genes named *Snn3-D1* and *Snn3-B1*, which both mediate recognition of the *P. nodorum* NE SnTox3 to confer susceptibility to SNB. These genes belong to a class different from *Tsn1* and *Snn1*. *Snn3-D1* was not found in hexaploid wheat, and it was recently acquired by *Aegilops tauschii* through a 200 kb insertion that likely occurred along the west bank of the Caspian Sea. *Snn3-B1* is prevalent among hexaploid wheat accessions, but somewhat rare

among tetraploids. Like *Tsn1* and *Snn1*, *Snn3* transcriptional expression is regulated by light. However, *Snn3* expression patterns are the opposite of *Tsn1* and *Snn1* in that expression increases under darkness and decreases under light. Yeast two-hybrid study indicated that the Snn3-D1 protein does not interact with SnTox3 directly. The cloning of the *Snn3* genes and the characterization of their novel features advances our understanding of the wheat-*P. nodorum* pathosystem and plant-pathogen interactions in general. This work also provides knowledge for strategic development of SNB-resistant wheat varieties through marker-assisted elimination of susceptibility genes or gene editing.

#### **W1107: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**

##### **Multiple Pathogen Recognition at the *Mla* Locus in Barley**

**Helen J. Brabham**<sup>1</sup>, Jan Bettgenhaeuser<sup>1</sup>, Inmaculada Hernández-Pinzón<sup>1</sup>, Jennifer Lorang<sup>2</sup>, Tom Wolpert<sup>2</sup>, Jodie Taylor<sup>3</sup>, Matthew Smoker<sup>3</sup>, John Ferguson<sup>4</sup>, Matthew Gardiner<sup>3</sup>, Andrew Dawson<sup>3</sup>, Kazuhiro Sato<sup>5</sup>, Patrick M. Hayes<sup>2</sup>, Hiromasa Saitoh<sup>6</sup>, Motoki Shimizu<sup>7</sup>, Peter Emmrich<sup>8</sup>, Amelia Hubbard<sup>9</sup>, Rosemary Bayles<sup>10</sup>, Brande B.H. Wulff<sup>11</sup>, Eric Ward<sup>12</sup>, Ryohei Terauchi<sup>7</sup> and Matthew J. Moscou<sup>1</sup>, (1)The Sainsbury Laboratory, Norwich, United Kingdom, (2)Oregon State University, Corvallis, OR, (3)The Sainsbury Laboratory, United Kingdom, (4)University of Nottingham, United Kingdom, (5)IPSR, Okayama University, Okayama, Japan, (6)IBRC, Japan, (7)Iwate Biotechnology Research Center, Kitakami, Japan, (8)BecA-ILRI Hub, Kenya, (9)NIAB, Cambridge, United Kingdom, (10)NIAB, United Kingdom, (11)John Innes Centre, Norwich, United Kingdom, (12)AgBiome

A primary threat to stable food production is epidemics caused by plant pathogens. However, it remains unclear how plant breeding approaches may increase the vulnerability of agricultural systems. The majority of NLR-encoding resistance genes recognise single pathogen species; few NLRs have the capacity to recognise multiple pathogens. The *Mla* locus has over 30 described alleles conferring isolate-specific resistance to *Blumeria graminis* f. sp. *hordei* (powdery mildew), contains three NLR encoding gene families (*RGH1*, *RGH2*, and *RGH3*), and is associated with resistance to multiple pathogens including *Puccinia striiformis* f. sp. *tritici* (wheat stripe rust; *Rps7*) and *Magnaporthe oryzae* (rice blast; *Rmo1*). In addition, sensitivity to the *Cochliobolus victoriae* toxin victorin (*Lov1*) is in coupling with *Mla3*—in *Arabidopsis*, sensitivity to victorin is mediated by an NLR-dependent plant immune response. We found that only three of eleven *Mla* haplotypes had functional resistance to *P. striiformis* f. sp. *tritici*, including *Mla7*, *Mla8*, and *Mla15*. Using *Agrobacterium*-mediated transformation, we have established that *Mla8* confers resistance to *P. striiformis* f. sp. *tritici*, but not *P. striiformis* f. sp. *hordei*. In the *Mla3* haplotype, we performed a high-resolution recombination screen and confirmed the genetic coupling of *Mla3*, *Rmo1*, and *Lov1*. Using sequence capture and RNAseq, we discovered copy number variation and high expression levels for *Mla3* (*RGH1*; three copies), with one expressed copy (*Mla3D6*) containing a 6 bp deletion in the LRR region. Evaluation of stable transgenic barley found that *Mla3* alone conditions *Blumeria graminis* f. sp. *hordei* and *M. oryzae* resistance. This work, coupled with the recent discovery of direct interaction of MLA and AVR<sub>a</sub> effectors (Saur *et al.*, 2019), suggests that MLA has the capacity to recognize molecular structures conserved among plant pathogen effectors. The role of *Mla* in conferring multiple pathogen recognition implies that selection for resistance to a host pathogen can directly influence resistance to non-adapted pathogens. This observation impacts our understanding of the evolution of plant immunity and how breeding resistance to one pathogen may have unintended consequences.

#### **W1108: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**

##### **MutChromSeq Identifies a Plant Immune Receptor with a Non-Canonical Domain Architecture that Confers Race-Specific Resistance to Wheat Powdery Mildew**

**Javier Sanchez-Martin**<sup>1</sup>, Victoria E.L. Widrig<sup>2</sup>, Gerhard Herren<sup>3</sup>, Helen Zbinden<sup>2</sup>, Jonatan Isaksson<sup>2</sup>, Coraline R. Praz<sup>4</sup>, Markus C. Kolodziej<sup>3</sup>, Thomas Wicker<sup>3</sup>, Julien Gronnier<sup>2</sup>, Burkhard Steuernagel<sup>5</sup>, Hana Šimková<sup>6</sup>, Jaroslav Doležel<sup>6</sup> and Beat Keller<sup>3</sup>, (1)University of Zurich, Zollikerstrasse 107, Switzerland, (2)University of Zurich, Switzerland, (3)University of Zurich, Zurich, Switzerland, (4)University of Zurich, Zürich, Switzerland, (5)The Sainsbury Laboratory and John Innes Centre,

Norwich, United Kingdom, (6)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic

The molecular identification of resistance genes is essential to understand the plant immune system and to implement effective breeding strategies. Here we used MutChromSeq<sup>a</sup>, where sequences of flow-sorted chromosomes of EMS-mutants are compared to the wild-type, to clone *Pm4*, a race-specific resistance gene of wheat to powdery mildew. We found that the *Pm4* gene undergoes alternative splicing generating two isoforms, which share an N-terminal domain but differ in the C-terminal domains. While the Pm4V1 isoform has a single C2 domain, Pm4V2 contains a different C2 domain. Nearly identical expression levels during pathogen attack suggest an equal contribution to resistance of the two isoforms. VIGS, EMS mutant analysis and transgenic validation demonstrate that both isoforms are essential for *Pm4*-mediated resistance. Intriguingly, split-luciferase complementation and co-immunoprecipitation assays showed that the two isoforms interact with itself and each other, suggesting a possible oligomerization/dimerization-based mode of action of Pm4. Our work reveals a race-specific resistance controlled by a novel class of proteins, with a domain architecture unique among all plant disease resistance genes cloned so far. This finding expands our knowledge of both receptor architecture and underlying molecular modes of action in host resistance.

<sup>a</sup>Sánchez-Martín, J. et al. Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* **17**, 221 (2016)

**W1109: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**  
**Characterization of the Barley Spike Transcriptome during Development using Laser Micro Dissected Meristems**

**Ravi Koppolu**, IPK, OT Gatersleben, Stadt Seeland, Germany

The gross morphology of an organism can be traced to its early developmental events, particularly to the changes in genes controlling development. In plants, specification of various organ primordia, such as roots, leaves, and flowers, is majorly driven by the transcriptional regulation at the site of their specification. Hence, understanding the precise control of organ specification, necessitates the need to dissect the transcriptional regulation at the site of organ initiation. The barley inflorescence called spike has a unique structure called triple spikelet (TS) [(one central (CS) and two lateral spikelets (LS)] along the inflorescence axis. The CSs are always fertile. The fertility of LSs at the TS distinguishes barley spikes into two- (sterile LSs) and six-rowed (LSs fertile). To understand the transcriptional landscape specifying the opposing fates of LSs and CSs, we precisely isolated immature LS and CS organs in the two-rowed cv. Bowman by applying laser-capture microdissection across seven spike developmental stages and subjected these samples for RNA-seq analysis. Besides, we also analyzed apical inflorescence meristem, spike pro-vascular tissue, apical root, and basal leaf meristems. Our analysis of differentially regulated genes between CS and LS tissues revealed the involvement of known and unknown regulators of LS fate and development. By using mutational and phenotypic analyses of three of the novel genes, we validated their differential transcriptional regulation between CS and LS. In summary, we have developed a high-resolution tissue-specific transcriptome atlas of meristems from developing barley spikes, illuminating the precise regulation of spike development.

**W1110: UCSC Genome Browser - tools and visualizations for all organisms**

**UCSC Genome Browser History and Overview**

**Robert Kuhn**, UCSC Genome Browser, Santa Cruz, CA

For 20 years the UCSC Genome Browser has been providing a visual display for genomic data from a large number of sequenced animals, now numbering 180 assemblies of more than 120 organisms. This workshop will demonstrate the Browser and a selected set of the most useful accompanying tools. Users may load their own data into the Browser to interpret alongside UCSC-hosted data. Importantly for the PAG community, the Assembly Hub mechanism will be reviewed as it allows anyone with sequence to

display their organism and use the full feature set of the UCSC platform, even if UCSC does not host that organism.

**W1111: UCSC Genome Browser - tools and visualizations for all organisms**

**Assembly Hubs and New API Data Access**

**Brian Lee**, Univ. Calif. Santa Cruz, Santa Cruz, CA

Researchers generating plant and animal genomes can visualize their new assemblies on the UCSC Genome Browser using Assembly Hubs. Assembly Hubs begin by converting a FASTA file, and desired related annotation, to binary indexed genomic data ("big" files) that are hosted remotely at institutional or university servers. These genome-wide data sets are then sent from the external host in an on-demand transfer to the UCSC Browser website as users navigate to different genomic coordinates within the genome.

Assembly Hubs enable visualizing a number of file types, but also programmatic access to the underlying data with UCSC's new Application Programming Interface (API) with JavaScript Object Notation (JSON) output. The new API provides direct access to different data including annotations and sequence data for both native Genome Browser assemblies as well as novel organisms that can be visualized within user-generated assembly hubs and accessed by the API. Some functions possible with the API include listing all available UCSC Genome Browser assemblies, listing chromosomes contained in a UCSC Genome Browser assembly or novel user-generated assembly hub, listing all data tracks in a specific assembly, extracting any track data from native Genome Browser assemblies or novel user-generated assembly hubs. A complete list of what data can be accessed, as well as examples of how to access the data, is available on the API help page: <http://genome.ucsc.edu/goldenPath/help/api.html>

**W1112: US National Animal Genome Research Program (NRSP8)**

**High Throughput Phenotypic Screening in Livestock Species**

Tim Regan<sup>1</sup>, Chen Hsui Lee<sup>1</sup>, Bruce Whitelaw<sup>2</sup>, Ross D. Houston<sup>2</sup>, Tom Burdon<sup>1</sup>, Paul Digard<sup>3</sup>, Sam Wilson<sup>4</sup>, Kenneth Baillie<sup>1</sup> and **Finn Grey**<sup>3</sup>, (1)The Roslin Institute, United Kingdom, (2)The Roslin Institute, Edinburgh, United Kingdom, (3)The Roslin Institute, Midlothian, United Kingdom, (4)University of Glasgow, United Kingdom

Genome editing has the potential to revolutionise livestock production through generation of pathogen resistant animals. To fully leverage this potential, high confidence targets for editing need to be identified. High throughput phenotypic screens are a powerful approach to identify host genes that are essential for pathogen infection. However, until recently the tools required to perform such screens did not exist for livestock species. As part of an integrated pipeline, combining high throughput systematic genotypic and phenotypic screens for the identification of host genetic elements underpinning resistance against infection, the Roslin Institute has developed genome-wide CRISPR Cas9 libraries for multiple livestock species including, pig, chicken, cow and salmon. In addition, we have developed arrayed interferon stimulated gene expression libraries to dissect the interferon response during viral infection in pigs and chickens. We are using these screens to identify species-specific host restriction factors to influenza virus. In addition, through multiple collaborations we are using the screens to identify host factors important for replication of the most important livestock viruses, including ASFV, FMDV, IBDV, IBV, MDV and NDV among others. The technologies involved and the future directions of systems approaches will be discussed.

**W1113: US National Animal Genome Research Program (NRSP8)**

**The 200 Mammals Project: Evolutionary Conservation at Single Base Resolution**

**Elinor K. Karlsson**, UMass Med School & Broad Institute, Cambridge, MA and The 200 Mammals Project Consortium

Evolutionary constraint is among the most powerful markers of genome function, and can distinguish functionally important variation. Existing resources identify genomic regions under constraint, but lack the precision needed to prioritize individual variants for functional follow up. To address this, the 200 Mammals Project aims to dramatically increase the number of mammalian species sequenced and aligned, with the goal of identifying individual bases conserved across eutherian mammals.

The first phase of the 200 Mammals Project is now complete. We sequenced and assembled new genomes for 130 species from across the eutherian (placental mammalian) tree, including 43 previously unrepresented families. For most genomes, we made shorter contiguity assemblies using a method that did not require intact cells and uses less than two micrograms of medium-quality DNA, enabling us to include species that are difficult to access. We upgraded 10 genomes to higher contiguity using proximity-ligation techniques, focusing on species from orders with no high-contiguity genome assemblies available.

We combined our new genomes with other publicly available genomes and aligned using Cactus, which does not require a reference genome, enabling us to detect both pan-mammalian and lineage-specific elements, including those unique to non-human lineages. Our alignment of 240 mammalian genomes spans approximately ~110 MY of eutherian mammal evolution. From this alignment, we have generated a UCSC human genome track of evolutionary constraint that achieves near single base resolution. We demonstrate that the 200 Mammals constraint scores represent a critical new resource for annotating functional genomic elements, particularly when paired with other genomic annotations.

#### **W1114: US National Animal Genome Research Program (NRSP8)**

##### **Leveraging Genotype-by-Environment Interactions across Discrete Climate Regions to Select More Sustainable Beef Cattle**

**Harly J. Durbin<sup>1</sup>**, Troy N. Rowan<sup>1</sup>, Jared E. Decker<sup>1</sup> and Stephen P. Miller<sup>2</sup>, (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Angus Genetics Inc., Saint Joseph, MO

Compared to other livestock systems, beef cattle are produced in a wide range of climates with minimal management interventions. While environmental heterogeneity is a strength of the beef cattle industry, it also increases the potential for genotype-by-environment interactions. This may present issues for sustainable beef production, especially for breeders in stressful environments purchasing genetics from outside their region.

Random regression models using a single environmental variable may not capture all stressors unique to a region (i.e., local pathogens and forages) and their interactions. Therefore, we generated environmental region-specific genetic predictions. Using k-means clustering on 30-year-normals for temperature, rainfall, and elevation, we assigned growth trait records from cattle registered in the American Angus Association to one of 7 discrete climate regions based on herd zip code.

Data was collected by American Angus breeders from 1973-2018 and included 7,080,318 weaning weight and 3,468,487 post-weaning gain measurements. We then calculated breeding values from univariate & multivariate (multiple regions) restricted maximum likelihood models. We find that while genetic trends are similar between regions, there is significant re-ranking of breeding values. Genetic correlations between regions ranged from 0.75 to 0.92 for weaning weight direct, 0.64 to 0.95 for maternal milk, and 0.63 to 0.91 for post-weaning gain. In the future, these approaches could be incorporated into genetic evaluations to help beef cattle producers identify animals better suited to their environment.

#### **W1115: US National Animal Genome Research Program (NRSP8)**

##### **QTL and Joint-Association Mapping Reveal Loci Associated with Acute Low Salinity Tolerance in the Eastern Oyster (*Crassostrea virginica*)**

**Alexandra J McCarty**, University of Maryland Center for Environmental Science, Cambridge, MD

Oyster aquaculture is increasing globally, but varying environmental parameters pose challenges to farmers. In the Maryland-portion of the Chesapeake Bay, and in other estuarine and coastal systems, low salinity regions and large freshwater inflow events leave oysters exposed to salinities < 3 ppt for periods of time, which decreases oyster productivity and results in subsequent mass mortalities. While low salinity regions may hinder growth, they are enticing to farmers because they provide a refuge from common oyster diseases. A breeding program for extreme low salinity (< 3) waters would allow aquaculture to expand further into these reduced-disease areas, but the genetic framework underlying this trait is currently unknown. To determine the heritability of acute low salinity tolerance, we exposed fifty half-sibling families to an acute low salinity exposure (< 3) and determined that this trait, mortality at low salinity, is indeed heritable ( $h^2 = 0.4$ ). To determine genomic regions underlying this tolerance, we exposed four recombinant oyster families and individuals from a wild population to a similar acute low salinity exposure and generated genome-wide SNP data via ddRADseq techniques, genotyping all individuals (both alive and dead). We created QTL maps for each family and the wild population, as well as a joint-association map for both day-to-death and survival (live versus dead). Results indicate an important region on chromosome 1 related to survival at acute low salinity. Moderate heritability estimates and the identification of a highly significant QTL provide good support for the potential success of a low salinity breeding program.

**W1116: US National Animal Genome Research Program (NRSP8)**

**The Frequency of Loss of Function Alleles in the Equine Population**

**Sian Ann Durward-Akhurst**, University of Minnesota, St. Paul, MN

Identification of disease-causing alleles is a fundamental goal of medical genetics and can facilitate disease diagnosis and improve disease understanding. Studies in humans have increasingly demonstrated the value of databases of genetic variation derived from genome sequencing (WGS) for disease-causing allele(s) identification. A surprising finding from these studies is the high number of alleles computationally predicted to lead to loss of function (LOF) of the affected gene in healthy adults. LOF variants with a high frequency in healthy individuals are unlikely to be disease-causing. Here, we investigate LOF alleles present in the equine population.

We mapped WGS from 535 horses to the EquCab3 reference genome. Single nucleotide polymorphisms (SNPs) and short insertions/deletions were identified using GATK-HaplotypeCaller and SAMtools, and annotated using ANNOVAR and SnpEff. The intersect was used to identify LOF variants i.e., nonsense, frameshift, and splice site disrupting variants, or deletions removing the first exon or >50% of protein coding sequence. Average depth of coverage was 11.5x (range 1.4 - 46.7x). 29,882,273 variants were identified, with 8,683 predicted to be detrimental, including 5,673 LOF variants affecting 3,810 genes enriched for olfactory reception and immune related pathways. On average, each horse carried 829 (range 211 - 1,182).

Overall, we demonstrate that similar to humans, LOF alleles are present in the horse population. We will further validate the LOF alleles using hand annotation and produce a list of LOF alleles that can be excluded from candidate disease-causing allele discovery approaches due to their high frequency in the general population.

**W1117: US National Animal Genome Research Program (NRSP8)**

**Chasing Colors: Identifying the Genetic Variants Responsible to Coat Color Variation in Sheep**

**Christian J Posbergh**<sup>1</sup>, Elizabeth A. Staiger<sup>1,2</sup> and Heather J Huson<sup>1</sup>, (1)Cornell University, Ithaca, NY, (2)Auburn University, Auburn, AL

White wool is the dominant product in the commercial wool market and commands a higher price than non-white wool. However, non-white wool can bring significantly higher prices than white wool in the hand spinning wool market, particularly in the Northeastern United States. Several studies have

examined the phenotypic variation and genetic inheritance patterns at specific loci including the Agouti, Extension, and Brown loci in sheep. However, there has been little work to determine the molecular cause of these different phenotypes. The objective of this study was to identify variants associated with various coat color variation observed within and across sheep breeds, such as dilution, red head and legs, and white spotting. Using Illumina 150 base pair paired-end reads we generated whole genome sequences (approximately 20x coverage) from 18 sheep across the Romeldale, Romney, Jacob, and California Red breeds that were selected to represent a variety of coat color patterns. Reads were aligned to the Oar\_v4.0 genome assembly using the Burrows-Wheeler aligner, and variants called following Genome Analysis Toolkit's "Best Practices" workflow. To date, only genes known to be associated with color variants in other species have been investigated. We have identified a variant responsible for the lilac dilution in Jacob sheep within *MLPH*, and a variant within *TYRP1* which is associated with the red hair color observed within the California Red and Tunis sheep. Work is still ongoing to identify variants around the *ASIP* region due to the wide variety in phenotypes attributed to the *ASIP* locus.

**W1118: US National Animal Genome Research Program (NRSP8)**

**Anthrax Toxin Receptor 1 Knockout Pigs are protected from Senecavirus A Infection**

**Paula R. Chen**, University of Missouri, Columbia, MO

Senecavirus A (SVA) has been the cause of numerous cases of vesicular disease in swine across the world in recent years. Studies investigating the oncolytic properties of SVA in humans revealed anthrax toxin receptor 1 (ANTXR1) as its receptor. The objective of the current study was to determine if ANTXR1 functioned as the receptor for SVA in pigs by employing the CRISPR/Cas9 system to edit exon 1 and create a premature stop codon. Two founder *ANTXR1* knockout pigs and two age-matched wild type pigs were challenged with SVA. Serum, fecal swabs, and nasal swabs were collected throughout the duration of the study. Presence of viral nucleic acid was determined by PCR, and SVA antibody responses were assessed. *ANTXR1* knockout pigs exhibited distinct anatomical features, including frontal bossing and wide, short statures, which is characteristic of GAPO syndrome in humans. The knockout pigs did not develop vesicular lesions while the wild type pigs had coronary band lesions after SVA infection. Moreover, SVA nucleic acid was not detected in serum from either *ANTXR1* knockout pig, but virus was present in fecal and nasal swabs of one knockout pig. The same pig demonstrated evidence for production of SVA-specific antibodies; however, both knockout pigs did not exhibit virus neutralizing activity. Because founder pigs created by microinjection of the CRISPR/Cas9 system can have mosaic genotypes, a study on F1s is warranted. Overall, knocking out *ANTXR1* appears to confer protection against SVA infection in pigs, and modulation of this region may be needed to correct the phenotype associated with the edit.

**W1119: US National Animal Genome Research Program (NRSP8)**

**Protein Levels in Blood of Young Healthy Pigs as Indicators of Disease Resilience**

**Yulu Chen**, Iowa State University, Ames, IA

Disease resilience is the ability to maintain performance under pathogen exposure. Selection for disease resilience is difficult because nucleus breeding populations must be kept in a high health environment. Biomarkers for disease resilience that can be measured at an early age in high-health conditions could overcome this limitation. Proteins act as a working force to determine the organism's phenotype and the blood proteome has been used to identify biomarkers for some human diseases. Our objective was to explore the blood proteome of young healthy pigs for potential biomarkers of disease resilience. Seven batches of 60-75 healthy weaned Yorkshire x Landrace barrows (n=405) were entered into a quarantine nursery and blood samples were collected around 27 days of age. One week later, the pigs were moved to a nearby natural challenge facility, which was established by seeding the barn with pigs that were naturally infected with multiple pathogens and maintained using a continuous flow system. The levels of 481 proteins were quantified in the blood samples using the Tandem Mass Tag based LC-MS/MS method. Associations of the protein levels with performance and disease resilience were evaluated using mixed linear models and several proteins showed suggestively significant associations ( $p < 0.05$ ). The

abundance of 95 proteins was found to be moderately heritable ( $h^2 > 0.1$ ). Estimation of genetic correlations of protein abundance with resilience is underway. In conclusion, protein levels in blood of young healthy pigs show potential as biomarkers for disease resilience. Funding from USDA-NIFA, Genome Canada, PigGen Canada.

**W1120: US National Animal Genome Research Program (NRSP8)**

**Genomic, and Virulence Comparisons of Different Bacterial Isolates from BCO Lesions in Broilers**

**Nnamdi Simon Ekesi**, University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR

We used embryo lethality assay (ELA) to examine the virulence of bacteria isolates from bacterial chondronecrosis with osteomyelitis (BCO) lesions, the leading cause of lameness in broiler chickens. Lameness poses serious animal health and welfare issues, as well as, significant economic losses. Our hypothesis is that bacteria cross epithelia, and some survive in the blood to colonize the proximal growth plates through weaknesses in the vascular of the rapidly growing leg bones. We compared *Escherichia coli*, and *Staphylococcus* species, when we induced lameness using the wire-flooring model, as well as lameness outbreaks at commercial broiler farms. Differences in ELA, especially among the *E. coli* strains, prompted us to examine phylogenies using whole genome comparisons. *E. coli* isolates 1409 and 1413, from neighboring farms of the same integrator, showed very different ELA results and affiliate with divergent *E. coli* clades. *E. coli* isolate 1527, from a different farm and integrator, had similar ELA results to and a genome very similar to *E. coli* 1413. Isolate *Staphylococcus aureus* 1516 represented a common BCO isolate from the third farm and is mildly virulent in the ELA relative to *E. coli* 1413, *E. coli* 1527 and a very pathogenic human *S. aureus* isolate. The genome of *S. aureus* 1516 is most similar to isolates from deep wounds/lesions from chickens in Poland and more distantly to chicken isolates derived from human *S. aureus* in the United Kingdom. ELA allows virulence comparisons of distinct isolates when containment facilities are not available for live bird work.

**W1121: US National Animal Genome Research Program (NRSP8)**

**Analysis of Copy Number Variation in Jersey Dairy Cattle using Whole Genome Sequencing**

**Beth M Lett**, University of Wisconsin Madison, Madison, WI

Copy number variation (CNV) plays a role in disease resistance/susceptibility, fertility, milk production, and phenotypic variation. The objective of this study was to 1) identify CNVs in Jersey cattle using whole genome sequencing (WGS) and 2) validate predicted CNVs located within genes. WGS Illumina paired-end short reads at 15x coverage were generated for twenty Jersey AI sires and four USDA-MARC twinner sires, the latter representing a genetically diverse, outbred population. All sequences were aligned to the ARS-UCD1.2 reference genome assembly using BWA-MEM. CNV discovery utilized four methods: CNVnator, DELLY, LUMPY-Single sample, and LUMPY-Population. Consensus CNVs were identified by comparing results across the four methods within and among all samples (n=1,269 CNVs), Jersey only samples (n=740 CNVs), and non-Jersey only samples (n=59 CNVs). Putative Jersey-only CNVs were considered for validation after restriction to CNVs in functional gene regions (n = 86) and CNVs of low frequency outside functional regions (n=41). Validation consisted initially of comparing CNV locations with those reported in the database of genome variants (DGVa) from Ensembl and other published reports (included: 66/86 and 22/41). Putative CNVs not in DGVa or publications were validated using PCR to amplify deletion (duplication) or non-deletion (non-duplication) products. 3/20 (17 in progress) were validated by PCR. Among these validated, functional region CNVs, six were found to have no deletion homozygotes in our preliminary sample of 20 Jersey sires. Further investigation is warranted to determine if absence of homozygotes remains true in a larger sample of Jersey animals, suggesting the possibility of embryonic lethality.

**W1122: Weedy and Invasive Plant Genomics**

**Introgression of Cultivar Genes into Wild Carrot Populations**

**Johanne Brunet**, USDA-ARS, Madison, WI and Luciano Palmieri, ORISE, Madison, WI

Wild carrots are widespread in the USA, can be weedy and have been declared invasive in some states. Wild and cultivated carrots can easily hybridize as they are commonly found in close physical proximity and belong to the same species. Cultivar genes are introduced into wild carrot populations via pollinator-mediated gene flow and can spread within and among wild populations in a process called introgression. The extent of cultivar genes into wild US carrot populations has not been quantified. This is a critical question because, with the deployment of new gene editing technologies, the likelihood that genetically modified carrot cultivars will be released in the future has increased and some of these cultivar genes could magnify the invasiveness of wild carrots. To address introgression of cultivar genes into wild carrot populations, we sampled four wild carrot populations near (< 300m) and four far away (> 1000m) from cultivated carrots. In addition, we sampled wild populations at incremental 300m distances up to 1800m along three transects between one of the near and one of the far populations. Leaf tissue was sampled from 20 individuals per population, DNA was extracted and single nucleotide polymorphisms (SNPs) were identified against the carrot genome following genotyping by sequencing. We used fastSTRUCTURE on the SNP data to determine the genetic structure of the far and near wild populations, and the genetic structure of the cultivars grown at the West Madison Agricultural Research Station and the near and far wild populations. Cultivated and wild carrots were genetically differentiated. Moreover, near wild populations were more genetically diverse than far wild populations. To identify hybrids between cultivated and wild carrots, we ran STRUCTURE with 2 genetic clusters (K=2), forcing the cultivars as K=1 and the far wild populations as K= 2. We ran these analyses first on the near populations and then on all populations in the transects. We detected hybrid individuals in near wild populations. These results start addressing the spread of cultivar genes into wild carrot populations. The next steps will include pursuing these analyses on a larger scale while also determining whether specific regions of the genome are more prone to introgression.

## **W1123: Weedy and Invasive Plant Genomics**

### **The Whole Genome Sequence of the Obligate Root Parasitic Plant *Orobanche cumana* (sunflower broomrape)**

Jérôme Gouzy<sup>1</sup>, Nicolas Pouilly<sup>1</sup>, Luyang Hu<sup>1,2</sup>, Philippe Delavault<sup>3</sup>, Philippe Simier<sup>3</sup>, Marie-Claude Boniface<sup>1</sup>, Johann Louarn<sup>1</sup>, Olivier Catrice<sup>1</sup>, Sébastien Carrère<sup>1</sup>, Ludovic Cottret<sup>1</sup>, Ludovic Legrand<sup>1</sup>, Erika Sallet<sup>1</sup>, Stephane Cauet<sup>4</sup>, Céline Jézioriski<sup>4</sup>, William Marande<sup>4</sup>, Hélène Bergès<sup>4</sup>, Sonia Vautrin<sup>4</sup>, Álvaro Calderón González<sup>5</sup>, Begoña Pérez-Vich<sup>5</sup>, Leonardo Velasco<sup>5</sup>, Olivier Bouchez<sup>6</sup>, Cécile Donnadiou<sup>6</sup>, Clotilde Claudel<sup>7</sup>, Marie Coque<sup>7</sup>, Sebastien Faure<sup>7</sup>, Xavier Grand<sup>1,7</sup>, Nicolas Ribière<sup>7</sup>, Weijun Zhou<sup>2</sup>, Jun Zhao<sup>8</sup>, Christophe Jestin<sup>9</sup> and **Stéphane Muñoz**<sup>1</sup>, (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRAE/CNRS/Université de Toulouse, France, (2)College of Agriculture & Biotechnology, Zhejiang University, China, (3)Laboratoire de Biologie et Pathologie Végétales, SFR 4207 QUASAV, Université de Nantes, Nantes, France, (4)French Plant Genomic Center CNRGV - INRAE, France, (5)Instituto de Agricultura Sostenible - CSIC, Córdoba, Spain, (6)INRAE GeT Genomics Facility, Castanet-Tolosan, France, (7)Innolea, France, (8)Inner Mongolia Agricultural University, China, (9)Terres Inovia, Thiverval-Grignon, France

*Orobanche cumana* (sunflower broomrape) is an obligate parasitic plant that specifically infects sunflower (*Helianthus annuus*). It is one of the main limiting factors of sunflower crop in Eastern Europe, Spain and Asia. In 2007, the first infested fields have been reported in France. Breeding for resistance in sunflower was successful but new more virulent races of *O. cumana* often overcame the resistance genes.

The first developmental stages of *O. cumana* occur underground. The germination of the seeds is first stimulated by sunflower root exudates before entering the host root through a haustorium. Without roots nor photosynthesis activity, *O. cumana* depends on sunflower for water and nutrients supply. It connects to the vascular system of the sunflower root and store metabolites in a tubercle before emerging a single flowering shoot. A better understanding of the interaction will help in breeding new resistant sunflower varieties.

In the frame of a collaborative project between France and Spain, we have sequenced the 19 chromosomes (1.42 Gb) of the *O. cumana* genome by combining PacBio data, optical and genetic maps. Sixty RNASeq libraries corresponding to 20 developmental stages from seed to flower were used for annotating the genome and for describing the biochemical pathways.

Our strategy to assemble and annotate the genome sequence as well as results on worldwide populations diversity will be presented.

## **W1124: Weedy and Invasive Plant Genomics**

### **Canada Fleabane Genome Sequence**

**Martin Laforest**, Agriculture and Agrifood Canada, Saint-Jean-sur-Richelieu, QC, Canada, Eric Page, AAC-AAFC, Harrow, ON, Canada and Sara Martin, AAC-AAFC, Ottawa, ON, Canada

Canada fleabane is one of the most economically important weed species worldwide and infestation of herbicide resistant biotypes can result in significant yield losses across a range of cropping systems. As such, Canada fleabane has been the subject of many studies to understand how it has become resistant to herbicides. These studies have not yet identified a mechanism that explains non-target site glyphosate metabolic resistance. We have provided the first report of a chromosome-scale genome sequence for Canada fleabane. Third generation sequencing technology was used to create a genome assembly of 426 megabases, of which 9 chromosome-scale scaffolds cover more than 98% of the entire assembled sequence. This provides the information necessary to allow for the use of powerful genetic tools to detect and map the genes responsible for herbicide resistance. The knowledge gained with the aid of this new tool will be useful to create genetic tests for early diagnostic of resistance and, eventually, for control of this problematic weed. Additionally, the genome sequence will be a resource for studying genetic traits in asters, a large family that represents 10% of the diversity of flowering plants.

## **W1125: Weedy and Invasive Plant Genomics**

### **Genome-Wide Association and Genomic Prediction of Herbicide Resistance in Annual Ryegrass (*Lolium rigidum*) using Pool Sequencing Data**

**Jefferson F. Paril**, The University of Melbourne, Parkville, VIC, Australia and Alexandre Fournier-Level, The University of Melbourne, x, VIC, Australia

Herbicide resistance in weeds is a consistent problem in minimum tillage cropping systems, exacerbating both control costs and yield losses. Modern computational genomic approaches can deliver large-scale monitoring tools to tackle this issue at the whole-species level. These tools include the identification of the genetic basis of resistance using genome-wide association studies (GWAS) and the prediction of resistance from genomic data using genomic prediction (GP) models. GWAS and GP complementarily elucidate trait architecture and leverage high-throughput genomic data to predict phenotypes for weed management applications. Despite the declining costs of genotyping, resource allocation still needs optimization to maximize the power to detect quantitative trait loci (QTL) and prediction accuracy without the need for an exhaustive population sampling within a species. Here, we present our work on optimizing weed sampling design to capture genetic diversity within and among populations. We first perform a simulation study over theoretical landscapes using Approximate Bayesian Computation (ABC). It shows that for the simultaneous optimization of GWAS and GP performance, sampling more populations over the landscape with less individuals per population is better than extensively characterizing only a few. We applied these models to 62 annual ryegrass (*Lolium rigidum*) populations collected across South-Eastern Australia in 2018 and 2019, in conjunction with five herbicides (clethodim, glyphosate, sulfometuron, terbuthylazine and trifluralin). Genome-wide association using pool sequencing (Pool-GWAS) in the first two populations highlighted a number of putative QTL. Genomic prediction using pool sequencing (Pool-GP) of the same two populations for training and cross-validation showed 94.68% correlation between observed and predicted resistance. However, whole population Pool-GP on the other 60 populations resulted in 5-fold cross-validation prediction accuracies of 50% to 67% correlation coefficients. Least squares and ridge regression Pool-GP models performed better than the parsimonious variable selection models, which suggests that these herbicide resistance traits are highly polygenic. These results show that GWAS and GP are valuable predictive tools for monitoring herbicide resistance in the field. As the amount of genomic data continues to increase, there is a clear opportunity to adopt genomic diagnostics in the management of herbicide resistance.

## **W1126: Weedy and Invasive Plant Genomics**

### **The eccDNA Replicon, Adaptive Potential, and Functional Genomics in *Amaranthus palmeri***

**Christopher A. Saski**, Clemson University, Clemson, SC

Palmer amaranth (*Amaranthus palmeri* (S) Wats.) is one of the most competitive weed species that threatens crop production in the U.S., and is spreading across the globe. Palmer amaranth has an intriguing adaptive potential to respond to abiotic stresses, including herbicides, in short evolutionary time scales. Palmer amaranth has evolved resistance to seven different herbicide mechanisms of action (MOAs) including inhibitors of long chain fatty acid synthesis, acetolactate synthase, , photosystem II, mitosis, protoporphyrinogen oxidase, hydroxyphenylpyruvate dioxygenase, inhibitors of diverse growth responses (synthetic auxins), and 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS). Combined, these factors contribute to the aggressive nature of Palmer amaranth, resulting in significant yield losses across multiple cropping systems. Upon exposure and continued selection with glyphosate, the *Amaranthus palmeri* genome has undergone extensive shuffling to form an episome-like structure, termed the eccDNA replicon, that harbors the 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) gene and 58 other encoded genes. The eccDNA replicon is a massive, ~400kb extrachromosomal circular DNA (eccDNA), that autonomously replicates to increase copy number and expression of crucial genes required for plant survival under stress. Gene expression analysis under glyphosate stress showed transcription of 41 of the 59 genes, with high expression of *EPSPS*, aminotransferase, zinc-finger, and several uncharacterized proteins. The structure, expression profiles, replication mechanisms of the eccDNA replicon, and a method for functional genomics in Palmer amaranth will be discussed.

#### **W1127: Weedy and Invasive Plant Genomics**

##### **Role of CYP81A P450s in Metabolism-Based Cross-Resistance in *Echinochloa phyllopogon***

**Satoshi Iwakami**, Kyoto University, Kyoto, Japan

Extensive use of herbicides for weed management has led to the evolution of resistance in weeds, resulting in a great threat to agriculture. Among the diverse mechanisms of resistance that weeds have acquired, enhanced herbicide metabolism is the most threatening in agriculture, as the resistant plants could exhibit resistance to many herbicides from various chemical groups. Limited information on the enzymes involved in herbicide metabolism has hindered the prediction of metabolism-based cross-resistance in weeds. Cytochrome P450 superfamily is the enzyme system which is well-recognized to play a major role in herbicide metabolism although little is understood about responsible genes for herbicide metabolism. Members of CYP81A subfamily in plants including multiple-herbicide resistant *Echinochloa phyllopogon* were previously reported to metabolize diverse herbicide chemistries, suggesting a critical role of CYP81As in endowing unpredictable cross-resistances in weeds.

In this study, herbicide-metabolizing functions of all its 9 putative functional *CYP81A* genes of *E. phyllopogon* to 33 herbicides from 24 chemical groups were characterized by ectopic gene expression in *Arabidopsis thaliana* and *Escherichia coli*. The results concerning the difference of metabolizing functions among the CYP81A members and its practical application to cross-resistance prediction will be presented and discussed.

#### **W1128: What Early Career Scientists Need to Know: Developing and Executing Successful Broader Impact Programs for Current and Future Grants**

##### **Mentoring for Science, Mentor for Life**

**Jamie A. O'Rourke** and Michelle A. Graham, USDA-ARS-MWA-CICGRU, Ames, IA

Successful broader impacts should be rewarding for both the participants and the mentor. For some students, this may be their first opportunity to work in a laboratory setting. However, their time in the lab will be defined not just by the basic skills they learn, but by the lab environment and relationships they build with fellow interns, lab-mates and you, their mentor. This presentation will highlight ways to improve both the student and mentor experience.

#### **W1129: What Early Career Scientists Need to Know: Developing and Executing Successful Broader Impact Programs for Current and Future Grants**

## **The Dynamic Genome Program: An Experiential Learning Experience.**

**Alejandro Cortez**, UC-Riverside, Riverside, CA

The aims of the Dynamic Genome Program are to increase the retention of undergraduate life science majors and to be a launching point for their pursuit of careers in STEM. Dynamic Genome primarily consists of a laboratory course for first year students that strives to deliver an authentic research experience. Now in its eight year, the program can serve over 500 students on an annual basis. Students learn how to use bioinformatics and molecular tools in genome research. This two-unit course is offered every quarter and consists of six hours a week of direct instruction per section. Students also learn how to maintain a lab notebook (electronic). The initial four weeks of the course focus on basic lab techniques and core concepts while the second half of the course incorporates collaborative projects directed by faculty members from various departments in a “plug-and-play” model. A recent project included screening a full length cDNA library to create a suppression screen in *A. thaliana*. Another involved creating guide RNAs for knocking out gene targets in *A. thaliana* via CRISPR/Cas9. The course is taught by a lead instructor (DG staff), a graduate student teaching assistant, and an undergraduate laboratory assistant.

## **W1130: What Early Career Scientists Need to Know: Developing and Executing Successful Broader Impact Programs for Current and Future Grants**

### **Using Citizen Science to Communicate and Catalyze Plant Chemical and Genomic Research**

**Lucas Busta**, University of Nebraska Lincoln, Lincoln, NE

A perhaps unexpected challenge facing scientists is the public's perception and understanding of our enterprise. In part, it falls to we the scientists to remedy this problem. One avenue for communicating that we are actively working to solve real world problems is to engage the public in the scientific method itself. A bonus of enlisting such citizen scientists in a research project is that they can generate useful data in addition to gaining experience and insight into the research process.

This brief presentation will highlight one example of how citizen scientists can generate otherwise difficult-to-obtain datasets and how that same data can have substantial impact on a research program as a whole. In this example, citizen scientists received bioprospecting kits through the mail and used them to prepare samples of plants with rare, visible chemical phenotypes which they return to the laboratory for mass spectrometric analysis. By interacting with more than a dozen citizen scientists across the country, we have together discovered more than 50 plant species with unique abilities in chemical synthesis. These plants have in turn impacted multiple aspects of the project coordinator's chemical genomics research program.

## **W1131: Yam Genomics**

### **Varietal Change and Consumer Preferences in Roots, Tubers and Bananas: Yams in Côte D'Ivoire**

**Graham Thiele**, CGIAR Research Program on Roots, Tubers and Bananas (RTB), led by CIP, Lima, Peru

Varietal change in RTB crops in Sub-Saharan Africa has been relatively slow compared with other parts of the world. Yam in Côte d'Ivoire is presented as a case study to understand the different drivers of adoption. The study concludes that lack of attention to consumer preferred traits by breeding programs has contributed to the slow uptake of modern varieties (MV) and low varietal turnover, although other traits are also important. More attention needs to be given to understand the nature of consumer preferences in shaping demand and driving adoption, and incorporating these in product profiles for varietal development. Furthermore insufficient attention to gender differences in understanding consumer preferences for quality and post-harvest traits has contributed to inadequately described product profiles and is hence also linked to slow uptake of MVs. The importance of paying attention to demand and consumer preferences is very consistent with the new breeding paradigm of demand-led breeding with well-structured product profiles that consider gender differences and preferences of farmers as well as

preferences of rural and urban consumers. Building linkages from consumer preferences to trait identification and high throughput protocols for screening for quality traits is confirmed as a high priority for RTB and this can inform strategies to apply genomics-assisted breeding.

**W1132: Yam Genomics**

**Phenotypic and Genetic Variability in a Panel of White Guinea Yam (*Dioscorea rotundata*) Genotypes Based on Joint Analysis of Morphological and Molecular Diversity**

**Asrat Asfaw Amele**, IITA, Abuja, Nigeria and Kwabena Darkwa, Paterne Agre, Bunmi Olasanmi, Iseki Kotharo, Adrian Powel, Guillaume Bauchet, Ryo Matsumato, David Dekoeyer, Lukas Mueller, Suzy Strickler, Satoru Muranaka, Patrick Adebola, Robert Asiedu and Terauchi Ryohei

A better understanding of the nature and extent of phenotypic and genetic diversity among germplasm in a breeding program is vital for making progress with genetic improvement. This study assessed variability in a panel of white Guinea yam (*D. rotundata*) genotypes using joint analysis for phenotypic and genotypic diversity. The efficiency of different analysis methods to dissect the diversity in yam germplasm pool was assessed using 136,426 SNP markers and 23 morphological attributes. The degree of resemblance between the original distance and result from cluster configuration varied among the different dissimilarity matrices and hierarchical clustering methods. The average (UPGMA) method showed high goodness-of-fit with Gower, and the Identity by the state (IBS) distances as it produced the most significant cophenetic correction values. The diversity estimated among the 173 white yam accessions was higher with molecular than phenotypic data. The grouping of genotypes into useful clusters showed a high inconsistency between the phenotypic and molecular data due to the non-overlapping information among the dissimilarity matrices. However, joint analysis for phenotypic and molecular data produced higher diversity indices that good-enough to capture existing genetic variability in the yam germplasm pool. The results from our study have provided valuable insights to inform breeding strategies and identify promising divergent parents for the development of improved white yam varieties with acceptable end-user qualities.

**W1133: Yam Genomics**

**Plant Sex Prediction in White Guinea Yam (*Dioscorea rotundata*)**

**Patrick Olusanmi Adebola**, IITA Ibadan Nigeria, Ibadan, Nigeria and Paterne Agre, Chidinma Nwachukwu, Bunmi Olasanmi, Queen Obi, Adewumi Adeyinka, Idris Adejumobi, Jude Obidiegwu, Emmanuel Nwachukwu, Asrat Asfaw and David Dekoeyer

Plant sex in yam crop is complicated, with a monoecious and dioecious flowering pattern expressed by different genotypes. Plant sex determination with visual observation at the time of anthesis alone during the growth period can be very lengthy and time-consuming. Hence, early plant sex prediction at the seedling stage through molecular markers will be of great importance for designing an efficient hybridization plan by enabling the selection of parental clones with defined sex information. The objectives of this study were to optimize and validate the appropriate leaf sampling method for high-quality DNA extraction and predict yam plant sex at the early growth or seedling stage in *D. rotundata* genotypes using molecular markers. Five leaf tissue sampling and preservation methods: liquid nitrogen, silica gel, 95% ethanol, dry ice, and oven-drying before DNA extraction were assessed for quality DNA extraction. Plant sex prediction was attempted in one hundred and ninety (190) genotypes using Sp16 SNP and Dr-Actin marker at the seedling stage. The marker predicted yam plant sex was validated with the visual flower sex phenotype score at the blooming stage. Liquid nitrogen, silica gel, dry ice, and oven drying gave the best quality DNA for leaf sample preservation before DNA extraction. Plant sex prediction at seedlings stage via the sp16 marker revealed more ZW genotypes (female/monoecious phenotypes) in the studied materials than ZZ genotypes (male phenotype) with prediction accuracy of 81.50%. These results have highlighted the potential applicability of the SP16 marker for plant sex prediction in the white yam breeding program.

**W1134: Yam Genomics**

## Genetic Fidelity of Yam Planting Materials Produced with Novel High Ratio Propagation Technologies Based on Morphological, Ploidy and SSR Markers

**Morufat O. Balogun**, International Institute of Tropical Agriculture /University of Ibadan, Ibadan, Nigeria and Oluwatobi Oladejo, Norbert Maroya, Agre Paterne, Muranaka Satoru, Victoria Dan and Robert Asiedu

### W1135: Yam Genomics

#### Effect of Ploidy Levels on Flowering Behavior in Yams (*Dioscorea* spp.)

**Ranjana Bhattacharjee**, International Institute of Tropical Agriculture, Ibadan, Nigeria and Cobes P. Gatarira, Asrat Asfaw, Alzbeta Nemeckova, Eva Hribova, Jaroslav Dolezel, Victor Adetimirin

Yams (*Dioscorea*spp.) are staple food and cash crop for over 300 million people in the tropical and subtropical regions of the world. The genus comprises about 600 species, of which ten are commonly cultivated. Flowering - poor to no flowering and lack of synchronization - is among the major challenges for genetic improvement of the crop through cross breeding. The objective of this study was to evaluate the ploidy level among commonly cultivated species and their wild relatives and correlate with their flowering behavior. For this, 274 accessions across eight *Dioscorea*spp. were evaluated using three different methods including chromosome counting, flow cytometric estimation of nuclear DNA amounts, and DArTseq genotyping. The flow cytometry and chromosome counting showed the presence of diploids, triploids and tetraploids with 40, 60 and 80 chromosomes, respectively. For cultivated species, *D. rotundata* and *D. alata*, majority of the accessions were diploid, while other species such as *D. cayenensis*, *D. dumetorum* and *D. burbifer* were triploids and all *D. esculenta* with a few *D. alata* accessions were tetraploids. Interestingly, for a few *D. rotundata* accessions, there was inconsistency between nuclear DNA amounts and chromosome counts. This discrepancy indicated that these accessions actually belonged to *D. cayenensis*, or these are inter-specific hybrids between *D. rotundata* and *D. cayenensis* and their probable progenitors such as *D. abyssinica* and *D. prahensis*. Genotyping using DArTseq confirmed these results and indicated the complexity in yams in West Africa. This was further confirmed through flowering behavior wherein profused flowering was observed in tetraploids followed by diploids and very low flowering in triploids. The fertility viability test among flowering accessions showed lowest pollen germination among triploids. The results from this study confirmed the importance of ploidy analysis in yam breeding scheme for successful intra- and inter-specific crossing.

### W1136: Yam Genomics

#### Development of Genomic and Genetic Resources for Water Yam, *Dioscorea alata* L.

Jessen V. Bredeson<sup>1</sup>, **Jessica B. Lyons**<sup>1</sup>, Ibukun Ogunleye<sup>2</sup>, Eva Hřibová<sup>3</sup>, Matthew Parker<sup>4</sup>, Shengqiang Shu<sup>5</sup>, Joseph W. Carlson<sup>6</sup>, Robert Kariba<sup>7</sup>, Samuel Muthemba<sup>8</sup>, Katarzyna Knop<sup>4</sup>, Nicholas Schurch<sup>4</sup>, Antonio Lopez-Montes<sup>9</sup>, Michael Abberton<sup>10</sup>, Robert Asiedu<sup>11</sup>, Ramni Jamnadass<sup>12</sup>, Alice Muchugi<sup>13</sup>, Chiedozi N. Egesi<sup>14</sup>, Asrat Asfaw Amele<sup>15</sup>, Jonathan Featherston<sup>16</sup>, Gordon Simpson<sup>4</sup>, Jaroslav Doležel<sup>17</sup>, Prasad Hendre<sup>18</sup>, Allen Van Deynze<sup>19</sup>, Jude Obidiegwu<sup>20</sup>, Ranjana Bhattacharjee<sup>11</sup> and Daniel Rokhsar<sup>5</sup>, (1)University of California, Berkeley, Berkeley, CA, (2)International Institute of Tropical Agriculture, Nigeria, (3)Institute of Experimental Botany, Prague, Czech Republic, (4)University of Dundee, United Kingdom, (5)DOE Joint Genome Institute, Walnut Creek, CA, (6)DOE Joint Genome Institute, Berkeley, CA, (7)World Agroforestry, Kenya, (8)World Agroforestry (ICRAF), Kenya, (9)International Institute of Tropical Agriculture, Ibadan, Nigeria, (10)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, (11)International Institute of Tropical Agriculture, Ibadan, Nigeria, (12)World Agroforestry, Nairobi, Kenya, (13)World Agroforestry (ICRAF), Nairobi, Kenya, (14)National Root Crops Research Institute (NRCRI) Umudike, Umuahia, Nigeria, (15)IITA, Abuja, Nigeria, (16)Agricultural Research Council, South Africa, (17)Institute of Experimental Botany, Olomouc, Czech Republic, (18)WORLD AGROFORESTRY (ICRAF), Nairobi, Kenya, (19)University of California, Davis, CA, (20)National Root Crops Research Institute (NRCRI), Umudike, Nigeria

The water (or greater) yam *D. alata* is the most widely cultivated species of yam, distributed across much of the Pacific Islands, tropical Asia, the Caribbean, South America, and Africa. *D. alata* also distinguishes itself among the yams by its ability to yield in marginal soil, ease of propagation, high nutritional content, low glycemic index, and low post-harvest losses. Optimal growth and production are constrained, however, by traits such as starch quality and tuber flesh oxidation, and threats including the fungal disease anthracnose. Rigorous understanding of the genetic basis of important traits such as these will aid in the improvement of water yam to better meet the needs of consumers and farmers.

To facilitate the efficient elucidation of genotype-phenotype relationships in this species, we used DArTseq to genotype over 1200 *D. alata* progeny from ten mapping populations. We have optimized methods for working with this data type, such as filtering markers for quality and implementing pedigree-based phasing. Leveraging our chromosome-scale *D. alata* genome assembly version 2, we have generated genetic linkage maps from individual crosses; the integration of these maps into a composite framework map, and their use to perform QTL studies for anthracnose resistance and quality traits, will be discussed. An update on the reference genome assembly and annotation will also be given.

This work is funded by a grant from the BREAD program of the National Science Foundation to UC Berkeley (CA, USA), IITA (Ibadan, Nigeria) and NRCRI (Umudike, Nigeria).

#### **W1137: Yam Genomics**

##### **Genome-Wide Association Mapping of Tuber Flesh Oxidation and Dry Matter Content in Water Yam (*Dioscorea alata*) Panel**

**Angelot Paterne AGRE**, International Institute of Tropical Agriculture (IITA/IBADAN), Ibadan, Oyo, Nigeria and Pierre Cobes Gattariara, Ryo Matsumoto, Patrick Adebola, Robert Asiedu and Asrat Asfaw

Yam (*Dioscorea* spp.) is a versatile tuber crop grown in the tropics and sub-tropics as a preferred staple with many nutritional and medicinal significance. Among the cultivated yam species, water yam (*D. alata*) is most widely distributed and known for its good agronomic and tuber quality traits. A panel of 100 accessions was genotyped using DArTSeq and phenotyped for tuber dry matter content and tuber flesh oxidation at three different locations to identify a region of genome involved in controlling the variation for tuber quality traits that would facilitate the development of new cultivars of water yam with improved food quality. A population structure analysis using 7442 SNP markers covering the yam genome identified four sub-genetic groups with 70% admixture within them. Genome-wide linkage disequilibrium (LD) analysis demonstrated that the average LD was about ~6 kb. A marker-trait association analysis in a linear mixed model that involved four different gene actions: additive, general, dominance alternative and dominance reference with the admixture group as a covariate identified hotspot regions significantly associated with tuber dry matter content in chromosomes 15 and 19 that cumulatively explained 22.40% of the total phenotypic variation for the trait. Likewise, eight regions of the genome spread across six different chromosomes (1, 2,3,13,15, and 17) showed statistically significant associations with the variation in tuber flesh oxidation. Gene annotation for the regions with significant marker-trait associations revealed the presence of Peptidase C1A (IPR012599) and DEAD/DEAH (IPR011545) genes, which were previously reported as responsible for oxidation in many plants. An additional 28 candidate genes were identified in the peak SNP sites (or adjacent to these sites) for both tuber dry matter and tuber flesh oxidation with unknown functions. These results have elucidated the genetic architecture of dry matter content and tuber flesh oxidation in yam and revealed the suitability of GWAS in the identification of SNP variants associated with tuber food quality traits potentially applicable in yam breeding programs.

#### **W1138: Yam Genomics**

##### **Effect of Vine Position of Aeroponics Mother-Plant on the Performance of Seed Yam Production**

**Norbert Godonou Maroya**, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria and Morufat Balogun, Beatrice Aighewi, P. Lava Kumar, Mikail Gbadegesin, Lasisi Jimoh, Djana Mignouna, and Robert Asiedu

Yams, *Dioscorea rotunda*, tropical tuber crops are known for multiplication using entire or sett tuber. The project Yam Improvement for Income and Food Security in West Africa (YIIFSWA) has developed a new system for seed yam tuber production using single-node vine seedlings generated from Aeroponics System (AS). After three consecutive years of tubers harvest at six months from vine seedlings, it was observed a high variation among tubers with six categories of more than 1kg, between (501 -999g), (301 -500g); (151 -300g); (50-150g) and below 50g. Planted in AS, virus-free, and endophytic clean plantlets from the Temporary Immersion Bioreactor System of two white yam varieties Kpamyo and Asiedu revealed significant differences among the varieties for the number of leaves and height of plants in AS. At four months, single-node vines were cut from the top, the middle, and the bottom of the AS plants and planted in the nursery. At the end of five weeks, there was a significant difference in the survival and production of new shoots among the varieties and the three types of vines. The survival of the single-nodes vine from the top, middle, and bottom give respectively for Asiedu 94%, 55%, and 64%, and Kpamyo has 84%, 74%, and 70%. The single nodes rooted vines planted in the field were harvested at six months. The varieties and the vines have significant differences in the number of plants at harvest and the tubers' productivity. The yields (t/ha) from vines from the top, middle, and bottom were respectively for Asiedu 7.95t/ha, 0.87t/ha, and 0.31t/ha and for the variety Kpamyo 6.04t/ha, 1.27t/ha and 1.42t/ha. Considering the initial proportions of single node vine generated from the top 86%, and middle 7% and down 7%, and the difference in yields of these three types of vines, the study revealed that vines from middle and bottom were the sources of smallest tubers harvested from single-node vine seedlings. In conclusion, single-node vine seedlings for yam tubers production should be generated only from vines cut in the top portion of AS plants.

#### **W1139: Yam Genomics**

##### **Expanding Global Food Supply by Improving Shelf Life of Perishable Produce using Patented Gene Technology**

**Venkatramana Pegadaraju**, Agribody Technologies, San Diego, CA and Jerry Feitelson

Globally 1/3<sup>rd</sup> of the global food produced is wasted each year, which accounts for 1.3B tons and a loss of \$1 Trillion in retail value. Nearly, 40-50% of the wasted food is in roots and tuber, fruits and vegetable, with ~70% of losses in developed countries and ~30% in the emerging countries. Wasted food and its effects on people, the environment, and the economy has become a major topic of international conversation, and for good reasons. Consumers demand access to fresh food that can sustainably nourish them in an environmentally friendly manner. Food wastage can be significantly reduced by improving the shelf-life of perishable produce. A combination of superior plant genetics along with existing conventional chemical and physical solutions will offer an attractive solution to this complex global challenge.

Agribody Technology has a unique strategy that offers a compelling value proposition to our customers. As opposed to conventional approaches, our methods involve knocking down activity of a critical regulatory gene controlling plant cell fate that causes perishable crops to quickly decay. Since, this a built-in genetic mechanism, it offers multi-fold value in maintaining freshness of the produce from farm to fork. The efficacy of this technology is consistently proven in multiple crops under both greenhouse and field conditions. Currently, several licensing and co-development projects are underway in potato, tomato, sweetcorn, lettuce, canola, rice and others. Results from these various studies will be discussed.