W001: Abiotic Stress

Integrating Genomics and Phenomics to Study Genetic Control of Salinity Tolerance Traits

Mark A. Tester, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Technologies to enable increasingly powerful high throughput plant phenotyping are emerging at the same as high-density genotyping has become increasingly possible through cheap genome sequencing. As phenotyping now moves from being powerful to being cheap and powerful, so plant genetics is being transformed. In this talk, one example of how some of these changes in technology will be described, with work from our own research on the genetic basis of traits contributing to salinity tolerance. We will focus on the use of forward genetics to discover genes affecting salinity tolerance in barley, rice, and tomatoes, along with some recent genomics in quinoa, a partially domesticated crop with high salinity tolerance. Rather than studying salinity tolerance as a trait in itself, we use high-throughput phenotyping to dissect salinity tolerance into a series of components that are hypothesised to contribute to overall salinity tolerance. The application of this approach provides opportunities to significantly increase abiotic stress tolerance in crops and thus contribute to increasing agricultural production in many regions.

For barley, two consecutive years of field trials were conducted at a site with sandy soil and very low precipitation. Drip irrigation systems allowed the control of soil salinity. Both association mapping and nested association mapping populations were used to dissect physiologically and genetically complex traits in response to salt stress. Ten traits related to yield and yield components (e.g., days to flowering, harvest index, 100-seed mass) were recorded. Remarkably, one locus in common was identified in both populations that affected the ability of plants to maintain some yield components under saline conditions. This is now the subject of fine mapping, validation, introgression into commercial lines, and testing for effects on yield and salinity tolerance in field trials.

For tomatoes, the focus is on the genetics of tolerance in wild tomatoes, specifically *Solanum galapagense, Solanum cheesmaniae*, and *Solanum pimpinellifolium*. An association genetic approach is being taken. High-quality genome sequences have been made and genotyping-by-sequencing undertaken. Tomatoes have been phenotyped in The Plant Accelerator® and in the field for three years, and analyses are currently in progress.

For quinoa, the genome has been sequenced to high quality, and now 2,000 lines are being skim sequenced. Up to 1,300 of these lines are also being phenotyped in The Plant Accelerator and field trial sites in Australia, China, Germany, USA and elsewhere, with the aim of identifying natural variation in a range of domestication and tolerance traits.

W**002: Abiotic Stress**

Translational Regulation of *Root Length Density 1* Improves Root System Architecture through Auxin Pathways Leading to Better Soybean Yield Under Waterlogging and Drought Conditions

Heng Ye[1], Li Song[1], Lijuan Zhou[1], Liakat Ali[2], Chengjun Wu[3], Leandro A. Mozzoni[3], William Schapaugh[4], Pengyin Chen[2] and Henry T. Nguyen[1], (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

Genetic improvement of root system architecture (RSA) is essential for breeding crop varieties adapting to abiotic stress environments. Previously, a soybean quantitative trait locus (QTL) on chromosome 3 was found to regulate waterlogging (soil-flooding) tolerance through modification of root length density (RLD) under the stress. The favorable allele of this QTL was also found to have positive role in maintaining water status under water-limited condition in the greenhouse. Here, we demonstrate modification of RLD improves both waterlogging and drought tolerance through map-based cloning and characterization of *Root Length Density 1* (*RLD1*) underlying this QTL. Transgenic analysis confirmed *RLD1* was responsible for the natural variations of this QTL. The favorable allele of *RLD1* was found to have an insertion of 11-bp poly-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. *RLD1* was further identified to regulate auxin levels in pericycle cells to control root initiation and elongation. Introducing *RLD1* 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.

W**003: Abiotic Stress**

Systems Genetics Analysis of Heat Stress Tolerance in Rice

Andy Pereira, Crop, Soil and Environmental Sciences Department, University of Arkansas, Fayetteville, AR

W**004: Abiotic Stress**

Dissection of Physiological and Molecular Mechanisms of Drought Tolerance in Chickpea

Smith Millicent, ICRISAT, Hyderabad, India

W**005: Abiotic Stress**

A Novel NAC83 Transcription Factor from *Kalanchoe fedtschenkoi* enhances Drought and Salt Tolerance in *Arabidopsis*

Kumudu Rathnayake, University of South Dakota, Vermilion, SD, Travis Michael Garcia, Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Reno, NV, John C. Cushman, University of Nevada, Reno, Reno, NV and Bernard W.M. Wone, University of South Dakota, Vermillion, SD

Crassulacean acid metabolism (CAM) plants are considered well adapted to water-limited environments because CAM enhances water-use efficiency (WUE) compared to C_3 photosynthetic plants. Still, the underlying regulatory basis of CAM is essentially unknown. We have identified several candidate transcription factors (TFs) as potential key regulators of the CAM pathway and/or water-deficit stress response in the obligate CAM plant, *Kalanchoe fedtschenkoi*. Of these candidate TFs, one belongs to the NAC family. Many members of this family have been shown to function in plant growth and development and in the regulation of the transcriptional reprogramming associated with plant stress responses. The function of the *K. fedtschenkoi* NAC83 TF (*KfNAC83)* is not known in CAM or C_3 photosynthesis plants, but its *Arabidopsis thaliana* orthologue suggests roles in abiotic stress responses and development. We have functionally characterized this TF *via* overexpression in

Arabidopsis thaliana to determine its role in abiotic stress responses and development. Four independent transgenic lines overexpressing the *KfNAC83* gene were generated and used to explore the function of the TF in plant growth and development, drought, and salt tolerance. Subcellular localization imaging showed that the *KfNAC83* TF localized to the nucleus of root cells in transgenic *A. thaliana*. Overexpression of *KfNAC83* TF in *A. thaliana* exhibited positive impacts on plant growth and development, including significantly increased rosette size, leaves in the mature rosette, shoot biomass, number of siliques, and lateral roots compared to wildtype (WT). Under acute drought stress condition, *KfNAC83* overexpressed transgenic lines survived up to 15 days, indicating enhanced drought tolerance in transgenic lines of *A. thaliana* carrying the *KfNAC83* gene. Interestingly, these transgenic lines also showed significant increases in integrated WUE compared to WT. *KfNAC83* overexpressed lines also showed tolerance to 150 mM NaCl. Collectively, our results indicate that the *KfNAC83* TF functions as an important regulator of plant growth and development, and responses to water-deficit and salinity.

W006: **Abiotic Stress**

Combined Approach of Quantitative Genetics and Crop Modelling to understand Sunflower Tolerance to Drought

Nicolas Langlade[1], Pierre Casadebaig[2], Brigitte Mangin[1] and SUNRISE Consortium, (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France, (2)AGroécologie, Innovations et TeRritoires (AGIR) – INRA/CNRS, TOULOUSE, France

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

Demystifying Machine Learning

Nirav Merchant, University of Arizona, Tucson, AZ

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

Accelerating Crop Improvement with AI and Data Integration: CropOS

Jesse Hoff, Benson Hill Biosystems, St. Louis, MO

CropOS is a predictive platform that helps plant breeders and researchers unlock global genetic potential and realize crop performance improvements through novel and natural variation. Within the scope of our work developing CropOS™ we have catalogued and integrated genetic, phenotypic and genomic data from more than 100K accessions, from across corn, soybean, rice, barley, wheat, sorghum as well as model grass species *Seteria viridis,* to enable CropOS-Edit application.

The CropOS Edit application is a unique data integration and analytics engine that simultaneously integrates and analyzes population genetic, phenotypic, transcriptomic and metabolic data to present high probability gene and allele targets for Genome Editing for target phenotypes of interest. Results from Edit can be leveraged for either marker-assisted breeding or for Genome Editing. Platform's data architecture and capabilities will be presented with a use case.

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

Bioinformatics Analytic Tool Discovery Leveraging Text Mining and Natural Language Processing

Carol Lushbough, University of South Dakota, Vermillion, SD and Etienne Gnimpieba, University of South Dakota, Sioux Falls, SD

Bioinformatics and computational biology play an important role in biological research. As researchers design their experimental projects, one major challenge is to find the most relevant analytic tools that will lead to new knowledge discovery from their data. The Bio-TDS (Bioscience Query Tool Discovery Systems, http://biotds.org/) was developed to assist researchers in retrieving the most applicable analytic tools by allowing them to formulate their questions as free text. The Bio-TDS is a flexible retrieval system that allows users from multiple bioscience domains the ability to query over 15,000 analytic tool descriptions integrated from well-established, community repositories. One of the primary components of the Bio-TDS is the ontology and natural language processing workflow for annotation, curation, query processing, and evaluation. To enable accurate searching of the Bio-TDS repository information, it is important to have meaningful annotations describing the tools and their features. A rule-based (or predicate) semi-automatic curation process has been integrated into the Bio-TDS by combining human inspection and data mining methods. When users query the Bio-TDS, their queries are tokenized and processed against the NCBO BioPortal (http://bioportal.bioontology.org/) to extract ontology terms along with their URIs in order to enrich the query token set. Natural Language Processing is applied to the query (e.g. stemming, tagging, English synonym mapping) for further augmentation. The annotated query is then submitted to the Bio-TDS search engine leveraging the indexed, annotated analytic tool descriptions to retrieve the related tools ranked by relevance.

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

Developing Deep Learning Models and Reusable Machine Learning Workflows for Genomics in the Cloud: From Single Cell Images to Variant Effects

Jason Chin, DNAnexus, Mountain View, CA

Both bioinformatics computation and machine learning tasks need diversified software and hardware environments. Cloud computing has the advantage of providing scalability and flexibility that is suitable for diversified demand. Even so, proper integration of batch job systems for large data process and interactive exploratory data analysis utilizing open source tools remain the main focus of most engineering work for bioinformatics and data science. It is particularly challenging to manage the dependence of toolchains when the required software is in its infancy stage or the open source ecosystem becomes more complicated. DNAnexus has successfully developed a platform that can perform batch processing tasks from large-scale alignments, variant calling to more complex plant and animal genome assembly processes efficiently.

In contrast to the batch processing work, interactive computation needs more flexible computation environment management. We discuss how we can leverage the same platform on performing research computation for machine learning and deep learning tasks. More, we focus on utilizing Docker, Conda and Jupyter Lab to simplify interactive and reproducible research workflow with examples including microscope image processing with deep learning and variant effect evaluation with deep learning models in a cloud computing environment.

W011: African Orphan Crops

Plant Breeding Capacity Building in Africa: Focus on Underutilized Crops

Julia Sibiya, School of Agricultural, Earth & Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Efforts by various training programs including the Alliance for a Green Revolution (AGRA)-funded programs namely the Improved Masters in Cultivar Development (IMCDA), African Centre for Crop Improvement (ACCI), and West African Centre for Crop Improvement (WACCI), Regional Universities Forum (RUFORUM), Makerere Centre for Crop Improvement (MaRCCI) among many others have improved plant breeding capacity in the national agricultural systems (NARs), and Universities in some countries in Africa, for some crops. However, increasing population growth, and food and nutrition insecurity remain among the biggest challenges in sub-Saharan Africa, despite the rich agricultural diversity in the continent that can assist in overcoming hunger and poverty. Thus, demand for skilled plant breeders is still critical to fill gaps in plant breeding capacity in many countries in Africa. In addition, research programs have neglected orphan crops, which have the capability to sustain agricultural resilience against diseases, drought and other stresses. Thus, the UC Davis African Plant Breeding Academy (AfPBA) provided a platform for collaboration among participants, which resulted in the Mobreed (Mobility for Breeders in Africa) grant award from the Intra – Africa Mobility Scheme, a set up under the Pan-African Programme (Development Cooperation Instrument) managed by the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Union. Partners of MoBreed are senior researchers and alumni of the AfPBA, from University of Abomey-Calavi (Benin), University of KwaZulu-Natal (South Africa), University of Namibia (Namibia), Ebonyi State University (Nigeria) and Jimma University (Ethiopia), working with a European partner CIRAD from Montpellier, France. MoBreed is committed to train breeders (10 PhD and 38 MSc) who will develop, improve and promote 10 selected orphan crops, from the listed 101 underutilized food crops in Africa by the African Orphan Crop Consortium (AOCC) using the latest technologies of plant breeding. Furthermore, MoBreed aims at harmonizing training programs in plant breeding among its founding partners which will allow the recognition of curriculum and credits among the partner universities. To this effect, the e-learning platform (Plant Breeding E-Learning in Africa - PBEA) developed by Iowa State University in collaboration with the IMCDA partner universities (University of KwaZulu-Natal, Makerere University and Kwame Nkrumah University of Science and Technology) was adopted. Staff of these universities will also take part in staff exchange in order to advance their knowledge on the orphan crops targeted by MoBreed.

W012: African Orphan Crops

Bringing Yam (*Dioscorea alata*) into Genomics Age

Jude Obidiegwu, National Root Crops Research Institute (NRCRI), Umudike, Nigeria

Yam (*Dioscorea* spp.) is a vegetatively propagated crop cultivated for its underground edible tubers. It is a very important food and income source for millions of producers, processors and consumers across the world especially in West Africa. Within the *Dioscorea* genus *D. alata* has an advantage of broad environmental adaptability, early growth vigour, high yielding/ multiplication ratio, resistance to mosaic viruses, and low glycaemic index. The increasing global demand for *D. alata* is threatened by deteriorating soil structure and fertility, increasing levels of field and storage pests and diseases (e.g. nematodes, mealybugs, scale insects, and the fungal disease anthracnose); and high tuber losses in storage, tuber flesh oxidation, poor starch and textural quality and limited food form utilization. Breeding efforts are focusing on development of new breeding tools and strategies, trait capture and gene discovery, pre-breeding for new traits, development of new varieties incorporating consumer-preferred characters, while aligning research with farmer and end-user priorities. Our project is timely and within the last 2.5 years we have focused on the development of genomic resources in *D. alata* for accelerated breeding and improvement. Key achievement of our research effort includes the generation of a high-quality, chromosome scale reference genome. On-going genotyping effort of ten mapping populations, comprising over 1400 progeny, seeks to develop a composite map. We are presently phenotyping economically important traits including post-harvest and cooking qualities, starch quality, pest and disease profiles with a view to determine QTL. Global collection of *D. alata* has been assembled and understanding of sequence diversity is on course. This work is funded by a grant from the BREAD program of the National Science Foundation to PI Dan Rokhsar (UC Berkeley, USA), and co-PIs Ranjana Bhattacharjee (IITA, Ibadan, Nigeria) and presenting author (NRCRI, Umudike, Nigeria); and is in partnership with the African Orphan Crops Consortium.

W013: African Orphan Crops

Leveraging Genomic Tools for Sorghum Improvement in Sudan

Mayada Mamoun Beshir, Agricultural Research Corporation (ARC), Khartoum North, Sudan, Nafisa Elmahi Ahmed, Agricultural Research Corporation (ARC), Wad Medani, Sudan, Abdelbagi M. Ali Ghanim, International Atomic Energy Association (IAEA), Vienna, Austria, Patrick Okori, The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Lilongwe, Malawi and Damaris Achieng Odeny, The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Nairobi, Kenya

The agriculture sector is the main driver of the economy in Sudan with over 80% of the population engaged in agriculture and agriculture related production activities. Sorghum (*Sorghum bicolor* L.) is the staple cereal with increasing demand, which is reflected in the trend for area under sorghum production over the last years. However, crop productivity has not kept pace with increasing demand, due mainly to overwhelming production constraints. Climate change additionally influences the spread of foliar diseases such as turcicum leaf blight (TLB) and anthracnose caused by the fungus *Exserohilum turcicum* and *Colletotrichum graminicola*, respectively. Understanding the genetic variability and patterns of agro-morphological variation among Sudanese germplasm, alongside the use of advanced genomic tools, would lead to more efficient sorghum improvement in Sudan and mitigate food scarcity caused by biotic and abiotic factors. In the current study, 140 sorghum recombinant inbred lines at F12, derived from a cross between resistant and susceptible parents, were used to map quantitative trait loci (QTLs) associated with both TLB and anthracnose resistance. The mapping population alongside parents were phenotyped for both diseases at Wad Elturabi and Wad Medani research fields in Sudan and genotyped using DART-sequencing. We used 2,143 high quality polymorphic SNP markers evenly distributed across the sorghum genome to perform QTL analysis using R/qtl in R software 3.3.3. Preliminary results revealed that nine and six QTLs were significantly associated with TLB and anthracnose respectively, and that the phenotypic variation explained by each QTL ranged from 4 to 24%

and 3 to 13% for TLB and anthracnose, respectively. QTLs identified in this study corresponded to disease resistance QTL reported previously. These SNP markers showed potential usefulness in the characterization of sorghum genotypes carrying TLB and anthracnose resistant QTL.

W014: African Orphan Crops
Breeding Bambara Groundnut for West Africa

Stephen Amoah, Council for Scientific and Industrial Research - Crops Research Institute, Kumasi, Ghana

Bambara groundnut is an indigenous African grain legume grown for human consumption. It is well adapted to harsher conditions and constitutes an important part of the local diet, culture and economy. However, Bambara groundnut's geotropic pod development makes breeding through artificial hybridization difficult and is still cultivated from landraces rather than varieties developed for specific target environments or traits. Here we describe progress of Bambara groundnut improvement and techniques used. To identify conserved ortholog set markers Bambara groundnut DNA was hybridized to the Affymetrix ATH1 GeneChip®. This technique, called cross-species (x-species) is designed to use the high-density oligonucleotide arrays developed for model organisms to study other species. Several thousands of COS markers were identified using this technique. Other methods including mutation breeding using gamma irradiation have been used to generate variability and shorten generation cycle. Progress made with the use of genome-wide association studies (GWAS) for marker discovery in Bambara groundnut is also discussed.

W015: African Orphan Crops
Toward the Genome-Enabled Improvement of Shea Tree for Sub-Saharan Africa

Agyemang Danquah[1], Iago Hale[2], Francis Kwame Padi[3], Jean-Marc Boffa[4], Michael Barnor[3], Prasad Hendre[5] and Ramni Jamnadass[4], (1)West Africa Centre for Crop Improvement (WACCI), University of Ghana, Legon, Ghana, (2)University of New Hampshire, Durham, NH, (3)Cocoa Research Institute of Ghana, Eastern Region, Ghana, (4)World Agroforestry Centre, Nairobi, Kenya, (5)WORLD AGROFORESTRY CENTRE, Nairobi, Kenya

Shea (*Vitellaria paradoxa*) is arguably, the most economically and culturally important tree species to small holder farmers, especially women, across its natural range: the so-called "Shea Belt," a semi-arid region spanning 21 countries, from Senegal to Uganda. Despite its importance to rural livelihoods, shea remains a semi-domesticated tree species with virtually no history of systematic genetic improvement. The shea tree is not traditionally planted. Rather, the region's vast shea parklands result from the selective protection by smallholders of volunteer seedlings. Presently, these parkland populations are under threat due to agricultural land use intensification and competing resource needs. With global demand for shea on the rise, there is a clear need to convert this semi-domesticated tree into a more actively managed and improved agroforestry species. Shea has a very long juvenile period (up to 25 years). Thus, any such improvement strategy requires modern breeding tools, supported by publicly available foundational genetic and genomic resources. In this talk, we discuss ongoing efforts to develop a suit of breeding resources, including a reference genome and sequence variant, an AM and NAM populations, and also build regional capacity to facilitate the translational use of these resources in the long-term improvement of the species. Obviously, a sustained and coordinated investment in these regional efforts will hasten the domestication of shea and lead to the development of shea cultivars/varieties.

W016: African Orphan Crops
Genetic Diversity in Okra

Dorcas Ibitoye, U. Orkpeh, O.O. Oguntolu, C.O. Anyaoha, J.A.T. Olofintoye, P.E. Akin-Idowu, O.I. Aderonmu, J.I. Ikoro, O.A.K. Olomide and V.O. Ebre, National Horticultural Research Institute, Ibadan, Nigeria

Okra (*Abelmoschus esculentus* L.) is valued for its green fruits and it plays an important role in the diet of many Nigerians. Little attention however, has been paid to the improvement of this important vegetable despite its high nutritional and economic importance. The main objective of this study was to be able to pre-breed the Okra germplasm for efficient utilization in our breeding for improved novel Okra varieties having increased nutritive values, higher yielding and more resilient to biotic and abiotic stresses. In order to achieve this, okra genotypes were collected across the country. Among the 120 genotypes that were added to the genebank collection, 50 were selected to be evaluated for morphological variability using a 5 x 10 Randomized Incomplete Block Design (α – lattice). Quantitative traits such as 50 percent to flowering, number of branches per plant, number of ridges per fruit, length and breadth of fruit, number of fruits per plant and yield per plant (g) were studied while qualitative traits such as fruit pubescence, colour of petiole, fresh fruit colour and leaf shape were recorded for each accession. The study revealed high variability for the observed traits. High heritability estimates was recorded for number of fruits per plant and yield. The result of this study revealed that there is high variability that can be further exploited in our breeding programme for okra crop improvement.

W017: Analysis of Complex Genomes
Genome Assembly Gone Wrong: Lessons from the Vertebrate Genome Project (VGP)

Sergey Koren[1], Arang Rhie[1], Brian P Walenz[2], Sarah B Kingan[3], Edward S. Rice[4], Derek M. Bickhart[5], Benjamin D. Rosen[6], Stefan Hiendleder[7], Jessica L. Petersen[8], John Williams[7], Doreen Ware[9], Timothy P.L. Smith[10], Adam M. Phillippy[1] and VGP Assembly Group, (1)NHGRI, NIH, Bethesda, MD, (2)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (3)Pacific Biosciences, Menlo Park, CA, (4)University of Nebraska/Department of Animal Science, Lincoln, NE, (5)Dairy Forage Research Center, USDA-ARS, Madison, WI, (6)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (7)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia, (8)University of Nebraska, Lincoln, NE, (9)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (10)USDA, ARS, USMARC, Clay Center, NE

A complete and accurate genome sequence forms the basis of all downstream genomic analyses. However, even the human reference genome remains incomplete, which affects the quality of experiments and can mask true genomic variations. For most other species, high-quality reference genomes do not exist. Typically, reference projects focus on a single, typically inbred, individual to minimize heterozygosity and simplify assembly. This approach mixes haplotypes, hides variation, and introduces false duplication which causes errors in downstream analysis. "Trio binning" is designed specifically for heterozygous genomes resulting in a complete diploid reconstruction. We applied this

method to several human trios, fish, bird, and mammal genomes, including an F1 cross between two Bovinae species. As a result, we completely assembled both parental haplotypes with NG50 haplotig sizes >65 Mbp and 99.998% accuracy, surpassing all current mammalian assemblies. Trio binning may also be applicable to separating ancient species fusions to simplify complex genome assemblies. These haplotype-resolved assemblies enable precise surveys of structural variation to create more representative references, providing opportunities to study complex variation.

W018: Analysis of Complex Genomes

Novel Bioinformatic Resources and Multi-Omics Approaches enable the Characterization of Early Response to *Ca.* Liberibacter *asiaticus* Infection in *Citrus*

Surya Saha[1], Elizabeth Chin[2], John Ramsey[3], Mirella Flores-Gonzalez[1], Darya Mishchuk[2], Juan Chavez[4], Kevin Howe[5], XueFei Zhong[4], Jodi L. Humann[6], Elizabeth Mitrovic[7], MaryLou Polek[8], Kris Godfrey[9], Dorrie Main[6], Lukas A. Mueller[1], Jim Bruce[10], Michelle Cilia[3,11], Carolyn Slupsky[2] and PAG Authors' Group, (1)Boyce Thompson Institute, Ithaca, NY, (2)University of California, Davis, CA, (3)USDA ARS, Ithaca, NY, (4)Department of Genome Sciences, Seattle, WA, (5)Functional & Comparative Proteomics Center, USDA-ARS, Cornell University, Ithaca, NY, (6)Washington State University, Pullman, WA, (7)Con, Davis, CA, (8)USDA ARS, Nat'l Clonal Germplasm, Riverside, CA, (9)Contained Research Facility UC Davis, Davis, CA, (10)Department of Genome Sciences, University of Washington, Seattle, WA, (11)Cornell University, Ithaca, NY

A major challenge in studying eukaryotic gene expression on a genome-wide scale is the lack of tools for working with non-model systems that do not have a sequenced genome. Citrus greening disease is a devastating disease that is ravaging the citrus industry in the US and world-wide. Although several species of citrus have sequenced genomes, many varieties with commercial importance do not have high quality genomics resources. Thus, studies to understand varietal-specific impacts of citrus greening on tree metabolism and tree health are hampered. We collaborated with a team of researchers at the USDA ARS in Ithaca, NY, and the University of California, Davis to investigate the impact of citrus greening in a longitudinal study using transcriptomics, proteomics and metabolomics in an integrated, systems biology approach. The same trees were sampled over time for comprehensive 'omics profiling. One of the varieties used in the study, Lisbon Lemon, had no genomic resources. Thus, for the Lisbon lemon transcriptome assembly, we evaluated three approaches (a) *de novo* transcriptome assembly (b) reference-based transcriptome assembly and (c) assembly from available citrus reference genomes. We decided to use the *C. sinensis* v2 reference genome for determining gene expression and metabolic regulation in both Lisbon lemon (*C. limon*) and Navel orange (*C. sinensis*) to enable comparison of results and to avoid false positives from partially assembled transcripts in the *de novo* assemblies.

To facilitate data analysis, we created two open-access tools for data mining and visualization. A metabolic pathway database using BioCyc Pathway Tools software for the *C. sinensis* v2 reference genome was developed to visualize and analyze large-scale genomics, transcriptomics, proteomics and metabolomics data. We also developed Citrus Expression Network (CEN), which is an open-access and interactive web tool to analyze expression patterns in citrus. CEN facilitates effective data analysis by enabling simultaneous visualization of co-expressed genes to develop novel hypotheses in addition to candidate gene identification. These tools can be accessed at https://citrusgreening.org/ and https://www.citrusgenomedb.org/. The systems analysis showed that citrus greening targeted similar metabolic pathways and that the timing of induction of these pathways were slightly different. Overall, the optimized data analysis strategy we developed for non-model citrus species will facilitate gene expression and regulation studies on the impact of citrus greening disease and remediation strategies in commercially important citrus varieties without the need to sequence the genomes of each new citrus variety.

W019: Analysis of Complex Genomes

Recombination in AD-Genome Cottons

Luis M. De Santiago, Texas A&M University, College Station, TX

W020: Analysis of Complex Genomes

Identifying Transgenic Alleles and Locations in Crops using a Nanopore Sequencing Technology

Chi Zhang, Bin Yu and Thomas E Clemente, University of Nebraska - Lincoln, Lincoln, NE

Identification of the transgene insertion location in an efficient and high-throughput manner is still a challenge although transgenic technologies are powerful tools to improve traits of various organisms. Here, we report a Nanopore-based sequencing method to map transgene insertions in the soybean genome. After pulling down the target DNA sequences, nanopore MinIon sequencing was employed to generate long reads that cover both target sequences and soybean genome sequences. A bioinformatics pipeline was developed to identify target sequences and flanking sequences in long reads, and transgene insertions could be located by the genome-wide comparison between flanking sequences and the reference genome. Using long read sequencing, one can get long flanking sequences so that the insertion locations could be identified accurately in spite of the complexity of plant genomes. This method can be executed in a high-throughput manner - several ten samples could be pooled together in one flow cell, and multiple insertions in one genome could be discovered. This nanopore-based method is rapid, convenient, reliable, cost-efficient and high throughput, and can be directly extended to various organisms.

W021: Analysis of Complex Genomes

IsoPhase: Haplotyping using Full-Length Transcript Sequencing in a F1 Maize Hybrid Reveals Allele-Specific Expression

Elizabeth Tseng[1], Bo Wang[2], Kevin Eng[3], Primo Baybayan[3] and Doreen Ware[2], (1)PacBio, Menlo Park, CA, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)Pacific Biosciences, Menlo Park, CA

An important need in analyzing complex genomes is the ability to separate and phase haplotypes. While whole genome assembly can deliver this information, it cannot reveal whether there is allele-specific gene or isoform expression. Full-length transcript sequencing using long read technology has enabled researchers to characterize alternative splicing events and improve genome annotation. The PacBio Iso-Seq method, which can produce high-quality transcript sequences of 10 kb and longer, has been used to annotate many important plant and animal genomes. However, the general Iso-Seq algorithm ignores SNP-level information, focusing instead on identifying alternative splicing differences. We present an algorithm called IsoPhase that post-processes Iso-Seq data for transcript-based haplotyping.

For each gene, IsoPhase gathers the associated full-length reads, each representing a single transcribed molecule. It then calls SNPs and is able to infer the haplotype of the reads due to the full-length nature of the sequencing. The reads are sufficiently high-quality (usually >99%), thus requiring minimal error correction.
We applied IsoPhase to a maize Iso-Seq dataset consisting of two homozygous parents and two F1 cross hybrids. We validated the majority of the SNPs called with IsoPhase against matching short read data and identified cases of allele-specific, gene-level and isoform-level expression. We investigated causes of SNP calling errors in both short and long read data, which are often associated with multi-mapping and highly similar genes. Finally, we explored the computational challenges with extending haplotyping to tetraploid species using both simulated and real Iso-Seq data.

W022: Analysis of Complex Genomes

Identification of Candidate Domestication-Related Genes with a Systematic Survey of Loss-of-Function Mutations

Davoud Torkamaneh[1], Istvan Rajcan[1], Jérôme Laroche[2] and **Francois Belzile**[3], (1)University of Guelph, Guelph, ON, Canada, (2)Laval University, Québec, QC, Canada, (3)Laval University, Quebec, QC, Canada

Summary
Domestication is an important key co-evolutionary process through which humans have extensively altered the genomic make-up and appearance of both plants and animals. The identification of domestication-related genes remains very arduous. In this study, we present a systematic analytical approach that harnesses two recent advances in genomics, whole-genome sequencing and prediction of loss-of-function (LOF) mutations, to greatly facilitate the assembly of an enriched catalogue of domestication-related candidate genes. Using whole-genome sequencing data for 296 cultivated (*G. max*) and 64 wild soybean accessions, we identified 8,699 LOF variants, and 116 genes that are uniquely fixed for one or more LOF allele(s) in domesticated soybeans. Existing soybean transcriptomic data led us to overcome analytical challenges associated with whole-genome duplications and to identify neo- or sub-functionalized genes. This systematic approach allowed us to identify 110 candidate domestication-related genes in an efficient and rapid way. This catalogue contains previously well-characterized domestication genes in soybean, as well as some orthologues from other domesticated crop species. In addition, it comprises many promising candidate domestication genes. Overall, this collection of candidate domestication-related genes in soybean is almost twice as large as the sum of all previously reported candidate genes in all other crops. We believe this systematic approach could readily be used in wide range of species.
Key words: domestication, whole-genome sequencing, loss-of-function mutations, whole-genome duplication, domestication-related genes

W023: Animal Epigenetics

Integrated Epigenome and Transcriptome Analysis for Reduction of Boar Taint in Pigs

Haja N. Kadarmideen, Technical University of Denmark, Kgs. Lyngby, Denmark

W024: Animal Epigenetics

Differences in Promoter Accessibility Responses to Hypoxia in Pulmonary Arteries of Ascites-Susceptible and Resistance Broiler Research Lines Detected using ATAC-Seq Technology

Douglas D. Rhoads, University of Arkansas, Cell and Molecular Biology Program, Fayetteville, AR

The aim of our study was to map genome-wide changes in chromatin accessibility of ascites- susceptible and ascites- resistant bird lines comparing growth under normal and hypoxic conditions. Ascites is a terminal result of pulmonary hypertension and is a significant metabolic disease of fast growing meat-type chickens. Pulmonary artery remodeling appears to be the main condition that leads to an increase of pulmonary vascular resistance, sustained arterial hypertension, right ventricular hypertrophy and ultimately death. Therefore, we investigated chromatin accessible regions in the pulmonary artery of both genders under normal and hypoxic conditions for two broiler experimental lines divergently selected for ascites phenotype via ATAC-seq technology. Transposition tagging was on nuclei from frozen pulmonary artery tissues. Libraries were sequenced to generate 50 million 2x150 PE reads and mapped to the Galgal5 assembly. A total of 23,444 open chromatin peaks were identified across all pulmonary artery samples. Our analysis showed that the peaks identified chromatin changes in promoters,, exons, introns, intergenic regions, and transcription start sites. Initial results demonstrate that there was a substantial increase in the chromatin accessibility throughout the genome of ascites- susceptible birds when challenged under hypoxic conditions in comparison with those at ambient oxygen levels. Contrastingly, we observed reduced changes in chromatin accessibility regions in ascites-resistant birds when challenged. When we limited our analyses to changes within 2 kb of transcription start sites we identified 1324 regions that become differentially accessible. In conclusion, we showed that chromatin accessibility is a key epigenetic factor influencing transcriptional regulation and a straightforward approach to identify functional genomic regulatory regions controlling complex diseases such as ascites in birds.

W025: Animal Epigenetics

Genome-Wide DNA Methylation Analysis Associated with Feed Efficiency in Angus and Hereford Beef Cattle

Shuli Liu, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

W026: Animal Epigenetics

Epigenetic Footprints in Chicken Marek's Disease Resistance

Jiuzhou Song, University of Maryland, College Park, MD

W027: Animal Epigenetics

Sperm Chromatin Dynamics and Bull Fertility

Erdogan Memili, Mississippi State University, Starkville, MS

Bull fertility is vital for cattle reproduction, production and product quality. Despite producing ample amounts of sperm with normal morphology and motility, some bulls have been afflicted with low fertility. The objectives of our studies were to test the hypothesis that sperm epigenetic characteristics of chromatin dynamics influenced by the nuclear proteins determine bull fertility. To accomplish this aim, nuclear proteins

Protamine 1 (PRM), Testis Specific Histone 2B (TH2B), Histone 3 (H3) and Histone 4 (H4) were analyzed in sperm from Holstein bulls with field fertility using flow cytometry, immunocytochemistry, Western blotting, and computational biology. The results showed that sperm levels of PRM and the Histones exhibit positive or negative correlations with bull fertility, respectively, and that proteins sequences of PRM1 and Histones are highly conserved across mammals. These results are significant because they help advance fundamental animal science and technology. In addition, because of the similarities in physiology and genetics between cattle and other mammals, the findings can be applied to study and reproductive biotechnology of other species including humans and endangered animals.

W028: **Animal Epigenetics**
The Role of DNA Methylation in CD4+ T-Cell Lineage Commitment in Immune Response Biased Dairy Cattle
Marlene Ann Paibomesai, Ontario Ministry of Agriculture, Food and Rural Affairs, Elora, ON, Canada

W029: **Animal Genomics and Adaptation to Climate Change**
Molecular Basis of Nutrient Utilization and Gene Expression of Transporters in Heat-Stressed Broilers
Samuel E. Aggrey[1], Walid S. Habashy[2], Kwaku Adomako[3], Marie C. Milfort[1] and Romdhane Rekaya[1], (1)University of Georgia, Athens, GA, (2)Damanhour University, Damanhour, Egypt, (3)Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
Heat stress (HS) causes nutrient metabolism, cellular and molecular changes that affect productivity and welfare of growing birds. The significant losses in productivity as a result of HS are expected to increase due to continuous rise in global temperatures. The effect of HS on intestinal integrity, digestibility of protein, fat and amino acids and the expression of nutrient and amino acid transporters were investigated in broilers. The total feed consumption and retention of protein, fat and amino acids were significantly lower in the HS group compared to the control group. The apparent ileal digestibility was consistently higher for all essential amino acids in the HS group but was not significant. The increase in endotoxin concentration and changes in expression of tight junction genes in the HS group points to compromised intestinal integrity. This may render heat-stressed birds susceptible to enteric infections. The dynamics of amino acid transporters in the *P. major* and ileum was influenced by HS. In *P. major* and ileum tissues at 1-d post HS, transporters FATP1 and SGLT1 were down regulated in the HS group while FABP1 and PepT1 were down regulated only in the ileum of the HS group. The nutrient transporter FABP1 at 12-d post HS was down regulated in the *P. major* and ileum but GLUT1 and PepT2 were down regulated only in the ileum and PepT1 was down regulated only in the *P. major* compared with the control group. In *P. major* and ileum tissues at 1-d post HS, transporters SNAT1, SNAT2, SNAT7, TAT1 and $b^{0,+}$AT, were down regulated in the HS group. Meanwhile, LAT4 and B^0AT were down regulated only in *P. major* in the HS group. The amino acid transporters B^0AT and SNAT7 at 12-d post HS were down regulated in the *P. major* and ileum but SNAT2 was down regulated only in the ileum and TAT1 was down regulated only in the *P. major* compared with the control group. These changes in nutrient and amino acid transporters could affect nutrient metabolism and subsequently affect body weight changes.

W030: **Animal Genomics and Adaptation to Climate Change**
Workshop Introduction
Susan J. Lamont, Department of Animal Science, Iowa State University, Ames, IA
Introduction to the workshop topics and speakers.

W031: **Animal Genomics and Adaptation to Climate Change**
Catfish Genetic Studies Relevant to the Changing Climate
Zhanjiang Liu[1], Xiaozhu Wang[2], Yulin Jin[2], Yujia Yang[2], Tao Zhou[2] and Dongya Gao[2], (1)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY, (2)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL
With a changing climate, it is likely that we will experience weather patterns with more extreme conditions such as extremely high and low temperatures. Such weather patterns will directly impact agricultural production of plants and animals. With aquaculture, increased temperature would translate into higher water temperature, which in turn, lead to lower dissolved oxygen in the water. As a result, fish are required of greater tolerance to high temperature as well as to low dissolved oxygen.
In this lecture, I will summarize our recent genetic studies on whole genome associate studies of heat tolerance, low oxygen tolerance, and epigenetic regulation of sex reversal in catfish. Three significant associated SNPs were detected by GWAS analysis. The SNP located on linkage group 14 explained 12.1% of phenotypic variation. The other two SNPs, located on linkage group 16, explained 11.3% and 11.5% of phenotypic variation, respectively. Genes with the QTL regions included *TRAF2, FBXW5, ANAPC2, UBR1* and *KLHL29*— with known functions in protein degradation through the ubiquitination pathway. Genetic architecture for tolerance to low oxygen appeared to be extremely complex, with multiple interspecific, intraspecific, and within strain QTL. Analysis of gene pathways related to the low oxygen tolerance suggested that multiple effector pathways of oxygen metabolism accounted for the complex genetic architecture of QTL for low oxygen tolerance.
High temperature not only affect fish survival, but also their reproduction. Catfish generate excessive percentages of females under high temperatures during early development and sex differentiation. This process is highly likely to be regulated at epigenetic levels, and analysis is ongoing to demonstrate epigenetic control of sex determination and sex differentiation in catfish. Preliminary results will be presented, and discussed.

W032: **Animal Genomics and Adaptation to Climate Change**
Genetics of Heat Stress Response in Cattle
Raluca Mateescu, Department of Animal Sciences, University of Florida, Gainesville, FL

W033: **Animal Genomics and Adaptation to Climate Change**
Molecular Response to Heat Stress in Turkeys

Gale Strasburg, Michigan State University, East Lansing, MI

W034: **Animal Genomics and Adaptation to Climate Change**
Molecular Basis of Nutrient Utilization and Gene Expression of Transporters in Heat-Stressed Broilers
Samuel E. Aggrey, University of Georgia, Athens, GA

W035: **Animal Genomics and Adaptation to Climate Change**
Genetic Factors Associated with changes in Feeding Behavior due to Elevated Temperature
Gary A. Rohrer[1], Amanda J. Cross[2], Brittney N Keel[1], Tami Brown-Brandl[1] and Joseph Cassady[3], (1)USDA, ARS, USMARC, Clay Center, NE, (2)DNA Genetics, Columbus, NE, (3)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 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). Females had a greater reduction in feeding time due to elevated temperature than males. Breed of sire differences were also observed as Duroc-sired pigs were marginally affected, Landrace-sired pigs were severely affected, and Yorkshire-sired pigs were intermediately affected by elevated temperatures. To avoid population stratification effects in 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. Genetics differences in feeding behavior of grow-finish pigs due to increased temperatures were observed. Selection targeting these genetic differences can produce pigs that are more tolerant to elevated ambient temperature and genetic markers will expedite this selection response.
USDA is an equal opportunity provider and employer.

W036: **Aquaculture**
FAASG - Functional Annotation of All Salmonid Genomes
Ben F. Koop, University of Victoria, Victoria, BC, Canada
As reference genomes for new species are added across the tree of life and as we transition towards the ability to generate full-length representations chromosomes,new opportunities in aquaculture can be obtained through comparative biology and the differential functional analysis of species. This includes documenting variation across the species through genome re-sequencing and analyzing genetic variation to gene expression patterns in various environments.
Salmonids with over 70 different economically and culturally import species provides an opportunity to examine Whole Genome Duplications (WGD) and mechanisms in the i) evolution of species, ii) specialization of molecular functions, iii) appearance of new functions, and iv) generation of system redundancy and adaptability. Central to this study is genome annotation that includes chromosomes order and structural variation, SNPs, gene identification, expression, and regulation patterns as well as epigenetic factors that can link environmental factors to genetic programs. Core assays proposed by FAANG include gene expression studies (RNA-Seq), chromatin accessibility and methylation (ATAC-Seq/BS-Seq) and chromatin histone modification (H3K4me3, H3K27me3, H3K23ac, H3K4me1 through ChIP-Seq). With the exception of RNA-Seq, these assays have not been used extensively in fish and their characterization and robustness have not been established. We will explore functional annotation of genomes and where this can lead. Successful annotation of salmonid genomes will provide essential links between genetic and environmental systems and their relationships to physiological, biochemical processes in health, disease, aquaculture, conservation and environmental protection strategies.

W037: **Aquaculture**
Exploiting the Dimensionality of Genomic Information in Channel Catfish
Ivan Pocrnic[1], Daniela A.L. Lourenco[1], André L. S. Garcia[1], Geoff Waldbieser[2], Brian Bosworth[2] and Ignacy Misztal[1], (1)University of Georgia, Athens, GA, (2)USDA-ARS Warmwater Aquaculture Research Unit, Stoneville, MS
Genomic selection (GS) is successfully applied in animal breeding based on single-nucleotide polymorphism (SNP) markers. While the number of SNP can be very large, especially with sequence data, the dimensionality of gene content is limited by limited effective population size and subsequently by limited number of the independent chromosome segments. When number of SNP markers and/or number of genotyped animals is large enough, it is possible to estimate all chromosome segments precisely, leading to perfect accuracy of GS. In that case, the dimensionality can be obtained via the number of non-negligible singular values of gene content, or the number of non-negligible eigenvalues of the genomic relationship matrix (GRM) that explain 98% of the variation. When a small amount of data is available, the "98%" is depressed and 4 times the "90%" number is a better estimate. In channel catfish, the available genomic information consisted of 2911 fish genotyped for 57k SNP. Interpolated number of eigenvalues explaining 10, 50, 80, 90 and 98% of variation in GRM were 3.8, 71.9, 571.2, 1100.9 and 2140.9. Subsequently, the number of independent chromosome segments is around 4,400. Knowing the dimensionality of genomic information can be a useful tool to estimate the optimal number of SNP markers and/or genotyped animals needed for GS and variant discovery.

W038: **Aquaculture**
Polygenic Nature of Genomic Loci Associated with Early Maturation in Two Pacific Salmonids, Coho Salmon (*Oncorhynchus kisutch*) and Chinook Salmon (*Oncorhynchus tshawytscha*).
Michelle T.T. Crown, Simon Fraser University, MBB Department, Burnaby, BC, Canada

Michelle T.T. Crown[1,2], Kris A. Christensen[2,5*], Sarah J. Lehnert[3*], Ruth E. Withler[4], Janine Supernault[4], Eric B. Rondeau[2,5], Ben F. Koop[5], Daniel D. Heath[3], Robert H. Devlin[2§] and William S. Davidson[1§]
[§,*] Both pairs of authors contributed equally to this work.
(1)Simon Fraser University, Molecular Biology and Biochemistry Department, Burnaby, BC, Canada,
(2)Fisheries and Oceans Canada, CAER, West Vancouver, BC, Canada,
(3)Great Lakes Institute for Environmental Research, University of Windsor, Windsor, ON, Canada
(4)Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada,
(5)University of Victoria, Victoria, BC, Canada
Reproductive success of salmon is influenced by multiple characteristics, including: morphological features, physiological performance, and mating strategies. Variation in the number of years spent at sea before spawning, age of maturity, is an evolutionary stable strategy that can also influence reproductive success where fitness is balanced between early and late maturing salmon. The genetic basis of early maturation (jacking in Pacific salmon and grilsing in Atlantic salmon) is not well understood, and it is unclear if the genetic basis for early maturation among salmonids is evolutionarily conserved. Although recent work has found that 39.4% of the phenotypic variation can be explained by a single locus *Vgll3* in European Atlantic salmon, other data suggests this genetic association may not be fully conserved in North American Atlantic salmon. Our current study aims to identify the loci associated with early maturation in two Pacific salmonid species, coho (*Oncorhynchus kisutch*) and Chinook (*O. tshawytscha*) salmon. We conducted a genome-wide-association-study on six families of coho and three families of Chinook salmon from British Columbia. Using the GBS approach, *EcoT22I* reduced representation libraries were generated for 716 coho and 474 chinook individuals and sequenced on the Illumina HiSeq platform. Current data analyses revealed highly polygenic associations with jacking is distinct from *Vgll3* in coho and Chinook salmon populations. Further sanger sequencing and TaqMan analyses on *Vgll3* confirms the lack of association in coho. This suggest that the molecular mechanisms underlying age of maturity may not be evolutionarily conserved among Atlantic salmon, Pacific coho and Chinook salmon.

W0**39: Aquaculture**

Crustacean Genome Exploration Reveals the Evolutionary Origin of Deadly Shrimp Virus

Satoshi Kawato, Hidehiro Kondo and Ikuo Hirono, Tokyo University of Marine Science and Technology, Tokyo, Japan
White spot syndrome virus (WSSV) is a crustacean-infecting, double-stranded DNA virus and is the most serious viral pathogen in the global shrimp industry. WSSV is the sole recognized member of the family *Nimaviridae*, and the lack of genomic data on other nimaviruses has obscured the evolutionary history of WSSV. Here, we investigated the evolutionary history of WSSV by characterizing WSSV relatives hidden in host genomic data. We surveyed 14 host crustacean genomes and identified five novel nimaviral genomes. Comparative genomic analysis identified 28 “core genes” that are ubiquitously conserved in *Nimaviridae*; unexpected conservation of 13 uncharacterized proteins highlighted yet unknown essential functions underlying the nimavirus replication cycle. The ancestral *Nimaviridae* gene set contained five baculoviral *per os*infectivity factor homologs and a sulfhydryl oxidase homolog, suggesting a shared phylogenetic origin of *Nimaviridae* and insect-associated double-stranded DNA viruses. Moreover, we show that novel gene acquisition and subsequent amplification reinforced the unique accessory gene repertoire of WSSV. The expansion of unique envelope protein and nonstructural virulence-associated genes may have been the key genomic event that made WSSV such a deadly pathogen. Our work redefines the previously poorly-characterized crustacean virus family and reveal the ancient genomic events that preordained the emergence of a devastating shrimp pathogen.

W0**40: Aquaculture**

Genetics of Pacific Oyster Uniformity in Different Environments

Konstantin Divilov, Blaine Schoolfield and Chris Langdon, Oregon State University, Newport, OR
Uniformity in size and shape of Pacific oysters grown for the half-shell market would reduce labor costs associated with grading oysters and lost income from discarding oysters at harvest that are too small to sell. Forty-eight Pacific oyster families were planted in Oyster Bay, Washington and Willapa Bay, Washington in 2015 and harvested in 2017 to test whether family uniformity of oysters was under genetic control. Twenty individual oysters from each family in each location were phenotyped for whole oyster weight, wet meat weight, dry meat weight, and shell depth/length ratio, which has been previously shown to be an indicator of ideal shell shape. A Bayesian Gaussian process regression model with heteroscedastic noise was used with each trait to estimate the posterior mode of variance for each family after accounting for genetic effects, which were estimated using pedigree information. Subsequently, these posterior modes were used as phenotypes in another Bayesian Gaussian process regression model to obtain 95% credible intervals of heritability for uniformity for each trait. Wet meat weight uniformity in Oyster Bay and whole meat weight and depth/length ratio uniformity in Willapa Bay were significantly heritable. When combining data from both environments, only wet meat weight uniformity was significantly heritable, and the wet meat weight uniformity breeding values had a significant negative correlation with wet meat weight breeding values. We expect uniformity in important oyster traits to be difficult to achieve through breeding given environmental effects and undesirable genetic correlations.

W0**41: Aquaculture**

Editing for Animal Welfare and Environmental Sustainability: Are These Traits Important?

Tad S. Sonstegard, Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN

W0**42: Aquaculture**

Development of a Universal Sex Assay and Identification of y-Chromosome Haplotypes in Chinook Salmon.

Garrett J McKinney, University of Washington, Seattle, WA
Sex has important influence on behavior, growth rate, and age at maturity in salmon. As a result, sex is an important consideration for researchers, managers, and hatcheries. For researchers, it is important to control for the sex of individual fish in analyses as sexes may differ in traits or response to treatments. For managers, sex of an individual can inform whether there are sex-biases in stage-specific mortality, differences in return timing, differences in age composition of a stock by sex, and sex-bias in harvest. For hatcheries, knowledge of sex is

important when broodstock are chosen prior to maturation and to identify if there are sex-biases in post-release mortality. Several molecular markers have been developed for genetic sex identification of Chinook salmon; however, these have displayed inconsistent accuracy when tested throughout the species range due to a combination of incomplete linkage between sex-associated markers and the true sex-determining region and the presence of non-functional male sex determining regions in female fish. To improve accuracy of sex identification we developed a GTseq assay to directly sequence the exons of SDY, the sex determination gene in Chinook salmon, and evaluate the accuracy of this assay in populations throughout the species North American range. In addition, we describe the discovery of y-chromosome haplotypes that are associated with variation in size and age at maturity in male Chinook salmon.

W043: Aquaculture

From Sea to Plate: Genomically Enabling the Australasian Snapper (*Chrysophrys auratus*) for Aquaculture

Maren Wellenreuther, Plant and Food Research, Nelson, New Zealand
Maren Wellenreuther[1, 2], David Ashton[1], Elena Hilario[3], Ross Crowhurst[3], Peter Morrison Whittle[1], Andrew Catanach[4], Jeremy Le Luyer[5], Cecilia Deng[3], Charles David[4], Peter Ritchie6, and Louis Bernatchez[7]
[1]The New Zealand Institute for Plant & Food Research Ltd, Nelson, New Zealand
[2]School of Biological Sciences, University of Auckland, Auckland, New Zealand
[3]The New Zealand Institute for Plant & Food Research Ltd, Auckland, New Zealand
[4]The New Zealand Institute for Plant & Food Research Ltd, Lincoln, New Zealand
[5]Ifremer, UMR 241 Ecosystèmes Insulaires Océaniens, Centre Ifremer du Pacifique, BP 49, 98725 Tahiti, Polynésie française
[6]School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand
[7]Institut de Biologie Intégrative et des Systèmes (IBIS), Pavillon Charles-Eugène Marchand, Université Laval, Québec, Canada
*corresponding author: maren.wellenreuther@plantandfood.co.nz

The Australasian snapper (Sparidae: *Chrysophrys auratus*) has been identified as a promising candidate for aquaculture, and a breeding programme was initiated in 2016 to select for faster growth. Here we report on recent developments to genomically enable this species for aquaculture breeding by 1) presenting a high-quality genome assembly – which integrates short-insert and mate-pair sequencing reads along with optical mapping and a linkage map - that represents a valuable resource to aquaculture programmes around the world using closely related bream species. We further present 2) a catalogue of SNPs and structural variants across the genome from 12 whole-genome re-sequenced wild snapper to both characterise the extent of genetic variation in this species and to quantify the extent of variant overlap with coding regions as well as those under putative selection. We then 3) use a GBS dataset of a hatchery reared and phenotyped pedigree to describe the first QTLs for growth in this species, and highlight some of the key findings from a growth-temperature rna-seq experiment on wild and domesticated snapper to elucidate the genes responsible for growth under different temperatures regimes. Finally, we will discuss how these genomic findings can be used to inform the future research directions of the breeding programme in this commercially and culturally important species.

W044: Aquaculture

Reverse Vaccinology: Functional Genomics Applied to the Development of Vaccines against *Caligus rogercresseyi*

Cristian Gallardo, Universidad de Concepcion, Concepcion, Chile
Cristian Gallardo-Escárate, Valentina Valenzuela-Muñoz, Gustavo Nuñez-Acuña, Diego Valenzuela-Miranda. Interdisciplinary Center for Aquaculture Research (INCAR), University of Concepcion, Chile.

One of the major diseases affecting the Chilean salmon farming is the "Caligidosis" caused by the copepod *Caligus rogercresseyi*. Our research group has intensively studied the host/parasite interactions between Atlantic salmon and *C. rogercresseyi* using functional genomics as an approach. The present study reports the efforts to develop a new vaccines against sea lice based on molecular information using high-throughput transcriptome sequencing technology. The life cycle of *C. rogercresseyi* comprises eight development stages: nauplius 1-2, copepodid, chalimus 1-4, and adult. Herein, twenty individuals from each instars of *C. rogercresseyi* were collected and separately sequenced by Illumina and Oxford Nanopore technologies. Short read sequences from Illumina was mapped against long RNA sequences obtained through a MinION. Then the transcripts were annotated according to Gene Ontology terms utilizing Blast2Go program. RNA-Seq analyses were conducted to detect genes differentially expressed. Following this, relevant candidate genes, mainly upregulated in lice infected stage were selected and tested as recombinant vaccines. Challenge trials were conducted in triplicate, considering vaccinated and unvaccinated fish. Furthermore, transcriptome sequencing in lice and fish exposed to the vaccine prototypes were also analyzed. The results obtained from the experimental trials revealed that individuals of Atlantic salmon immunized with specific antigens displayed protection up to 90% at 25 days post-infestation, evidenced the reduction of adult juveniles, females, and males. Notably, the transcription analysis of key genes related to ontogeny and reproduction in sea lice were highly downregulated in response to the vaccines. Furthermore, the gene profiling of immune related-genes in vaccinated fishes revealed that relevant transcripts were modulated. In turn, the current experimental evidence suggests that at least one prototype could be assayed in the field or under commercial aquaculture conditions.
Funding: This study was funded by FONDAP (1510027) and FONDECYT (1181867) CONICYT-Chile.

W045: Aquaculture

Gene Transcription Data for eQTL Analysis, Variance Component Analysis and GEBV Estimation in Atlantic Salmon

Anna K Sonesson, Nofima AS, AAs, Norway, Hooman Moghadam, Salmobreed, Bergen, Norway, Matthew D Baranski, Marine Harvest, Aas, Norway, Siri Storteig Horn, Nofima, AAs, Norway and Theodorus M. Meuwissen, Norwegian University of Life Sciences, AAs, Norway

Gene transcription data is a highly informative source of data. Here, we are presenting examples in Atlantic salmon, where transcriptomic data is used for mapping of expression QTL (eQTL), variance component analysis and genomic selection.
Differentially expressed genes were used for eQTL mapping for omega-3 fatty acid synthesis genes in the liver of Atlantic salmon from the SalmoBreed population. The results were verified in a selective phenotyping test. Selection of SNPs was important for the results in this study. In another study, the most significant SNPs from a GWAS and/or a transcriptomic study were added to a GBLUP analysis for Pancreas Disease in a GEBV analysis of 8000 Atlantic salmon from the Marine Harvest population. The accuracy of selection using pedigree data only was 0.63.

When adding the genomic data, accuracy of selection was up to 0.924. The fish in this study were genotyped with a 55k SNP chip. Also, the 1316 most important SNPs derived from the transcriptomic data explained 37% of the genetic variance for PD resistance.
Allele specific expression of ~500 SNPs explained 24% of the genetic variation for the omega-3 fatty acid DHA content in the fillet of Atlantic salmon from the SalmoBreed population.
Overall, these studies show that a small number of highly informative SNPs derived from transcriptomic data provide important information for selective breeding.

W046: Aquaculture

Genomics of New Zealand Trevally: Exploring the Genetic Basis of Quantitative Traits to Inform a Newly Developed Breeding Program

Noemie Valenza-Troubat, School of Biological Sciences, Victoria University Wellington, Wellington, New Zealand

Most diversity in phenotypic traits is due to a combination of variation in alleles at multiple Quantitative Trait Loci (QTL), environmental effects and their interaction (Mackay, 2001). Understanding the complex network of genes underlying phenotypic variation and their external modulation has been a major and longstanding challenge in genetics (Fisher, 1930). In animal breeding, one of the main goals is to identify individuals that have high breeding values for traits of economic interest and use them to produce offspring homogeneous for these traits within short time frames (Dekkers, 2012). Modern genomics-informed breeding programmes can help accelerate these generational gains by focusing directly on the inherited components of traits and using parentage assignment to maximize family representation and control inbreeding (Vandeputte and Haffray, 2014).
Here we present the first efforts to identify key QTLs that influence growth-related traits in a new species being developed for aquaculture, the New Zealand trevally (Pseudocaranx georgianus). Using a broodstock and F1 offspring combined with genome-wide DNA information and extensive phenotype data, we estimate the heritability of target traits, determine parental contribution levels, measure inbreeding and genetic diversity, and identify targets for selective breeding. This work represents the first genotype-phenotype map for P. georgianus, allowing us to gain fundamental insights into the genetic architecture of traits and the potential for selective breeding of this species

1. Mackay TFC (2001) The genetic architecture of quantitative traits. Annual Review of Genetics, 35, 303–339.
2. Fisher R 1930. The genetical theory of natural selection, Clarendon.
3. Dekkers JCM (2012) Application of genomics tools to animal breeding. Current Genomics, 13, 207–212.
4. Vandeputte M., Haffray P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. Front. Genet. 5:432. 10.3389/fgene.2014.00432

W047: Aquaculture

The Road to Diploidy is Paved with Dangerous Duplicates: Conservation of Ohnologs through Partial Tetrasomy in Salmonidae

Matthew C. Hale, Texas Christian University, Fort Worth, TX, Krista M. Nichols, NOAA Fisheries, Seattle, WA, Matthew A. Campbell, Department of Ecology and Evolutionary Biology, Santa Cruz, CA, Garrett J McKinney, University of Washington, Seattle, WA and Devon E. Pearse, Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA

Whole-genome duplications (WGDs) have occurred repeatedly and broadly throughout the evolutionary history of eukaryotes. However, the effects of WGD on genome function and evolution remain unclear. The salmonid WGD that occurred approximately 88 million years ago represents an excellent opportunity for studying the effects of WGD as ~10-15% of salmonid genomes still exhibit partial tetrasomic inheritance. Herein, we utilized the rainbow trout (*Oncorhynchus mykiss*) genome assembly and brain transcriptome data to examine the fate of gene pairs (ohnologs) following the salmonid whole-genome duplication. We find evidence for higher sequence identity between ohnologs located within known tetrasomic regions than between ohnologs found in disomic regions, and that tetrasomically inherited ohnologs showed greater similarity in patterns of gene expression than disomically inherited ohnologs. In addition, we find evidence that one pair of homeologous chromosomes previously thought to be inherited tetrasomically - Omy01q and Omy23 - show evidence of sequence similarity more in line with disomic inherited regions of the genome, suggesting they have returned to a diploid state. We also used enrichment testing for Gene Ontology terms to determine the functionality of ohnologs in tetrasomic and disomic regions. We identified 49 over-represented terms in tetrasomically inherited ohnologs compared to disomic ohnologs. However, why these ohnologs are retained as tetrasomic is difficult to answer. It could be that we have identified salmonid specific “dangerous duplicates”, that is, genes that cannot take on new roles following WGD. Alternatively, there may be adaptive advantages for retaining genes as functional duplicates in tetrasomic regions, as presumably, movement of these genes into disomic regions would affect both their sequence identity and their gene expression patterns.

W048: Aquaculture

Varied Transcriptomic Responses to Dermo Disease within an Eastern Oyster Breeding Population

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

Dermo disease continues to limit aquaculture production of the eastern oyster, *Crassostrea virginica*. Disease resistance is a top priority for eastern oyster breeding programs, but simple methods to measure the trait directly are lacking, thereby reducing the efficiency of genetic improvement. Here we describe our efforts to better characterize Dermo resistance phenotypes and understand the genetic mechanisms underlying the response to this devastating disease. Oysters belonging to six distinct families within a selective breeding population were challenged with *Perkinsus marinus*, the parasite causing Dermo disease, according to optimized laboratory protocols. In addition to monitoring survival, five individuals · family^{-1} · treatment^{-1} were censored at 6hr, 7d, and 28d post exposure to assess changes in parasite load and gene expression over time. Gene expression profiles for three families exhibiting divergent resistance phenotypes were generated using RNAseq and differentially expressed genes at each time point were detected with DESeq2 and the Bayesian ApeGLM shrinkage estimator. As expected, the transcriptomic response to parasite exposure varied across families. The fewest differentially expressed genes (DEGs) were detected in the most susceptible family at early time points (34 and 24 at 36hr and 7d respectively); however, by 28d post-exposure, close to 897 genes were

differentially expressed (most upregulated) between control and exposed individuals. We observed a more pronounced transcriptomic response in the Dermo-tolerant family immediately after exposure, but by 28d, expression levels for all but six genes were similar between the control and exposed treatments. Surprisingly few DEGs were detected in the Dermo-resistant family at each time point. The functional annotation of DEGs detected among the divergent phenotypes will further elucidate mechanisms of Dermo resistance and advance our ability to directly measure the trait within a selective breeding framework.

W0**49: Aquaculture**

Identification of QTL Associated with Aeromonas Disease Resistance in Catfish through a Genome-Wide Association Study

Wenwen Wang, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

The catfish industry in the United States has been severely impacted by the emerging disease motile *Aeromonas* septicemia (MAS) caused by a highly virulent bacterium, *Aeromonas hydrophila*. Understanding genetic resistance should provide opportunities for genetic enhancement of disease resistance through breeding. In this study, we conducted a genome-wide association study (GWAS) to identify quantitative trait loci (QTL) for MAS resistance. A total of 1,820 interspecific backcross individuals of seven families were used in challenge experiment, and 382 phenotypic extremes were used for genotyping using the catfish 690 K SNP array. Three QTL on linkage group (LG) 2, 26 and 29 were found to be significantly associated with MAS resistance. Within these regions, a total of 24 genes were identified with known functions in immunity. Among them, ten genes were involved in NF-κB signaling pathway, including *trim25*, *cul1*, *dusp14*, *jak2*, *prlr-like*, *malt1*, *surf4*, *card9*, *casp3-like*, and *trl3*. This study suggested that NF-κB signaling may play important roles for MAS resistance, providing insights into the molecular mechanisms of disease resistance, and setting the foundation for the identification of causal genes of MAS disease resistance.

W0**50: Aquaculture**

Use of Atlantic Salmon Gene Editing in Research and Development

Anna Troedsson-Wargelius, Institute of Marine Research, Bergen, Norway

W0**51: Aquaculture**

Estimates of Numbers of Diagnostic Markers Required to Identify Introgressions in Diploid Cross-Species Hybrids from Different Types of Inter and Backcross Populations

Joseane Padilha Da Silva, Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil and **Alexandre R. Caetano**, Embrapa Recursos Genéticos e Biotecnologia, Brasilia, DF, Brazil

The use of artificial cross-species hybridizations to generate animals and plants better suited for draft and food production has been important in agriculture for centuries, and is still widely used with increasing importance especially in aquaculture. Adequate analytic tools for correct identification of cross-species hybrids from intercrossed and backcrossed populations, based on marker panels with adequate numbers of independent markers, are increasingly necessary for accurate species-purity certification and management of commercial broodstocks, in addition to monitoring of wild populations, especially when resulting hybrids are fully fertile. A statistical framework was developed based on power analysis to estimate minimal numbers of di-allelic markers with species-specific alleles required to reliably identify hybrids in advanced intercrossed and backcrossed populations. Simulated populations were used to test accuracy of proposed estimates. Estimated numbers of required markers ranged from 5 to 191 (α=0.05), and from 7 to 293 (α=0.01), considering backcross 1 (BC1) to BC6 populations, respectively. Numbers of markers required for proper hybrid identification observed in simulated BC1 to BC6 populations ranged from 5 to 1,131 and 7 to 8,065, considering error rates $\leq$5% and $\leq$1%, respectively. Obtained results indicate that cost-effective assay panels could be developed to provide practical tools for use in commercial and research settings, for proper hybrid identification following even up to four generations of backcrossing (BC4).

W0**52: Aquaculture**

Whole Genome Resequencing Refines Genomic Regions Associated with Bacterial Cold Water Disease Resistance in Rainbow Trout

Sixin Liu, Guangtu Gao, Gregory D. Wiens and Yniv Palti, USDA-ARS-NCCCWA, Kearneysville, WV

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 objective of this study was to refine the genomic regions associated with BCWD resistance using whole genome resequencing. Two filtering steps were used to select parents of phenotyped offspring from two generations of the Troutlodge May Spawning nucleus population for sequencing. First, the families in each generation were sorted by BCWD survival rates. The parents of the top 20 families were assigned to a resistant group (R), and the parents of the bottom 20 families were assigned to a susceptible group (S). Then, the haplotypes in the two QTL regions were reconstructed in each parent using the genotypes from our previous study. Ten R parents homozygous for the favorable Omy8 haplotype and ten S parents without the favorable Omy8 haplotype were selected for sequencing in the 2015 generation. Similarly, ten R parents homozygous for the favorable Omy25 haplotype and ten S parents without the favorable Omy25 haplotype were selected for sequencing in the 2017 generation. Over 15 million SNPs were identified from resequencing in this population and the two major QTL were narrowed down to regions much smaller than those reported previously. We are currently examining the reference genome sequence to identify candidate genes for BCWD resistance within the narrowed QTL regions.

W0**53: Aquaculture**

Development of a SNP Baseline for Genetic Stock Identification in a Commercially Important Species of the South-East Pacific (*Genypterus chilensis*)

Valentina Rocio Cordova Alarcon, Universidad de Chile, Depto. de Produccion Animal, Santiago, Chile

W0**54: Aquaculture**

Assembling the Arctic Charr, Chinook, Sockeye, Pink, and Chum Salmon Genomes and Identifying Orthologs Among Species

Kris A. Christensen[1], Eric B. Rondeau[1], David R. Minkley[1], Jong S. Leong[1], Dionne Sakhrani[2], Ben F. Koop[1] and Robert H. Devlin[3], (1)University of Victoria, Victoria, BC, Canada, (2)Fisheries and Oceans Canada, Centre for Aquaculture and Environmental Research, West Vancouver, BC, Canada, (3)Fisheries and Oceans Canada, West Vancouver Laboratory, West Vancouver, BC, Canada

An outline of the progress and results of the project to sequence and assemble several of the Pacific salmon and Arctic charr genomes will be presented. In addition, the preliminary results of attempts to identify orthologs between all salmonid species will be discussed. The presented methodology for identifying orthologs utilizes synteny and protein similarity to identify orthologs. This methodology was developed to mitigate the difficulties in identifying orthologous genes between salmonid species, which can be complicated by multiple gene copies and differing nomenclatures between species. Instead of individual researchers independently identifying orthologous genes repeatedly, our hope is that they would be able to look up all orthologs from a table that used well documented methodologies. With the identified orthologs between species, we also hope to identify genes that have unique gene expression profiles within a species and may have influenced species specific traits.

W055: **Aquaculture**

NRSP8 Bioinformatics

James M. Reecy, Iowa State University, Ames, IA

W056: **Arabidopsis Informatics**

New Developments at TAIR

Eva Huala and **Tanya Z. Berardini**, Phoenix Bioinformatics, Fremont, CA

The Arabidopsis Information Resource (TAIR) is a continuously updated, online database of genetic and molecular biology data for the model plant Arabidopsis thaliana that provides the global research community with centralized access to data for over 30,000 Arabidopsis genes as well as visualization and analysis tools for those data. We will share an update on the work done in the past year relevant to gene function and phenotype curation based on the experimental literature. We'll also touch on other new developments, like using BAR's API to pull in eFP Browser images into Locus pages, hosting a Community Resources Portal, and starting a blog. An update on a related Phoenix project, Phylogenes, will also be included.

W057: **Arabidopsis Informatics**

The Bio-Analytic Resource for Plant Biology - 2018 Updates

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

The Bio-Analytic Resource has collaborated with Cristobal Uauy and colleagues in Saskatoon, Canada, to develop a Wheat “eFP Browser” for visualizing expression patterns of wheat genes and their homoeologs in 71 tissues of wheat, sampled at different stages of development (Ramirez-Gonzalez *et al.*, doi: 10.1126/science.aar6089). These data are now available in the BAR’s ePlant framework at http://bar.utoronto.ca/~dev/, which has also been used to set up ePlants for data from the kilometer to nanometer scale across 14 other agronomically-important plant species, including *Cannabis sativa*. The original Arabidopsis ePlant published by Waese *et al.* last year has been updated to include 2.8M protein-DNA interactions and ~40k experimentally-determined protein-protein interactions from BioGRID in the Interaction Viewer module. We have collaborated with TAIR to enable eFP images from an RNA-seq expression atlas by Klepikova *et al.* (doi: 10.1111/tpj.13312) to be embedded directly in TAIR Gene pages. Last, we have created an updated Maize eFP Browser based on new and published RNA-seq data and V4 gene models in collaboration with Robin Buell and Genevieve Hoopes at Michigan State University (Hoopes *et al.*, Plant Journal – accepted).

W058: **Arabidopsis Informatics**

1001 Genomes

Arthur Korte, Center for Computational and Theoretical Biology, University Wuerzburg, Wuerzburg, Germany

Abstract here

W059: **Arabidopsis Informatics**

The New ABRC Database: Stocks and Distribution for the 21st Century

Jelena Brkljacic, Ohio State University, Columbus, OH

The Arabidopsis Biological Resource Center (ABRC) collects, preserves, reproduces and distributes *Arabidopsis thaliana* and related species resources to the research and education communities. ABRC is funded by the National Science Foundation (NSF), as well as by the community through stock and order charges. The physical collection of close to one million stocks is hosted at The Ohio State University (OSU), while the role of a corresponding searchable database has been contributed by The Arabidopsis Information Resource (TAIR) since 2001. Prompted by a change in TAIR funding model in 2013, the ABRC is developing a new ABRC-hosted independent database and ordering system, located at OSU. We are happy to announce that the launch of the new ABRC searchable database will be in early 2019. New search functions of this stand-alone database will greatly improve the accessibility of our collection by making our non-Arabidopsis and non-seed/non-DNA stocks (such as cell cultures, antibodies, vectors) easier to find. In addition to enabling ordering and secure payment functions, the system will provide superior speed and access to quality control data and other new features. New capabilities for searching, along with new navigation and ordering tools will be presented.

W060: **Arabidopsis Informatics**

CrY2H-Seq

Renee Garza, The Salk Institute for Biological Studies, La Jolla, CA

W061: **Arabidopsis Informatics**
Tuxnet: A Simple Interface to Process RNA Sequencing Data and Infer Gene Regulatory Networks
Rosangela Sozzani, North Carolina State University, Raleigh, NC

W062: **Arabidopsis Informatics**
Phenomator: An R Package for the Analysis of Large-Scale Phenomics Data
Jarkko Salojärvi, Nanyang Technological University, Singapore, Singapore
With the recent introduction of automated high-throughput phenomics platforms, plant phenotyping is enterering into a new era. The platforms make it possible to do accurate phenotyping of thousands of individuals, facilitating the mapping of QTLs and identifying the effects of gene knockouts on the phenotype. However, it is difficult to make all the heterogeneous data from different phenotyping variables commeasurable. We are developing an R toolbox aimed for standardizing the data and comparing the phenotypic variation against internal controls. The package can use the output of most commercial phenotyping platforms.

W063: **Arabidopsis Informatics**
PlantCV: Open-Source Image Analysis Software for Models and Crops
Malia Gehan, Donald Danforth Plant Sciences Center, St. Louis, MO, Noah Fahlgren, Donald Danforth Plant Science Center, Saint Louis, MO, Sanaz Jarolmasjed, Donald Danforth Plant Science Center, St. Louis, MO and Haley Schuhl, Donald Danforth Plant Science Ce, St. Louis, MO
To tackle the challenge of producing more food and fuel with fewer inputs a variety of strategies to improve and sustain crop yields will need to be explored. These strategies may include: mining natural variation of wild crop relatives to breed crops that require less water; increasing crop temperature tolerance to expand the geographical range in which they grow; and altering the architecture of crops so they can maintain productivity while being grown more densely. These research objectives can be achieved with a variety of methodologies, but they will require both high-throughput DNA sequencing and phenotyping technologies. A major bottleneck in plant science is the ability to efficiently and non-destructively quantify plant traits (phenotypes) through time. PlantCV (http://plantcv.danforthcenter.org/) is an open-source and open development suite of image processing and analysis tools that could initially analyze images from visible, near-infrared, and fluorescent cameras. Here we present new PlantCV analysis tools including ones associated with the development of a hyperspectral imaging platform aimed at the identification of early abiotic stress response.

W064: **Arabidopsis Informatics**
High-Throughput Phenotyping Combined with RNAseq to Investigate the Regulatory Network of SIG6 in Arabidopsis
Hussien F. Alameldin, Plant Research Laboratory, Department of Energy (PRL-DOE), East Lansing, MI and Beronda Montgomery, Biochemistry & Molecular Biology and Microbiology & Molecular Genetics Department of Energy (DOE) Plant Research Laboratory Michigan State University, East Lansing,, MI
Sigma factors are proteins that bind to the core RNA polymerase and initiate transcription in the chloroplast. Sigma factor 6 (SIG6) plays an essential role in regulating plastid development in Arabidopsis seedlings, including serving as an important factor in chlorophyll synthesis and/or accumulation in seedling development. Furthermore, expression of *SIG6*was previously reported to be reduced in phytochrome mutant lines *phyA*and *phyB*. Despite this genetic association and prior research on the developmental role of *SIG6*, there is limited knowledge about the specific regulatory role of SIG6 in specific light-signaling pathways, including associations with photosynthesis.
We tested the potential role of Sig6 in the photosynthesis pathway in adult Arabidopsis rosettes using high throughput phenotyping . Our results show a reduction of photosynthetic efficiency (measured as ΦII) of *sig6*mutants relative to WT under dynamic light conditions. This impairment was reversed when we expressed the *SIG6*in a *sig6*mutant background, suggesting an important role of Sig6 in plastid-related processes of adult plants.
Targeted gene expression and whole-genome transcriptome profiling (using RNA-seq) of the *sig6*mutant line, Sig6 complemented line, and wild-type (Col-0 WT) under continuous, sinusoidal, and fluctuating light are being analyzed to uncover the regulatory network of Sig6 under dynamic light conditions.

W065: **Arthropod Genomics and Genome Engineering**
***Trans*-Complementing Gene Drive System uncovers fine Workings of the CRISPR-Based Homing Process**
Victor Lopez Del Amo, Alena L. Bishop and **Valentino M. Gantz**, Cell and Developmental Biology, University of California, San Diego, La Jolla, CA
CRISPR-based gene drive constructs can bypass Mendelian inheritance by copying themselves onto the companion chromosome and are this way inherited with Super-Mendelian frequencies. This technology offers tremendous promise for public health as a mosquito or disease control tool, in agriculture for pest suppression and in conservation to remove invasive species that are harming the environment. Here we show that: 1) a CRISPR gene drive can be split into two separate functionally complementing transgenes that can be independently propagated; 2) this system can be used to analyze different variables in the laboratory phase optimizing the constructs before field applications; 3) fine workings of the copying process need to be considered when designing a gene drive construct to ensure efficiency.

W066: **Arthropod Genomics and Genome Engineering**
Progress Towards Developing Genetic Pest Management Strategies for Spotted Wing Drosophila (SWD)
Anna Buchman, UC San Diego, La Jolla, CA

Spotted Wing Drosophila (SWD; *Drosophila suzukii*) is a major invasive pest of many small fruits, and has caused significant worldwide economic losses. SWD control measures have largely relied on prophylactic application of broad-spectrum insecticides, which is problematic, as repeated use of insecticides is expensive, has a serious impact on beneficial arthropods, and makes it inevitable that resistance will arise in the foreseeable future. However, there are no effective alternatives to managing SWD infestation, and it is likely that this pest will continue to spread. Genetic pest management is a promising alternative approach that could complement existing control methods, and multiple types of genetic pest management systems can theoretically be engineered in SWD. Here, we describe the creation of the first-ever synthetic Medea gene drive element in SWD, and discuss the potential use of such an element for SWD population control. We also describe the development of functional genomic CRISPR-based tools in SWD, and discuss the utility of these tools for building CRISPR-based genetic pest management strategies for SWD.

W067: Arthropod Genomics and Genome Engineering
ABC Transporter Knockouts for Functional Genomics in the Bollworm, *Helicoverpa zea*
Omaththage P. Perera[1], Nathan Little[1], Calvin Pierce[1], Heba Abdelgaffar[2], Lin Niu[2] and Juan Luis Jurat-Fuentes[2], (1)USDA-ARS Southern Insect Management Research Unit, Stoneville, MS, (2)University of Tennessee, Department of Entomology and Plant Pathology, Knoxville, TN

W068: Arthropod Genomics and Genome Engineering
Making It Better: Improvement of a Mature Genome Assembly with Long-Read and Long-Range Sequencing Data
Brenda Oppert, USDA ARS Center for Grain & Animal Health Research, Manhattan, KS, Timothy P.L. Smith, USDA, ARS, USMARC, Clay Center, NE and Marce Lorenzen, North Carolina State University, Raleigh, NC
The sequence of the genome of the red flour beetle, *Tribolium castaneum*, was significant as it was the first sequenced genome of a coleopteran species, as well as that of a major agricultural pest. *T. castaneum* was sequenced more than ten years ago using Sanger sequencing of BAC and fosmid libraries of a highly inbred strain, GA2, and was assembled using Atlas suite. However, despite its wide utility in functional genomics and integration into genetic models, the assembly lacks approximately 35% of the total base pairs of the 205 Mbp genome, and the sequence of the Y chromosome remains elusive, likely due to highly repetitive sequences. We therefore improved the assembly with additional high coverage (140x) long-read PacBio sequence from GA2 and long-range information from Hi-C data (218M read pairs of 2x151 bp, 210,294x coverage). The mean insert length of the SMRTbell library was 8,731 bases and insert N_{50} of 14,750 bases; Sequel sequencing provided 3,764,395 subreads totaling 30.0 Gb, with an average length of 7,970 bases. The CANU assembly of PacBio data contained 864 contigs and 200,197,996 bp, with an N_{50} of 887,206 bp. A reference guided assembly of error-corrected PacBio reads to the current reference Tcas5.2 found 533,742 SNPs that may provide increased accuracy in the genome assembly. A Dovetail Hi-Rise assembly of Hi-C data to the CANU-assembled contigs resulted in 4 breaks and 360 joins and provided an N_{50} of 13.1 Mb and total length of 200 Mb. I will discuss how the data has been used to improve the genome assembly of an important pest and model of coleopteran genetics.

W069: Arthropod Genomics and Genome Engineering
Visualization of Insect Vector-Plant Pathogen Interactions in the Citrus Greening Pathosystem using Genomics, Transcriptomics and Proteomics
Surya Saha[1], Prashant S Hosmani[1], Mirella Flores-Gonzalez[1], Marina Mann[2], Stephanie Hoyt[2], Angela Kruse[2], John Ramsey[3], Wayne Hunter[4], Susan J. Brown[5], Michelle Cilia[2,3], Lukas A. Mueller[1] and International Psyllid Annotation Consortium, (1)Boyce Thompson Institute, Ithaca, NY, (2)Cornell University, Ithaca, NY, (3)USDA ARS, Ithaca, NY, (4)USDA Agricultural Research Service, Fort Pierce, FL, (5)Division of Biology - Kansas State University, Manhattan, KS
The Asian citrus psyllid (ACP,*Diaphorina citri*) is the insect vector of the bacterium *Candidatus* Liberibacter asiaticus (CLas), the causal agent for citrus greening disease, which threatens the citrus industry worldwide. The Asian citrus psyllid genome project is a coordinated effort to define the psyllid genome, including the identification and annotation of every psyllid gene. This discovery of psyllid genes regulating CLas acquisition and transmission by the psyllid will transform future vector management strategies for controlling citrus greening. Advances in psyllid genome sequencing to improve genome assembly, including using Pacbio and long-range Hi-C scaffolding, resulted in the identification of 13 psyllid chromosomes, the first description of chromosome number for this economically important hemipteran insect vector. Together with Pacbio IsoSeq technology to sequence psyllid transcripts from different life stages and those reared on CLas + and - trees, approximately 20,000 putative full-length protein coding psyllid genes were identified. Student driven annotation resulted in more than 500 high quality models of genes involved in CLas-ACP interactions. New assemblies and annotations of the Florida strains of the ACP bacterial endosymbionts, *Wolbachia*, *Profftella,* and *Carsonella* were also characterized from the genome sequencing data.
Finally, we developed a data visualization platform, the Psyllid Expression Network (PEN), which is a user-friendly web-based tool for mining gene and protein expression patterns. PEN enabled us to identify tissue and host plant specific changes in ACP genes in response to CLas at the transcript and proteome level. The availability of a high quality reference genome, endosymbiont genomes and tools for analyzing transcriptomics, proteomics and metabolomics data in an integrated, systems biology approach will enable novel approaches to control the transmission of citrus greening disease. The new ACP genome assembly (Diaci v3), PEN and other tools are available on https://citrusgreening.org/ which is our portal for all omics resources for the citrus greening disease.

W070: Arthropod Genomics and Genome Engineering
Evolutionary Genomics of Insect Microbe Interactions
Allison Hansen, Department of Entomology, University of California, Riverside, Riverside, CA

W071: Avian Genomics - Going Wild!
Genetic Landscape of a Haplotype-Resolved Female Zebra Finch Genome

Arang Rhie, NHGRI / NIH, Bethesda, MD

W072: **Avian Genomics - Going Wild!**
Chromosome-Level Genome Assembly and Reconstruction of Evolutionary Events in Birds and Other Dinosaurs
Rebecca O'Connor, University of Kent, Canterbury, United Kingdom

W073: **Avian Genomics - Going Wild!**
Conservation Genomics in the Sagebrush Sea: Population Differentiation, Historic Demography, and Local Metabolic Adaptation in Sage-Grouse
Kevin Oh, Colorado State University & USDA-APHIS Wildlife Services, FORT COLLINS, CO

W074: **Avian Genomics - Going Wild!**
Methodological Advances Since the First Avian Phylogenomics Project
Siavash Mirarab, University of California, San Diego, La Jolla, CA

W075: **Avian Genomics - Going Wild!**
Coalescent Analyses of Genome-Scale Indel (insertion-deletion) Data Provide a Unique Source of Information about Avian Species Tree and Ancestral Population Sizes
Edward L. Braun, University of Florida, Gainesville, FL

W076: **Avian Genomics - Going Wild!**
Immunome DB: A Resource for Avian Immunogenomics
Ralf Christian Mueller, Max Planck Institute for Ornithology, Radolfzell, Germany, Richard Kuo, Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, Lel Eory, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, Jacqueline Smith, Roslin Institute, Edinburgh, United Kingdom and Robert H. S. Kraus, University of Konstanz, Department of Biology, Konstanz, Germany
Over the last two decades, high-throughput sequencing technologies have paved the way for discovery-driven studies. Improved sequencing quality and longer reads with decreasing costs at the same time allow for not only large-scale but also high-quality sequencing in non-model organisms. With the introduction of third-generation sequencing, it became also possible to study highly repetitive regions on the genomic and alternative splice variants on the transcriptomic level. The immunome--by our informal definition the totality of all genes that are up- or down-regulated due to an immune response--is arguably comprised of the most complex regions in vertebrate genomes for in humans it is the biggest source of genetic variation. While it keeps getting easier to generate large-scale sequencing data, it is also getting more challenging to organise this data. Pieces of evidence for immune genes are kept in several publicly available databases like Ensembl, Genbank, Uniprot, Pfam, Gene Ontology, and candidate immune genes are frequently published in gene expression profile experiments. Thereby, several pieces of evidence for the same gene are often dissociated due to different identifiers in the corresponding databases. In order to prepare studies in the field of avian immunogenomics, it is therefore cumbersome to identify immune genes in the first place and then compile them into target gene sets. It is our goal to reconcile pieces of evidence for immune genes in an immunome database to facilitate studies in the field of avian immunogenomics. With a gene-centric view, each gene will get a unique identifier while also retaining its original identifier for traceability. We will integrate immune gene evidence from Ensembl, Genbank, Uniprot, Pfam, Gene Ontology, as well as from recently generated 363 bird genomes within the framework of the B10K consortium. Eventually, the immunome database can simplify the process of setting up immune gene sets for expression experiments, as well as for studies on genetic and functional variation, gene presence/absence across multiple species, and molecular evolution.

W077: **Banana Genomics**
Chromosome-Scale Assemblies of Wild *Musa* Genomes using Nanopore Long Reads and Optical Maps
Jean-Marc Aury, CEA - Genoscope, Evry, France
Plant genomes are often characterized by a high level of repetitiveness and polyploid nature. Consequently, creating genome assemblies for plant genomes is challenging. The introduction of short-read technologies 10 years ago substantially increased the number of available plant genomes. Generally, these assemblies are incomplete and fragmented, and only a few are at the chromosome scale. Recently, Pacific Biosciences and Oxford Nanopore sequencing technologies were commercialized that can sequence long DNA fragments (kilobases to megabase) and, using efficient algorithms, provide high-quality assemblies in terms of contiguity and completeness of repetitive regions. However, even though genome assemblies based on long reads exhibit high contig N50s (>1 Mb), these methods are still insufficient to decipher genome organization at the chromosome level. Here, we describe a strategy based on long reads (MinION or PromethION sequencers) and optical maps (Saphyr system) that can produce chromosome-level assemblies and demonstrate applicability by generating high-quality genome sequences of a wild banana, *Musa schizocarpa*.

W078: **Banana Genomics**
Oligo Painting FISH Facilitates the Analysis of Chromosome Structure in Banana (*Musa* ssp.)
Denisa Šimoníková, Jaroslav Doležel and **Eva Hřibová**, Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic
Genus *Musa* contains about 70 different species, which have traditionally been classified based on plant morphology and basic chromosome number into four sections: Eumusa (x=11), Rhodochlamys (x=11), Australimusa (x=10) and Callimusa (x=9, 10). The section Eumusa is the largest, comprises a majority of edible seedless banana cultivars and also their seed-bearing progenitors, and originated by intra- and inter-specific crosses between wild diploids *M. acuminata* (A genome) and *M. balbisiana* (B genome). A minor group of edible banana clones known

as Fe'i originated from species of the Australimusa section. In this work we took the advantage of the availability of genome sequence of *M. acuminata* 'DH Pahang' (2n=22) to design large sets of chromosome-specific oligomers suitable for fluorescence *in situ* hybridization (FISH). The oligomers were designed for each of the pseudomolecule / chromosome arm using the Chorus program and synthetized by Arbor Biosciences. Pools of oligomers were labeled using reverse transcription either by 5'-biotin or 5'-digoxigenin-specific primer and used for FISH on mitotic metaphase plates in accessions representing different subspecies of *M. acuminata* and in two representatives of each of *M. balbisiana*. Apart from wild seed-bearing diploids, chromosome structure was analyzed in intra- and inter-specific edible hybrid clones. The results obtained demonstrated that chromosome oligo painting is suitable for anchoring pseudomolecules to chromosomes, creation of molecular karyotypes and identification chromosomal rearrangements in *Musa*. This work opens avenues for comparative analysis of structural chromosome changes that accompanied the evolution of genus *Musa* and the origin of cultivated banana clones.
Work was supported by National Program of Sustainability I (LO1204); bioinformatics analyses were supported by the National Grid Infrastructure MetaCentrum (LM2010005).

W079: **Banana Genomics**

Gene-Edited Cavendish utilizing the Draft Genomes of Wild and Domesticated Bananas

James Dale[1], Benjamin Dugdale[2], Pradeep Deo[3], Peter Waterhouse[4], Hyungtaek Jung[5], Fatima Naim[5], Michal Lorenc[5], David Kudrna[6], Jianwei Zhang[7], Anthony Brinin[3], Jennifer Kleidon[3], Robert Harding[2] and Upendra Shekhawat[2], (1)Centre for Tropical Crops and Biocommodities, Brisbane, Australia, (2)Queensland University of Technology, Brisbane, Australia, (3)Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Australia, (4)Centre for Tropical Crops & Biocommodities, Queensland University of Technology, Brisbane, Australia, (5)QUT, Brisbane, Australia, (6)Arizona Genomics Institute, Tucson, AZ, (7)Huazhong Agricultural University, Wuhan, China

Gene editing now has the potential to precisely alter traits in bananas in such a way that in at least some jurisdictions the resultant lines would not be considered as genetically modified. The advantages of gene editing include the ability to make very defined and precise changes to genes or the expression of genes, a greatly reduced regulatory path to market and the lack of a "GM label". However, gene editing to develop a banana with an altered trait that has practical benefits which would not be considered GM are, in reality, challenging. The challenges include demonstrating that gene editing works in bananas, editing a banana cell and regenerating a plantlet in the absence of a selectable marker and identifying the gene/allele to be editing and finally what is the target edit.

We have already demonstrated that the gene editing technology, CRISPR/Cas-9, works efficiently in bananas. Using CRISPR, we knocked out the phytoene desaturase (PDS) gene in Cavendish bananas with many of the edits being tri-allelic. However, these edits were facilitated by stable Agrobacterium mediated transformation. For the "non-GM" version, we are editing Cavendish protoplasts using transient, non-integrating components and regenerating plantlets in the absence of selection. Our first practical target trait is resistance to Fusarium wilt tropical race 4 (TR4). We have previously demonstrated that over-expression of an NB-LRR gene, RGA2, derived from a TR4 resistant *Musa acuminata ssp malaccensis* accession, in Cavendish provides high level resistance in the field. A homolog of RGA2 is present in the Cavendish genome. We have subsequently sequenced and assembled haplotype-resolved genomes of both Cavendish (Grand Nain) and the *M. acuminata ssp malaccensis* accession using >80X coverage PacBio long read sequences together with polishing using Illumina short reads. These assembled genomes will provide the platform for further editing targets.

W080: **Banana Genomics**

Genome Editing of Plantain (AAB genome) to Inactivate the Integrated Endogenous *Banana Streak Virus* (eBSV) from Host Genome

Jaindra Tripathi[1], Valentine Ntui[1], Mily Ron[2], Kariuki Muiruri[1], Anne B. Britt[2] and **Leena Tripathi**[1], (1)International Institute of Tropical Agriculture (IITA), Nairobi, Kenya, (2)University of California, Davis, CA

Banana including plantain is important staple food crop cultivated across tropical and subtropical countries. Just like any other crop, the productivity of banana is constraint by the several diseases and pests. Genetic improvement of banana by conventional breeding is key for development of varieties with resistance to diseases and pests and higher yields. *Banana streak virus* (BSV) is one of the major challenges in banana breeding and dissemination of plantain hybrids due to presence of integrated endogenous *banana streak virus* (eBSV) sequences, which can be activated into the infectious episomal BSV. It is a plant pathogenic badnavirus of the family *Caulimoviridae*, affecting production of banana (*Musa* spp.). BSV is non-covalently closed, bacilliform double-stranded DNA (dsDNA) virus with monopartite genome of approximately 7.2 to 7.8 Kb encoding three open reading frames (ORFs).

Banana cultivars are polyploid clones derived from *Musa accuminata* (A genome) or/and *Musa balbisiana* (B genome). The eBSV sequences are integrated in the B genome of banana. Many economically important sub-groups of banana, such as plantains (AAB), which are an important staple food in Africa, contain at least one B genome. Under stress conditions like *in vitro* propagation, hybridization or/and unfavorable conditions like water and temperature stress, the eBSV produces functional episomal viral genome and infectious viral particles and as a result the plant develops disease symptoms. The presence of eBSV in the B genome of plantain (AAB) has become a major challenge for breeding and dissemination of hybrids. Therefore, we have developed a strategy to inactivate the eBSV in the host by editing the virus sequences using CRISPR/Cas9. The regenerated genome-edited events of Gonja Manjaya showed mutations in the targeted sites with the potential to prevent proper transcription or/and translational into functional viral proteins. Seventy-five percent (6/8) of the edited events remained asymptomatic in comparison to the non-edited control plants under water stress conditions, confirming inactivation of eBSV into infectious viral particles. This study paves the way for the improvement of B genome germplasm that can be used in breeding programs to produce hybrids that can be globally disseminated.

W081: **Banana Genomics**

Comparative Analysis of Whole Flower Transcriptomes in the Zingiberales

Ana Almeida, California State University East Bay, Hayward, CA

W082: **Banana Genomics**

Gene Tree Discordance, Phylogenetic Inference and Introgressions in *Musa acuminata* Subspecies

Mathieu Rouard, Bioversity International, Montpellier, France

Rouard M[1], Droc G[2,3], Martin G[2,3], Sardos J[1], Hueber Y[1], Guignon V[1], Cenci A[1], Geigle B[4], Hibbins MS[5], Yahiaoui N[2,3], Baurens FC[2,3], Berry V[6], Hahn MW[5], D'Hont A[2,3], Roux N[1]

1 Bioversity International, Parc Scientifique Agropolis II, Montpellier Cedex 5, France.

2 CIRAD, UMR AGAP, Montpellier, France.

3 AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

4 Computomics GmbH, Tuebingen, Germany.

5 Department of Biology and Department of Computer Science, Indiana University, Bloomington, Indiana.

6 LIRMM, CNRS - Univ. Montpellier 2, 161 rue Ada, Montpellier Cedex 5, France.

Edible bananas result from interspecific hybridization between Musa acuminata and Musa balbisiana, as well as among subspecies in M. acuminata. Four particular M. acuminata subspecies have been proposed as the main contributors of edible bananas, all of which radiated in a short period of time in southeastern Asia. Clarifying the evolution of these lineages at a whole-genome scale is therefore an important step toward understanding the domestication and diversification of this crop. This study reports the de novo genome assembly and gene annotation of a representative genotype from three different subspecies of M. acuminata. These data are combined with the previously published genome of the fourth subspecies to investigate phylogenetic relationships. Analyses of shared and unique gene families reveal that the four subspecies are quite homogenous, with a core genome representing at least 50% of all genes and very few M. acuminata species-specific gene families. Multiple alignments indicate high sequence identity between homologous single copy-genes, supporting the close relationships of these lineages. Interestingly, phylogenomic analyses demonstrate high levels of gene tree discordance, due to both incomplete lineage sorting and introgression. This pattern suggests rapid radiation within Musa acuminata subspecies that occurred after the divergence with M. balbisiana. Introgression between M. a. ssp. malaccensis and M. a. ssp. burmannica was detected across the genome, though multiple approaches to resolve the subspecies tree converged on the same topology. To support evolutionary and functional analyses, we introduce the PanMusa database, which enables researchers to exploration of individual gene families and trees.

W083: **BER Plant Genomic Science**

Overview and Joint Genome Institute Plant Program Update

Jeremy Schmutz[1], Kerrie W. Barry[1], David M. Goodstein[1], Jane Grimwood[2], Jerry Jenkins[2], Ronan O'Malley[1], Daniel Rokhsar[1] and John Vogel[1], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL

The Department of Energy Joint Genome Institute is funded to enable scientific advances that benefit the DOE research areas of bioenergy, global carbon cycling and biogeochemistry. These are accomplished through collaborative projects with JGI users through the Community Sequencing Program and the three funded DOE BioEnergy Research Centers. The Plant Program is part of the JGI 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 www.phytozome.net. Recently, the plant program has focused on projects that elucidate function of genes through comparative transcriptomics and directed experiments in our JGI Plant Flagship genomes. For these Plant Flagship genomes, we are continuing to improve the accuracy and completeness of the genome sequence and add data to update and improve the reference annotation. We continue to sequence de novo genomes as comparators to the Plant Flagships, have introduced new advances into these pipelines, and have expanded our efforts on projects that use diversity of natural or structured populations to identify and link genotypes to phenotypes for plant traits important in biofuel crops.

W084: **BER Plant Genomic Science**

Comparative Genomic and Transcriptomic Analyses for Bioprospecting in the Green Lineage using KBase

Crysten E. Blaby, Brookhaven National Laboratory, Yaphank, NY

The development of 'omics technologies has provided an unprecedented opportunity for the study of organisms as complex systems. With the genome-wide data these technologies provide, nearly any species can be fast-tracked to a level of understanding that was previously attainable for only a few "model organisms". The present challenge is shifting from acquiring genomic data, such as whole-genome sequences and transcriptomes, to using that knowledge to enable predictive biology and the rational redesign of biosystems. With the successful implementation of an integrated systems biology approach to model, design and engineer high levels of biofuel precursors and value-added products, microalgae have the potential to become major sources of sustainable bioenergy and bioproducts. Toward this goal, over 100 algal whole-genome sequences are either presently available or are soon to be published. As fundamental resources, these data combined with thousands of published transcriptomes are precipitating a paradigm shift in the way we understand one of the most diverse, complex and understudied groups of photosynthetic eukaryotes. Remarkably, over half of the proteins encoded by algal genomes are of unknown function, highlighting both the volume of unique functional capabilities yet to be discovered and a fundamental knowledge gap that impedes successful biosystem design. We are taking an integrative holistic approach to deciphering this functional capacity. Employing comparative genomics and transcriptomics tools in U.S. Department of Energy Systems Biology Knowledgebase (KBase), we are performing pathway bioprospecting, while elucidating the genomic-basis for conserved and unique processes that enable algae to adapt and acclimate to the environment.

W085: **BER Plant Genomic Science**

Evolution of Sex Chromosomes in the Moss Ceratodon

Stuart McDaniel, University of Florida, Gainesville, FL

The bryophytes are a diverse and abundant group of plants, with important roles in structuring ecological communities and driving global biogeochemical cycling. Until relatively recently, bryophyte genomics was represented by the moss Physcomitrella patens. Comparative analyses of P. patens and angiosperms highlighted the remarkable conservation of major developmental pathways among all land plants. Understanding

the basis of the important trait variation in land plants, however, requires additional genomic data. To provide additional genomic resources in the bryophytes, JGI sequenced two genomes of the moss Ceratodon purpureus. This species is common in exposed sites in temperate and polar regions throughout the world. C. purpureus has chromosomally determined males and females, and undergoes frequent sexual reproduction. We used a combination of Illumina, PacBio, and Dovetail HiC sequencing to generate chromosomal-scale genomes for a male and female of the species. We have additionally generated transcriptome and resequence data for additional isolates from across the distribution of the species. As a test case, we have used these tools to understand the evolution of sexual dimorphism in mosses. These tools, combined with ongoing efforts to refine existing gene-targeting tools, position C. purpureus as an emerging model system for understanding the genetic basis of ecologically important trait variation in land plants.

W086: BER Plant Genomic Science
Bioinformatic Approach to Discovering Promoters Regulating Poplar Drought Response
Austin R Wyer, University of Tennessee - Knoxville, Knoxville, TN

W087: BER Plant Genomic Science
Highlights from Genome of Miscanthus
Therese Mitros, UC Berkeley, Berkeley, CA

Miscanthus is a productive C4 grass capable of high biomass yields even in cool climates. The chromosome-scale assembly of the *Miscanthus sinensis* genome allows us to examine the relationships within the *Miscanthus* genus and analyze the relationship within the Andropogoneae tribe. We have analyzed both sinensis and sacchariflorus accessions and their relationship to the highly productive triploid *Miscanthus* x *giganteus*. The *Miscanthus* genus has a whole-genome duplication relative to the tropical grass *Sorghum bicolor* due to an allotetraploidy event that has resolved itself to disomic inheritance. It shares an evolutionary branch with sugarcane, a genome that has undergone multiple rounds of duplication relative to their common ancestor with sorghum. An analysis of the structural and regulatory changes among these genomes offer insights into the evolution of rhizome development, nutrient recycling, and self-incompatibility traits.

W088: BER Plant Genomic Science
Transcriptome Changes during Leaf Senescence in Populus
Haiwei Lu, Oregon State University, Corvallis, OR

W089: BER Plant Genomic Science
Natural Diversity in Setaria and a Novel Gene for Shattering
Elizabeth A. Kellogg, Donald Danforth Plant Science Center, St. Louis, MO

Setaria viridis is becoming a valuable model for C_4 panicoid grasses because of its small stature, rapid life cycle, ease of growth, and transformation efficiency. Here we add to those resources a platinum-quality assembled genome (*S. viridis* version 2.0) and a large, sequenced diversity panel. The genome was assembled using a combination of PacBio and short-read sequencing to produce 395.1 Mb of sequence with an N50 of 11.2 Mb. The sequence consisted of 75 contigs, with the vast majority (99.95%) of the assembled bases assigned to chromosomes. The diversity panel includes over 600 lines collected primarily from North America and sequenced to an average depth of 41.8x coverage. These assembled genomes had an average contig N50 of 16.2 Kb and 322.5 Mb of assembled bases. The diversity panel was assessed for presence-absence variation (PAV) and for single-nucleotide polymorphism (SNPs). Population genetic analyses of these two data sets separately arrived at similar results, finding three clear subpopulations of North American *S. viridis*, plus a fourth population that was heavily admixed. To demonstrate the power of the new genomic resources, we undertook a genome wide associate study GWAS to determine the genetic basis of seed abscission (shattering), a key trait in domestication of crops. We discovered a gene called Less Shattering 1 (LeS1) that has not previously been implicated in shattering. The gene was disrupted with CRISPR-Cas9 technology, confirming that LeS1 is indeed responsible for a reduced shattering phenotype.

W090: Beyond Drosophila: Genomic advances in non-model Diptera
Genomic and Genetic Strategies for understanding Chemosensory Behavior in the Mosquito *Aedes aegypti*
Ben Matthews, HHMI-Rockefeller University, New York, NY

Aedes aegypti mosquitoes are deadly vectors of arboviral pathogens including Zika, dengue, and yellow fever, and breed in containers of freshwater associated with human habitation. I will describe the generation of AaegL5, a dramatically improved reference genome assembly for *Ae. aegypti.* AaegL5 was generated and validated using a combination of diverse genomic technologies, including long-read PacBio sequencing and Hi-C scaffolding and de-deduplication. I will discuss how AaegL5 and its associated geneset annotation provides new insights into the biology of chemosensory-drive behaviors in mosquitoes. For example, when compared to previous geneset annotations, we doubled the known members of the ionotropic receptor (IR) gene family of transmembrane receptors that gate chemosensory behaviors like host-seeking, blood-feeding, and egg-laying.

Finally, I will discuss a genome engineering strategy based on CRISPR-Cas9 that allows for flexible genetic access to molecularly-defined cell types in the mosquito nervous system. Using loss-of-function mutagenesis as well as targeted integration of a transcriptional activator into specific genomic loci, I will describe how we identified a critical role for a DEG/ENaC ion channel gene, *ppk301*, in egg-laying behavior in *Ae. aegypti*. We found that *ppk301* is expressed in a defined population of sensory neurons in legs and proboscis, appendages that directly contact water during egg-laying, and that *ppk301*-expressing neurons project to central taste centers within the mosquito brain and ventral nerve cord. When *ppk301* mutant females contact water, they do not lay eggs as readily as wild-type animals and are more likely to make aberrant decisions between freshwater and saltwater at concentrations that impair offspring survival.

Together, the availability of new genomic resources and the advent of precise and efficient genome-engineering tools will facilitate understanding of the genes, neurons and neural circuits that enable mosquitoes to perform the behaviors that make them among the deadliest animals on earth. Using our work in *Ae. aegypti* as a roadmap, the development of similar tools and resources in other insect species of public

health, agricultural, or ethological interest will broaden our view of insect biology and facilitate comparative studies of the genes and circuits underlying evolutionary adaptations in insects. These behaviors contribute to the spread of deadly pathogens and understanding the underlying biology of these behaviors will contribute to control efforts, in addition to providing an invaluable comparative window into the organization and function of insect nervous systems.

W091: Beyond Drosophila: Genomic advances in non-model Diptera
Next-Generation Gene Drive in Mosquitoes: Approaches to the Development of Drive-Resistant Targets

Christian Ogaugwu, University of California Irvine, Irvine, CA

Gene drives are promising for effective and sustainable insect population control. These could be applied especially for malaria control through modification or suppression of mosquito populations. Our laboratory pioneered the first Cas9-mediated gene drive system for malaria control via population modification in the Indian malaria mosquito *Anopheles stephensi*. However, this system encountered resistance due to unwanted non-homologous end-joining (NHEJ) events in the mosquito's genome. We have recently developed a more efficient gene drive system for population modification in the African malaria mosquito *Anopheles gambiae*. This new system shows absence of resistance allele development, offering lessons for avoiding resistances and great hopes for control of malaria.

W092: Beyond Drosophila: Genomic advances in non-model Diptera
Genomic Approaches for Diagnostic Phylogenetics of Invasive Pests

Julian R. Dupuis[1], Sheina Sim[2], Daniel Rubinoff[3] and Scott Geib[2], (1)University of Hawaii, Hilo, HI, (2)USDA-ARS, Hilo, HI, (3)University of Hawaii, Honolulu, HI

Each year, thousands of exotic and invasive species are intercepted at ports of entry or detected in the environment. Rapid species identification is crucial to management and eradication of these species, and prevention of their establishment. Here, I will discuss recent developments in genomics-based species identification tools for tephritid fruit fly pests, a highly diverse family that includes some of the world's most economically-important agricultural pests, but lacks robust morphological tools for species discrimination, particularly at the immature level. Our approach combines: 1) multifarious uses of foundational genomic and transcriptomic resources, and 2) novel, cost-effective approaches for generating phylogenomic datasets for phylogenetic inference and species identification. For the former we combine genomic resources from traditional methods with those from cutting edge technologies, including de novo genome assemblies from long-read sequencing of single insects on the Oxford Nanopore MinION and PromethION platforms and linked-read sequencing with 10X genomics. For the latter, we leverage these diverse genomic resources in a bioinformatic locus selection pipeline that identifies conserved, phylogenetically-informative exons, and then used a novel single tube, highly-multiplexed amplicon sequencing approach to generate a phylogenomic dataset of ~900 genes for 384 specimens. In addition to providing unprecedented systematic resolution of Tephritidae, this approach is rapid (2-3 hour library preparation) and inexpensive ($10-20 per specimen), which are vital characteristics for a diagnostic tool. It also provides a massive resource for developing diagnostic markers for particular questions/needs; I will discuss our ongoing progress in transitioning these markers to other technologies, including real-time sequencing and analysis at ports of entry (e.g. Oxford Nanopore MinION sequencing fed into cloud-based analysis pipelines) to provide immediate diagnostic capabilities.

W093: Beyond Drosophila: Genomic advances in non-model Diptera
Development of Split-Gene Drive Systems in the Global Human Disease Vector, *Aedes aegypti*

Ming Li, University of California San Diego, La Jolla, CA

Aedes aegypti is a major vector of Dengue fever that infect nearly 390 million people annually, in addition it also transmits yellow fever, chikungunya, and Zika virus. Its preference for anthropomorphic habitats and ability to survive a prolonged desiccation at the egg stage contributed to *A. aegypti*'s range expansion far beyond its original distribution area. Current methods for *Ae. aegypti* control, *Wolbachia*-based Insect Incompatibility Technique (*W*IIT), the release of insects carrying a dominant lethal (RIDL) and female specific (fsRIDL) systems, are not adequate, risky, or prohibitively expensive. Here we describe a non-invasive and self-limited gene drive for *Ae. aegypti*. To achieve this we split the CRISPR-based gene drive into two separate elements – Gene Drive element (GDe) and Cas9 endonuclease (Cas9) – that are bred as separate lines and crossed together to generate the an active split-gene drive. Because each element segregates away from each other at every generation, the split design self-limit its propogation. A super-conserved region of *white* locus was targeted with a guide RNA (*gRNAw*) to establish a viable and easily scorable loss-of-function phenotype. We identified four active U6 polymerase III promoters and generated four transgenic *wGDe* lines each containing the same two genes, *gRNAw* expressed from one of four U6 promoters and a marker gene, integrated at *white* locus. We screened for the most efficient combination of GDe and Cas9 lines by scoring the cleavage of wild type *while* allele and transmission of *wGDe* in the progeny. Both *wU6b-GDe/w+*; *nup50-Cas9/+* females and males resulted in the highest super-Mendelian inheritance of *wU6b-GDe.* We found that out of 100% of *w+* alleles present in female germ cells of *wU6b-GDe/w+*; *nup50-Cas9/+*, 78.8% ± 8.6% were cut, 62.9% ± 15.8% were converted into *wU6b-GDe*, and 15.9% ± 9.4% were mutated and became resistant to further cleavage, *wR*. In trans-heterozygous male germ cells, these numbers were 48.1% ± 10.6%, 36.9% ± 10.8%, and 11.1% ±5.1%, respectively.

W094: Beyond Drosophila: Genomic advances in non-model Diptera
Developing Genomic Resources for *Contarinia nasturtii* (Diptera: Cecidomyiidae) to Aid in the Search for Resistant Crop Varieties.

Boyd A. Mori, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

W095: Beyond Drosophila: Genomic advances in non-model Diptera
Hi-C Assembly and Reassembly at the DNA Zoo: Inexpensive, Accurate and Open-Source

Olga Dudchenko, SB Batra, AD Omer, MS Shamim, NC Durand, B St Hilaire, Z Colaric, NT Musial, R Mostofa, M Pham, W Yao, E Stamenova, A Presser Aiden and Erez Lieberman Aiden, Baylor College of Medicine, Houston, TX

Hi-C contact maps are valuable for genome assembly (Lieberman-Aiden, van Berkum et al. 2009; Burton et al. 2013; Dudchenko et al. 2017). Recently, we developed 3D-DNA, an automated pipeline for using Hi-C data to assemble genomes (Dudchenko et al. 2017), Juicebox, a system for the visual exploration of Hi-C data (Durand, Robinson et al. 2016), and "Assembly Tools," a new module for Juicebox, which provides a point-and-click interface for using Hi-C heatmaps to identify and correct errors in genome assemblies (Dudchenko et al. 2018). The genome assembly procedures we describe are fast, inexpensive, accurate and open-source, and can be applied to many species. To illustrate, we assembled the mosquito disease vectors *Aedes aegypti* and *Culex quinquefasciatus* (Dudchenko et al. 2017; Matthews, Dudchenko, Kingan et al. 2018) as well as a number of other non-model organisms from many clades. We share the resulting chromosome-length genomics resources as part of the DNA Zoo Consortium effort at dnazoo.org.

W096: Big Data: Manage your data before your data kills you
BigPint: Big Multivariate Data Plotted Interactively

Lindsay Rutter, Iowa State University, Ames, IA and Dianne Cook, Monash University, Clayton VIC, Australia

Despite the availability of many ready-made testing software, reliable detection of differentially expressed genes in RNA-seq data is not a trivial task. While the data collection might be considered high-throughput, data analysis has intricacies that require careful human attention. Researchers should use modern data analysis techniques that incorporate visual feedback to verify the appropriateness of their models. While some popular RNA-seq packages provide static visualization tools, the types and capabilities of visualization tools should be expanded and their meaningfulness should be explicitly demonstrated to users.

In this talk, we 1) compile a collection of examples that demonstrate to biologists why visualization should be an integral component of RNA-seq analysis, 2) introduce new interactive RNA-seq visualization tools. We use public RNA-seq datasets to show that our new visualization tools can detect normalization issues, differential expression designation problems, and common analysis errors, sometimes in ways undetectable with models. We emphasize that interactive graphics should be an indispensable component of modern RNA-seq analysis, which is currently not the case.

Our R package "bigPint" includes the plotting tools introduced in this talk, many of which are unique additions to what is currently available. The "bigPint" website can be found online at https://rnaseqvisualization.github.io/bigPint. We include ten short vignette articles on our website that introduce users to our package. One article provides a recommended pipeline for users to iterate between models and visualizations when performing RNA-seq analysis. Our software incorporates data structures that allow users to transition smoothly between our plots and popular models from packages like edgeR, DESeq2, and limmaVoom. All articles are written using reproducible code that new users can follow.

This talk and its corresponding software aim to 1) persuade users to slightly modify their RNA-seq analyses by incorporating effective statistical graphics into their usual analysis pipelines, 2) persuade developers to create additional complex and interactive plotting methods for RNA-seq data, possibly using lessons learned from our open-source codes. It is our hope that our work will serve a small part in upgrading the RNA-seq analysis world into one that more wholistically extracts biological information using both models and visuals.

W097: Big Data: Manage your data before your data kills you
Nomenclature Can be Groovy!

Lisa Harper, USDA-ARS Corn Insects and Crop Genetics Research, Ames, IA and AgBioData Consortium

Wouldn't it be nice to have all published information about each gene available to researchers through databases? Wouldn't it be nice to agree on only ONE name for each gene so we could find information about that gene? In this genomics era, the naming of individual genes has gone off the rails. In the maize community, we have a nomenclature committee, and strict rules for gene naming. These were followed for decades, until about 10 years ago, when cloning genes became so much easier. Scientists seem to want to give any gene they touched a new name. This is a bioinformatics nightmare. In this talk, examples will be given to illustrate the issue, and the downstream effects of how hard this makes finding information on genes, and some ideas to help scientists NOT re-name genes. This talk will not touch upon the differences in gene nomenclature between different groups- that's a nightmare for another day. However, genome nomenclature is at a point where if we act fast, we can have species- independent guidelines for naming genome assemblies, annotation sets, etc., thus reducing confusion. Groovy!

W098: Big Data: Manage your data before your data kills you
Peanut Genomics Resources

Peggy Ozias-Akins, University of Georgia, Tifton, GA

Cultivated peanut (*Arachis hypogaea* L.) is a polyploid of recent origin with subgenomes of high sequence similarity derived from diploid progenitors *A. duranensis* and *A. ipaensis*. The crop and progenitor genomes only recently have been sequenced though an initiative largely supported by the peanut industry and involving multiple US and international collaborators. In addition to high quality genome assemblies, resequencing data have enabled SNP discovery and the development of tools (informatics and arrays) essential for diversity studies and genetic mapping leading to association of genetic markers with traits of interest. As with most crops, trait priorities for growers are yield, quality and disease resistance. With its narrow genetic base, cultivated peanut has relatively few strong disease and pest resistance alleles; therefore, a concerted effort now is being made to introgress novel alleles from wild species, taking advantage of the rich germplasm resources housed at the Plant Genetic Resources Conservation Unit in Griffin, GA. Genome, gene expression, germplasm and mapping data from intra- and interspecific lines are publicly accessible through PeanutBase and eFP Browser. PeanutBase and public repositories such as GenBank that house peanut sequence data have facilitated numerous studies advancing peanut science across the global peanut community.

W099: Big Data: Manage your data before your data kills you
Down in the Beans: Challenges and Solutions to Recovering and using Historic Soybean Yield Trial Data

Benjamin W. Campbell[1], Liana Nice[1], Ana M. Poets[2], Peng Zhou[3] and Aaron Lorenz[4], (1)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (2)University of Minnesota, Saint Paul, MN, (3)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (4)University of Minnesota, St. Paul, MN

The Uniform Soybean Tests (Northern Region) cooperative regional yield trials have been conducted each year by public universities in the Upper Midwest since the first published report in 1941. These yield trials improve breeders' selection efforts by providing a way for public institution breeders to test their new experimental lines in more diverse environments. This historical data presents a valuable resource for future breeding efforts as this data could be used for trait mapping and to train genomic selection models. The talk will cover challenges faced, solutions found, and results produced during efforts to recover and use data from the past 25 years of the cooperative yield trials (1993-2017).

W100: Big Data: Manage your data before your data kills you

Reproducible and Interpretable Data Management: Looking out for Future You (and Other People Too)

Sarah Odell, University of California, Davis, Davis, CA

Making scientific data available is becoming increasingly common, and is crucial in ensuring reproducible research. Releasing data alone, however, does not make open research. Sharing the methods used to collect and analyze data are just as important. This workshop will discuss do's and don'ts of data management and documentation, as well as various tools that are helpful with these processes. Being mindful about these processes while one is doing them is in the best interest of everyone, including future you!

W101: Big Data: Manage your data before your data kills you

Helping Biologists Make Sense of Plant Variant and Annotation Data

Oliver Schliebs[1], Armin Scheben[1], Brent Verpaalen[1], **Philipp E. Bayer**[1], Yuxuan Yuan[2], Ajit Singh[3], Keywan Hassani-Pak[3], Kenneth Chon-Kit Chan[4], Jacqueline Batley[1] and David Edwards[1], (1)University of Western Australia, Perth, Australia, (2)University of Western Australia, Perth, Western Australia, Australia, (3)Rothamsted Research, Harpenden, United Kingdom, (4)Australian Genome Research Facility, Melbourne, VIC, Australia

Advances in sequencing technology have led to a rapid rise in the genomic data available for plants, driving new insights into the evolution, domestication and improvement of crops. Sets of Single nucleotide polymorphisms (SNPs) and genome annotations are invaluable in research and breeding programs. Although there are many public repositories for sequencing data, which is routinely uploaded associated with publication of the related analysis, there are no formal repositories for crop SNP array data, and no repositories linking different annotation versions. To make SNP array data more easily accessible, we have developed CropSNPdb (http://snpdb.appliedbioinformatics.com.au), a database for SNP array data produced by the Illumina Infinium™ hexaploid bread wheat (*Triticum aestivum*) 90K and Brassica 60K arrays. We currently host data from SNP array datasets covering 526 *Brassica* lines and 309 bread wheat lines, and provide search, download and upload utilities for users. CropSNPdb provides a useful repository for these data which can be applied for a range of genomics and molecular crop breeding activities. CropSNPdb is available at http://snpdb.appliedbioinformatics.com.au.

In a separate development, we have developed a database to compare genome annotations, allowing the identification of gene homologues within and between genome assemblies. Daisychain is an interactive graph visualization and search tool for custom-built gene homology databases. The main goal of Daisychain is to allow researchers working with specific genes to identify homologs in other annotation releases. The gene-centric representation includes local gene neighborhood to distinguish orthologs and paralogs by local synteny. Daisychain is available at daisychain.appliedbioinformatics.com.au.

W102: Bioenergy Grass Genomics

Using Plant Associated Endophytes to strengthen Sustainable Energy Crops

Kerrie Farrar, Aberystwyth University, Aberystwyth, United Kingdom

Miscanthus and other perennial energy crops are being developed for renewable energy and bioproducts as substitutes for fossil fuels. They are largely undomesticated and must produce high annual biomass yields on low-quality land without inputs such as water, fertiliser or pesticides. Bacterial endophytes, which live in plant tissues, interact with aspects of plant growth and development and can play a role in host plant resilience to multiple abiotic and biotic stresses. We have isolated diverse bacterial endophytes from Miscanthus seed and mature plant tissues, as well as from plants growing under abiotic stresses such as salinity and heavy metal contamination. Genomic and functional analyses of the endophyte collections are underway to identify strains for sustainable agriculture applications and to address fundamental questions about plant-microbe relationships.

W1**03: Bioenergy Grass Genomics**

Think 2050! Making Miscanthus Meet the Demands of a Fast Growing Bio-Economy

William Cracroft-Eley, TERRAVESTA, Lincoln, United Kingdom

International commitments on emissions and climate change are all framed on a number of target dates, with the furthest being 2050
30 years is the "Blink of an Eye" in which to develop new feedstocks, develop and implement new infrastructure, and **have it all working at scale** in order to deliver the required outcomes
So, for a company like Terravesta, how can greater understanding of the Genome enable decision making shortcuts in the selection of breeding parents that can knock years of the trial & error of phenotypic trialling, and help us to deliver the solutions demanded today for the sustainable achievements of the worlds 2050 goals?

W1**04: Bioenergy Grass Genomics**

The Advantage of Perennial Miscanthus in Biogas Yield and Further Improvement on Biomass Composition

Jason W.H. Kam[1], David Thomas[1], Sandra Pierre[1], Jon McCalmont[2] and Sarah Purdy[3], (1)IBERS, Aberystwyth, United Kingdom, (2)College of Life and Environmental Science, Exeter, United Kingdom, (3)I.A. Watson Grains Research Institute, Narrabri, Australia

Anaerobic digestion (AD) currently relies heavily on biomass crop feedstocks such as *Zea mays* (maize) to maintain a constant output. Because of its high yielding and easy access on non-structural carbohydrate (NSC) for microbial conversion, maize became one of the major annual crop feedstock. This has raised significant environmental concerns such as high agricultural inputs, autumn harvesting exposing soils to heavy winter rainfall and a requirement for higher quality food-growing land for cultivation.
The UK the commercial hybrid *Miscanthus x giganteus* (Mxg) is physiologically similar to maize. Although, it can be used for AD it has a lower biogas yield caused by lower NSC content and biomass. On the positive side, based on previously calculated comparison between Miscanthus

and maize agricultural input, biogas produced with Mxg require far less (~50%) energy input per L of gas. Being typically harvested in spring made Miscanthus a better choice in terms of soil protection.
A novel Miscanthus hybrid was shown to have higher biogas yield than the commercial standard (Mxg). A follow up experiment has also determined the contribution of NSC to the improved biogas yield and demonstrated that breeding effort on improving NSC contents should be focused upon.

W105: **Bioenergy Grass Genomics**
Forward Genetics Identification of Genes Responsible for Differential Carbohydrate Metabolic Profiles in Miscanthus
Jose De Vega, Earlham Institute, Norwich, United Kingdom
Miscanthus is a candidate lignocellulosic biofuel crop owing to its high productivity and low chemical input requirements. Currently, M. x giganteus, a sterile interspecific hybrid of the two Miscanthus species, M. sacchariflorus and M. sinensis, is the only commercially available Miscanthus genotype. Breeding programmes aim to produce new interspecific hybrids that can out-perform M. x giganteus. In previous works, height has been shown to be the trait that best correlates with final yield in *Miscanthus,* and significant correlations were observed between carbohydrate metrics and biomass traits. In this work, we performed a RNA-seq transcriptomic analysis comparing "tall" and "short" interspecific hybrids and their progenitors in order to clarify the transcriptional differences behind their distinctive carbohydrate metabolic profiles. To facilitate the analysis, we also assembled a reference genome for the M. sacchariflorus progenitor. We identified a complex network of differentially expressed loci involved in starch metabolism and carbohydrate biosynthesis, and specifically up-regulated in the stem of some groups of Miscanthus hybrids. The molecular components that explain complex agronomic traits are difficult to define and largely unknown in crops, this omic data is allowing us to characterise the predominant functions in different tissues and genotypes, as well as the recent evolution of the Miscanthus genome.

W106: **Bioenergy Grass Genomics**
Development of Genomic Resources to De-Tangle the Complex Genome of Sugarcane
Adam Healey, HudsonAlpha Institute For Biotechnology, Huntsville, AL

W107: **Bioenergy Grass Genomics**
Sorghum: A Sustainable Feedstock for the Bioeconomy
Thomas E Clemente, University of Nebraska-Lincoln, Lincoln, NE
Terrestrial-based feedstocks can serve as an alternative source for petro-based industrial products. Sorghum (*Sorghum bicolor* (L.) Moench) possesses many qualities that can positively address the three prongs of sustainability, societal impact, environmental footprint, and economic viability. These attributes include C4 photosynthesis, production systems compatible with current infrastructure, high degree of genetic variation within the germplasm, and can be grown on a rather large footprint, with relatively low inputs. These phenotypic attributes are complemented by a wealth of genomics resources including a high-quality draft genome, plethora of genetic maps and cataloged transcript profiling databases. Moreover, reliable genetic transformation systems are available for the crop as a means to further expand genetic variation in the species through exploitation of various genetic tools including reagents for genome editing and capacity to build transgene stacks implementing DNA synthesis and assembly platforms. The challenge is to further enhance the attractiveness of sorghum as a feedstock for the bioeconomy through improvements in selected input and output traits. The former as a means to mitigate its environmental footprint and the latter as a way to boost the farmgate value of the harvest, thereby improving on the economic viability of sorghum as a biomass feedstock. To this end, a multidisciplinary approach, employing a build-test-learn method is being implemented targeting genetic strategies to introduce variation for two input traits, water use efficiency and light penetration through the canopy, by reducing the number of stomata and alteration of leaf angle, respectively, along with one output trait, synthesis of vegetative lipids.

W108: **Bioinformatics**
100 Genomes in 100 Days: The Structural Variant Landscape in Tomato Genomes
Michael C. Schatz[1], Sebastian Soyk[2], Xingang Wang[2], Michael Alonge[1], Zachary H. Lemmon[2], Srividya Ramakrishnan[3], Sara Goodwin[2], Fritz Sedlazeck[4], W. Richard McCombie[2], Esther van der Knaap[5], Joyce Van Eck[6] and Zach Lippman[2], (1)Johns Hopkins University, Baltimore, MD, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)Cold Spring Harbor Laboratory, NY, NY, (4)Baylor College of Medicine, Cold Spring Harbor, NY, NY, (5)University of Georgia, Institute of Plant Breeding, Genetics & Genomics, Athens, GA, (6)Boyce Thompson Institute for Plant Research, Ithaca, NY
Structural variations (SVs) are known to be major drivers of quantitative variation in many species. Short-read sequencing has proven valuable for SNP discovery, but lacks power for more complex SVs. Long-read sequencing has the potential to illuminate this large reservoir of hidden variation, but has previously been too costly and time-intensive to apply at scale. Addressing this critical gap, we have optimized the Oxford Nanopore PromethION to produce read lengths averaging over 30kb and yields up to 100Gbp per flowcell, making it fast and affordable to sequence large numbers of samples.
We have used this technology to sequence and understand SVs in nearly 100 varieties of tomato. Here we discuss 3 important questions: (1) How to select the samples for sequencing that will capture the largest amount of variation; (2) How to rapidly sequence, basecall, and manage the data for 100 genomes; and (3) How to identify SVs across the population using a combination of *de novo* assembly and read-mapping approaches. Each sample shows ten to forty thousand variants, including many within gene sequences that were not detectable using short-reads. Furthermore, for some variants we have developed CRISPR-based mutants showing novel fruit morphologies and inflorescence branching phenotypes. We believe this approach will become the new gold standard for SV analysis in all species.

W109: **Bioinformatics**
Falcon-Phase Integrates PacBio and HiC Data for *de novo* Assembly, Scaffolding and Phasing of Diploid Genomes

Sarah B Kingan[1], Zev N. Kronenberg[2], Gregory T Concepcion[1], Paul Peluso[1], Arang Rhie[3], Sergey Koren[3], Stefan Hiendleder[4], Olivier Fedrigo[5], Erich D. Jarvis[6], John Williams[4], Timothy P.L. Smith[7], Adam M. Phillippy[3], Richard Hall[1] and Shawn Sullivan[2], (1)Pacific Biosciences, Menlo Park, CA, (2)Phase Genomics, Seattle, WA, (3)NHGRI, NIH, Bethesda, MD, (4)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia, (5)Rockefeller University, New York, NY, (6)The Rockefeller University, Vertebrate Genome Laboratory, and HHMI, New York, NY, (7)USDA, ARS, USMARC, Clay Center, NE

Haplotype-resolved genomes are important to understand how combinations of variants impact phenotypes. Phase is commonly resolved using familial data, population-based imputation, or by isolating and sequencing single haplotypes using fosmids, BACs, or haploid tissues. Because these methods can be prohibitively expensive, or samples may not be available, alternative approaches are required. The *de novo* genome assembly with PacBio Single Molecule, Real-Time (SMRT) data produces highly contiguous, accurate assemblies. Two approaches have recently been described for phased diploid genome assembly. Trio binning is implemented in Canu v 1.8 and requires Illumina data from parents and PacBio data from the offspring. The long reads from the child are partitioned into maternal and paternal bins using parent-specific k-mers; the two PacBio read bins are then separately assembled, generating two fully phased haplotypes. An alternative approach (FALCON-Unzip) does not require parental information and separates PacBio reads during genome assembly using heterozygous SNPs. The length of haplotype phase blocks in FALCON-Unzip is limited by the level and distribution of heterozygosity, the length of sequence reads, and read coverage. To avoid fragmenting the assembly, FALCON-Unzip primary contigs may typically contain haplotype-switch errors between phase blocks. We developed FALCON-Phase, which integrates Hi-C data to resolve these switch errors. We applied our method to three vertebrate samples for which trio data are available for validation: an F1 bull, a zebrafinch, and a Puerto Rican female (HG00733). Phasing algorithm accuracy was 80% for human but >96% for non-human samples versus the trio approach, improving with higher heterozygosity. We also assessed the accuracy of the FALCON-Phase contigs by calling parental SNVs and found 90% accuracy in the zebrafinch and 96% accuracy in the bull. After phasing, the contigs were scaffolded and phased with the HiC data using the Proximo method to generate chromosome-scale phased scaffolds.

W1**10: Bioinformatics**
How Much Challenge Remains in the Task of Genome Annotation
Mark Borodovsky, Wallace H Coulter Department of Biomedical Engineering, Georgia Tech, Atlanta, GA

W1**11: Bioinformatics**
Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2
Sam Kovaka, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

W1**12: Bioinformatics**
Genome-Wide Function Annotation, Gene Families and Reference Datasets: Optimising Function Prediction for Comparative Genomics
Heiko Schoof, INRES Crop Bioinformatics, University of Bonn, Bonn, Germany

W1**13: Bioinformatics**
Generalization of the Minimizers Schemes
Guillaume Marcais, Insitutte for Physical Science and Technology, College Park, MD

W1**14: Bioinformatics for Large Scale Genotyping Data Management and Analytics Platforms and Future**
Providing Breeding Informatics Support to Breeding Teams in the Developing World
Gary Atlin, Bill & Melinda Gates Foundation, Seattle, WA

W1**15: Bioinformatics for Large Scale Genotyping Data Management and Analytics Platforms and Future**
GOBii, a Scalable Genomics Data Management System with Rapid Data Extract Times and Integration with Downstream Genomic Selection Analysis Pipelines
Yaw A. Nti-Addae[1], Victor Jun M. Ulat[2], Dave Matthews[3], Guilhem Sempéré[4], Valentin Guignon[5], Pierre Larmande[6], Jon Renner[7], Adrien Petel[8], Elizabeth Jones[1] and Kelly Robbins[3], (1)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (2)Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT), Texcoco, Edo. de Mexico, Mexico, (3)Cornell University, Ithaca, NY, (4)CIRAD, UMR Intertryp, Montpellier, France, (5)Bioversity International, 34397 Montpellier Cedex 5, France, (6)University of Montpellier,, Montpellier, France, (7)University of Minnesota, Minnesota, MN, (8)CIRAD, Montpellier, France

The Genomic Open-Source Breeding informatics initiative (GOBii) has built a genomics data management system that is highly scalable and has focused on data extract performance for large genomics data files. We have benchmarked several SQL and noSQL open-source data management systems with a view to managing large scale genomics data, and have determined that the HDF5 file system outperformed other data management systems both in loading and extract times. In order to also accommodate metadata management, we have designed and developed a hybrid system based on Postgres for sample and marker metadata management and HDF5 for the large genomics files. The HDF5 genomics files are stored in two different orientations to enable rapid extract in either sample or marker-fast formats. The system is flexible enough to be used across different crops and with diverse marker and sequence-based platforms. We are now working to integrate the genomics data extracts with downstream genomic selection applications in Galaxy.

W1**16: Bioinformatics for Large Scale Genotyping Data Management and Analytics Platforms and Future**
Managing and Exploring Large Genotyping Data with Gigwa

Guilhem Sempéré[1], Adrien Petel[2], Mathieu Rouard[3], Julien Frouin[4], Yann Hueber[3], Fabien De Bellis[4] and Pierre Larmande[5], (1)CIRAD, UMR Intertryp, Montpellier, France, (2)IRD, Montpellier, France, (3)Bioversity International, Montpellier, France, (4)CIRAD, Montpellier, France, (5)IRD - UMR DIADE, Montpellier, France

With the decreasing cost of genome sequencing, many laboratories are increasingly adopting genotyping technologies as a routine component of their analytical workflows, generating large datasets (e.g. VCF files) of genotyping information. Nevertheless, manipulating such large datasets remains a challenge for many scientists. In this context, we developed Gigwa (Genotype Investigator for Genome-Wide Analyses) with the aim of providing a user-friendly system to meet the requirements of scientists who need to filter large datasets and export them into various formats for subsequent analyses.

Gigwa is species-agnostic, cross-platform, scalable and easy to deploy. It can be configured to run on a local computer or setup across servers to act as a data portal. It may be used to share data with collaborators while providing means to seek variants of interest based on location, functional annotations or genotype patterns.

Based on NoSQL technology, it supports very large datasets (up to tens of millions of genotypes) when configured on suitable hardware. Its most attractive features are: ergonomic interface including user management, numerous import and export formats, powerful filtering engine, interoperability via REST APIs and connection to online or standalone tools.

Gigwa now in version 2 (http://gigwa.southgreen.fr), is developed within the scope of South Green bioinformatics, a cross-institute platform and community dedicated to genetics and genomics of tropical and Mediterranean plants, based in Montpellier, France.

W117: Bioinformatics for Large Scale Genotyping Data Management and Analytics Platforms and Future

SNP-Seek II: A Resource for Allele Mining and Analysis of Big Genomic Data

Dmytro Chebotarov, International Rice Research Institute, Los Baños, Philippines

W118: Bioinformatics for Large Scale Genotyping Data Management and Analytics Platforms and Future

The Practical Haplotype Graph: Cheap Genomic Selection using a Pan-Genome Database with Diverse Taxa

Sarah Jensen, Cornell University, Ithaca, NY

Decreasing sequencing costs have significantly increased the amount of genomic data available for a variety of crop species, but managing such large quantities of data is a continuing challenge. Sample preparation costs and bioinformatics requirements remain prohibitive for small breeding programs and have limited the integration of genomics in breeding decisions. We developed a Practical Haplotype Graph (PHG) in sorghum that provides both a database for sequence storage and a tool for genomic analyses. Using low-coverage (<0.1x) sequence data and the PHG we imputed genotypes for 2,880 individuals in a sorghum breeding program at the Regional Technical and Knowledge Center (CHIBAS) in Port au Prince, Haiti, which were subsequently used for genomic prediction. The PHG can also be used to evaluate diversity in the 300 biomass, grain, and wild sorghums currently in the database. Our results demonstrate the potential of the PHG as a research and breeding tool that maintains variant information from a diverse group of taxa, stores sequence data in a condensed but readily accessible format, and provides a cost-effective option for genomic selection.

W119: Bioinformatics for Large Scale Genotyping Data Management and Analytics Platforms and Future

ICRISAT Genebank Large Scale Genotyping: Increasing the Knowledge about the Diversity, Optimising Conservation Strategies and Promoting the use of the Germplasm

Vania C R Azevedo, INTERNATIONAL CROPS RESEARCH INSTITUTE FOR THE SEMI-ARID TROPICS, HYDERABAD, India

The ICRISAT genebank conserves over 126,000 germplasm accessions of six mandate crops (sorghum, pearl millet, finger millet, groundnut, chickpea and pigeonpea) and five small millets (foxtail millet, proso millet, little millet, barnyard millet and kodo millet) from 144 countries. Advances in sequencing technologies have made it possible to genotype large numbers of germplasm against low production costs, which has opened the door to screen genebank collections more efficiently for DNA sequence variation. Genomic data of genebank accessions helps to assess genetic structure of collections and to improve the composition thereof by eliminating redundancies and help in more informed decisions about acquisition. Also linking genotypic data to measured traits, allow germplasm repositories to be searched for materials containing desired genetic elements or traits characteristics for direct use in crop improvement program (link conservation with utilization). Two paralel pilot projects are being developed aiming genomic characterization of genetic resources conserved at ICRISAT's Genebank, initially for sorghum, pigeonpea and pearl millet wherein high quality genome sequence is available and the three crops collections conserved at ICRISAT genebank represents 78,647 accessions, 62% of the entire ICRISAT's collection. We are currently genotyping 12,000 samples (1360 accessions 8-25 plants/accession)– with DArT- seq platform that will be utilized to decide number of individuals to be genotyped per accessions, and within accession diversity will be estimated. Utilizing the insight and knowledge generated from these projects, we will propose to genotype the entire collection using appropriate number of plants/accession and the most appropriate genomic technology (for example: DT rhAmpSeq technology/Illumina Chip or other suitable platform) that brings maximum efficiency, high genomic resolution and relatively lower cost, with reproducibility and repeatability for remaining germplasm later on. For selected platform, a set of SNPs will be defined, aiming to prioritize validated and widely used international SNPs, and diagnostic SNPs of known alleles of interest and high information content SNPs for studies of diversity, kinship and varietal identification. For this we will utilize existing sequence data of germplasm diversity sets such as core/mini core collections. A representative set of germplasm will be also phenotyped for important stress(es) and nutritional traits and will be utilized along with genotyping data. This proposed strategy will lead identify/define threshold for defining redundancy/putative duplication of accessions and thus improve management efficiency. We will use this common set having genotype-phenotype information, and predict for the untested/unphenotyped individuals. Once this approach is tested in three crops initially with a large diverse set of germplasm, this could be extended to remaining collection of sorghum and other crops collections conserved in the ICRISAT genebank. Results will allow identification of duplication and further rationalization of the collection reducing operational costs; estimation of genetic diversity; identification of potential accessions with certain traits to be phenotyped and used in breeding programs. This study will also allow the development of thematic sets, the stratification of the collection accordingly molecular diversity in combination with important and useful traits.

W120: **Brachypodium Genomics**

Genome-Wide Transcriptome Analysis of BdCBF3-Dependent and -Independent Responses in *Brachypodium distachyon*

Shui-zhang Fei, Iowa State University, Ames, IA

C-repeat binding factors (CBFs) are transcription factors that play key regulatory roles in the cold acclimation process, which dramatically increases freezing tolerance in plants. We show that *Brachypodium distachyon,* a model for cool season grass can successfully cold acclimate. We also identified a *CBF* gene family consisting of eight genes in a tandem array and are designated as *BdCBF1-8*. Expression analysis indicated that all eight *BdCBF* genes are induced by cold and the knockdown of *BdCBF1* and *3* genes by RNAi results in a significant reduction in survival after an exposure to freezing temperatures. RNA-seq transcriptomic analysis was conducted using the wild type and RNAi *cbf3* mutant plants under both normal and cold conditions. We identified 460, 3,213, 2,839 and 1,871 differentially expressed (DE) genes from pairwise comparisons of *cbf3* (23 °C) vs. WT (23 °C), WT (23 °C) vs. WT (4 °C), *cbf3* (23 °C) vs. *cbf3* (4 °C), and *cbf3* (4 °C) vs. WT (4 °C), respectively. A comprehensive analysis of DE genes in some of the enriched pathways provide important insights into possible mechanisms of plant response to cold in the *BdCBF3*-dependent, -independent or -compensation categories. These DE genes serve as good candidates for further functional characterization in relation to cold acclimation and freezing tolerance and for marker-aided selection in improving cold hardiness for cool-season grasses including important forage, turfgrass and cereal crops.

W121: **Brachypodium Genomics**

Homoeolog-Specific Activation of Genes for Heat Acclimation in *Brachypodium hybridum*

Keiichi Mochida, RIKEN CSRS, Yokohama, Japan and Kotaro Takahagi, Kihara Institute for Biological Research, Yokohama City University, Yokohama, Japan

Allopolyploid plants often show better growth and stress tolerance compared to their ancestors, which are likely resulted from fixed heterozygosity of their homoeologous genomes. The complex homoeologous transcriptome of allopolyploid plants may have contributed to their evolutionally advantageous. To elucidate interaction of the homoeologous genomes in an allopolyploid grass, *Brachypodium hybridum* generated through natural interspecific hybridization between *B. distachyon* and *B. stacei*, we analyzed its genome and transcriptome and compared them to those from the ancestral species. Through our comparative transcriptome analysis with the ploidy series of Brachypodium, we found that the expression patterns of approximately 26% and approximately 38% of the homoeolog groups in *B. hybridum* represented nonadditive expression and nonancestral expression, respectively, under normal condition. We also found that *B. hybridum* and *B. stacei* show long-term heat stress tolerance, unlike B. distachyon. Comparing their transcriptome under heat stress condition, we identified that *B. distachyon* showed similar expression patterns between normal and heat stress conditions, whereas *B. hybridum* and B. stacei significantly altered their transcriptome in response to heat after 3 days of stress exposure, suggesting that homoeologs inherited from *B. stacei* contribute to the stress response and adaptation ability to heat in *B. hybridum*. After 15 days of heat exposure, we observed that *B. hybridum* and *B. stacei* maintained their transcriptome similar to those under normal conditions. These findings demonstrate that a homoeologous genome specific earlier response to heat contributes to maintain its physiological state under long-term heat stress in *B. hybridum*, and provide insights into homoeo-transcriptome related to adaptation to wider environments in alloplyploid plants

W122: **Brachypodium Genomics**

Impact of Cell Wall Carbohydrates on Plant Growth

Sang-Jin Kim, Great lakes Bioenergy Research Center/ Michigan State University, East Lansing, MI

Plants are important sources for our civilization that requires various products from plants, such as crops, feedstock, oils, proteins and complex sugars. The plant-derived products have also high potential to be used as renewable fuels, lubricants, textiles and building materials of many kinds. Biomass is important as an initial input that could determine yield; a high plant digestibility is also important for feedstock that contains high amount of proteins buried in complex matrix polysaccharides and essential for downstream application such as biofuel production.

To increase easily digestible hemicellulose (mixed-linkage glucan; MLG) in plant, we focused on understanding the factors required to accumulate large quantities of MLG including genetic regulation of MLG synthesis, identification of tissue specific promoters and characterization and engineering of MLG synthases. Mixed linkage (1,3;1,4)-β-glucan (MLG) is one of the major carbohydrates of cereal grains, and occupies up to 80% of cell walls of the Brachypodium endosperm. The MLG biosynthesis depends on the biochemical activity of membrane spanning glucan synthases encoded by the CSLH and CSLF cellulose synthase-like gene families. However, relatively little had been known about their topology with respect to the biosynthetic membranes until recently when it was shown that catalytic domain of CSLF6 resides in cytoplasm. Along with the identification of the topology of CSLF6, we have demonstrated that a functional YFP fusion of BdCSLF6 is localized to the Golgi apparatus and that the Golgi localization of BdCSLF6 is sufficient for MLG biosynthesis using live cell imaging and immuno-electron microscopy analyses of tobacco epidermal cells expressing BdCSLF6. When we have performed immuno-localization analyses with the MLG-specific antibody in Brachypodium and in barley, we found MLG present in the Golgi, post-Golgi structures and in the cell wall. Accordingly, analyses of a functional fluorescent protein fusion of CSLF6 stably expressed in Brachypodium demonstrated that the enzyme is localized in the Golgi. We established that overproduction of MLG causes developmental and growth defects in Brachypodium as also occur in barley. To overcome the growth defect by over-accumulation of MLG, we generate plants with improved biomass by expressing ER stress sensor IRE1 and crossed with MLG over-accumulating Brachypodium plant. CSLF6 is a multi-spanning membrane protein associated with the Golgi apparatus and MLG follows the secretory pathway for deposition in the cell wall. Therefore, MLG biosynthesis is likely controlled by the UPR. Therefore, we hypothesized that the growth penalty observed in the BdCSLF6 over-expressors (BdCSLF6OX) could be suppressed by overexpression of IRE1. We found that the lines transgenic for both IRE1OX and CSLF6OX lines maintained the increased biomass accumulation of the IRE1OX lines supporting that IRE1 overexpression suppresses the growth penalty induced by BdCSLF6 overexpression. We next tested the levels of MLG in IRE1OX/CSLF6OX and found enhanced that MLG levels compared to wild type, suggesting that increased availability of IRE1 results in maintaining an increase in biomass accumulation with concomitant increase in MLG levels. Therefore the IRE1OX is a positive trait for plant biomass productivity.

W123: **Brachypodium Genomics**

Wild Monocots are Wired Differently: Hormone Action in *Brachypodium*

Christian S. Hardtke, University of Lausanne, Lausanne, Switzerland

The evolutionary young age of flowering plants bolsters the assumption that equivalent genetic scenarios in different plant species trigger essentially congruent phenotypic consequences. Observations in the prime plant model system Arabidopsis, a dicotyledon, therefore set expectations even for distantly related species, such as major monocotyledon staple crops. To test such expectations in a direct translational forward in crops is neither practical nor universal, and limited by genetic bottlenecks due to domestication. *Brachypodium distachyon* (Brachypodium) is a tractable wild monocotyledon whose experimental investigation is not constrained by scale. Brachypodium is thus suitable to accelerate discoveries and translational research in its close crop relatives, notably barley, rye and wheat. Here I present examples of Brachypodium mutants in the auxin and ethylene phytohormone pathways that display counterintuitive phenotypes based on their Arabidopsis precedents. These discrepancies can be explained by alternative wiring of a regulatory node in the crosstalk between the two pathways. Moreover, phenotypic variation between alleles and genetic backgrounds emphasizes the differential impact of transcriptional fine tuning. Our results highlight the advantage of trait evaluation in genetically less complex, wild relatives of crops, and contribute to the evolutionary framework of monocotyledon development.

W1**24: Brachypodium Genomics**

Revisiting *Brachypodium* Genomes through Whole-Genome Optical Maps

Tingting Zhu, Department of Plant Sciences, University of California, Davis, Davis, CA

Brachypodium distachyon (n = 5) is a diploid and has been widely used as a genetic model. *Brachypodium stacei* (n = 10) and *B. hybridum* (n = 15) are species that are related to *B. distachyon*, leading to an hypothesis that they are part of a polyploid series based on x = 5. Several lines of evidence suggest that this hypothesis is incorrect and that the genomes of the three taxa may have evolved by a more complex process. We constructed an whole-genome optical map for each species and did pairwise alignment of these maps. The alignment results showed that *B. distachyon* and *B. stacei* are both diploid, in spite of *B. stacei* having twice as many chromosomes as *B. distachyon*, and that *B. hybridum* is an allopolyploid formed from hybridization between *B. distachyon* and *B. stacei*. This study also demonstrated the use of optical maps in the detection and quantification of structural variants among the genomes.

W1**25: Brassicas**

Exploiting Long Read Sequence Technology to Resolve the Hidden Genomic Landscape of *Brassica* Species

Isobel Parkin[1], Sampath Perumal[1], Chu Shin Koh[2], Lingling Jin[3], Erin Higgins[1], Trisha Bender[1], Stephen J. Robinson[1], Miles Buchwaldt[1] and Andrew G. Sharpe[2], (1)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (2)Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)Thompson Rivers University, Kamloops, BC, 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 these often elusive genome differences. In order to study a large number of lines the technology needs to be both cost effective and preferably accessible to many labs. To test its 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 resulting assembly was compared to an assembly generated using an Illumina short-read sequencing based approach. The ONT assembly extended the original reference assembly by 91 Mb, covering ~94% of the expected genome size. The majority (85%) of the additional assembled sequence represented repetitive DNA, yet ~6,000 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. This technology is advancing rapidly and offers many opportunities for accessing some of the important structural variation in Brassica species; we are currently testing its efficacy for tackling the further complexity of the *Brassica napus* genome, a preliminary assembly for the allopolyloid will be discussed.

W1**26: Brassicas**

Long Reads Reveal Small Scale Genome Structural Variations in *Brassica napus*

Harmeet Singh Chawla, Suriya Tamilselvan-Nattar-Amutha, Andreas Welke, Subhadra Chakrabarty and Rod Snowdon, Department of Plant Breeding, Justus Liebig University, Giessen, Germany

There is increasing evidence that genome structural variation (SV) contributes significantly to phenotypic variation for many important agronomic traits. Most crop plants derive from ancient or recent polyploidisation events and carry extensive SV, however difficulties in describing this kind of variation with common genetic marker systems or short-read DNA sequencing mean that SV has been largely neglected when it comes to explaining observable phenotypes. Accurate long-read sequencing provides new opportunities to detect SV at the scale of single-genes and associate such variants to traits. Here we describe the use of Oxford Nanopore sequencing for precise detection of small scale SV and association with quantitative disease resistance and various other agronomically interesting traits in *Brassica napus* (canola, rapeseed), a recent allopolyploid species with a complex, rearranged genome which is today the world's second most important oilseed crop. Whole genome sequencing was performed for a commercial cultivar along with two synthetic *B. napus* lines carrying interesting trait variation for breeding, in order to identify the distribution and frequency of large-scale and small-scale SVs and investigate associations with known QTL in crosses between these lines. We established efficient and fast protocols for high molecular weight DNA isolation and library construction that enabled high data yields with optimal sequence quality on the Nanopore MinION system. Although DNA quality in *Brassica* species is often poor due to high levels of secondary compounds in the leaves, our protocols enabled consistent yields of 6-8 GBp per Nanopore flowcell, corresponding to around 5x – 6x coverage of the *B. napus* genome (1.3 Gb). Using this data we were able to identify large-scale and small-scale SVs (down to single-gene level) associated with quantitative disease resistance and other traits. Reduced costs and improved data quality from long-read

sequencing technologies provide new opportunities for analysis and gene discovery in polyploid crops where reference-based sequence mapping approaches fail to capture important SV.

W1**27: Brassicas**

Brassica Pangenomes as a Novel Source of Disease Resistance Genes

Armin Scheben[1], Philipp E. Bayer[1], Aria Dolatabadian[1], Agnieszka Golicz[2], Bhavna Hurgobin[1], Soodeh Tirnaz[1], Kenneth Chon-Kit Chan[3], David Edwards[1] and Jacqueline Batley[1], (1)University of Western Australia, Perth, Australia, (2)The University of Melbourne, Melbourne, Australia, (3)Australian Genome Research Facility, Melbourne, VIC, Australia

Brassicas are important crops used as oilseed, condiments, and vegetables. Every year, fungal and bacterial pathogens of Brassicas cause diseases that lead to substantial yield loss. Resistance to these diseases is mediated by resistance (R) genes that enable plants to recognise pathogens and activate inducible defences. The identification and genomic analysis of R genes often relies on reference genomes and annotations. However, several recently published Brassica genome annotations show strong differences in R gene content. We demonstrate that part of these differences in annotations is explained by repeat masking. As reference genomes also rarely capture the full genetic variation within a species, pangenome studies are useful for shedding light on R gene diversity. By analysing the pangenomes of *B. napus* and *B. oleracea,* we uncover high levels of presence/absence variation in R genes. Based on known QTL and markers associated with resistance to fungal and bacterial diseases, resistance gene analogs could be linked to these resistance traits. The R gene diversity found using carefully annotated reference genomes and pangenomes can facilitate breeding for disease resistance to increase yields.

W1**28: Brassicas**

Dissecting Vernalisation: Keeping the Greens on Your Plate

Judith Irwin, John Innes Centre, Norwich, United Kingdom

W1**29: Brassicas**

***Brassica oleracea*: The Dog of the Plant World**

Makenzie E. Mabry, Division of Biological Sciences, University of Missouri, Columbia, MO

The horticultural crop *Brassica oleracea* L. plays an important role in global food systems. *Brassica oleracea* is unique in that it has been domesticated into several morphotypes (cultivars), including broccoli, Brussels sprout, cabbage, cauliflower, kale, kohlrabi, and several lesser well known morphotypes, such as walking stick kale and marrow cabbage. These crops are widely used as leaf and root vegetables, as well as for animal feed. There are several hypotheses on the origin of these crops. However, cultivation likely originated in the Mediterranean region with additional domestications occurring around the world. One uniting characteristic of these vegetable crops is the presence of glucosinolates, bitter tasting compounds that are useful for their herbivory defense, and potentially have anti-carcinogenic properties. Using this system of diversity within *Brassica oleracea*, we aim to examine patterns of relationships among morphotypes and wild relatives, including signals of hybridization and introgression. We also plan to elucidate the wild progenitor of *B. oleracea* to determine its origin of domestication. Lastly, using association mapping techniques, we hope to possibly identify genes underlying quantitative phenotypic traits of economic importance.

W1**30: Brassicas**

Using *Brassica rapa* to understand Epigenetic Dynamics in the Seed

Rebecca A. Mosher, University of Arizona, Tucson, AZ

Mutants in the RNA-directed DNA Methylation (RdDM) have different phenotypes in *Arabidopsis thaliana* and *Brassica rapa*, despite the close evolutionary relationship of these species. Specifically, *B. rapa* RdDM mutants display specific and severe defects in seed development, manifest as high rates of abortion after fertilization. This phenotype is controlled by maternal sporophytic genotype, rather than the genotype of the filial tissues (embryo and endosperm), which suggests that small interfering RNAs from the maternal soma might perform an essential function in *B. rapa* filial tissues. We are currently investigating the epigenomic patterns of the three genetically-distinct seed compartments and any interactions between them.

W1**31: Buffalo Genomics**

Chromosome-Level Assembly of the Water Buffalo Genome Surpasses Human and Goat Genomes in Sequence Contiguity

Lloyd Low[1], Rick Tearle[1], Derek M. Bickhart[2], Benjamin D. Rosen[3], Sarah B Kingan[4], Thomas Swale[5], Francoise Thibaud-Nissen[6], Terence D. Murphy[6], Rachel Young[7], Lucas Lefevre[8], David A. Hume[7], Andrew Collins[9], Paolo Ajmone Marsan[10], Timothy P.L. Smith[11] and John Williams[1], (1)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia, (2)Dairy Forage Research Center, USDA-ARS, Madison, WI, (3)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (4)Pacific Biosciences, Menlo Park, CA, (5)Dovetail Genomics, LLC, Scotts Valley, CA, (6)National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, (7)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (8)The Roslin Institute and Royal (Dick) School of Veterinary Studies, Edinburgh, United Kingdom, (9)University of Southampton, Southampton, United Kingdom, (10)Università Cattolica, Piacenza, Italy, (11)USDA, ARS, USMARC, Clay Center, NE

Rapid innovation in sequencing technologies and improvement in assembly algorithms have enabled the creation of highly contiguous mammalian genomes. Here we report a chromosome-level assembly of the water buffalo (*Bubalus bubalis*) genome using single-molecule sequencing and chromatin conformation capture data. PacBio Sequel reads, with a mean length of 11.5 kb, helped to resolve repetitive elements and generate sequence contiguity. All five *B. bubalis*sub-metacentric chromosomes were correctly scaffolded with centromeres spanned. Although the index animal was partly inbred, 58% of the genome was haplotype-phased by FALCON-Unzip. This new reference genome improves the contig N50 of the previous short-read based buffalo assembly more than a thousand-fold and contains only 383 gaps. It surpasses

the human and goat references in sequence contiguity and facilitates the annotation of hard to assemble gene clusters such as the Major Histocompatibility Complex (MHC).

W132: Buffalo Genomics

NCBI Annotation of the Water Buffalo Genome

Francoise Thibaud-Nissen, Terence D. Murphy and the Eukaryotic Genome Annotation Team, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

A reference assembly and a quality gene annotation are critical for understanding the diversity within a species and for harnessing this diversity toward the genetic improvement of crop and livestock. In the past year, a substantially improved water buffalo (*Bubalus bubalis*) assembly, UOA_WB_1 (GCF_003121395.1), was made available, replacing the fragmented assembly, UMD_CASPUR_WB_2.0 (GCF_000471725.1). The National Center for Biotechnology Information (NCBI) produces and maintains consistent gene annotation in RefSeq for over 500 eukaryotes of medical, economical or agricultural relevance. These annotations are the product of an automated pipeline, and are based on experimental evidence, such as ESTs, proteins, and RNA-Seq. We have annotated UOA_WB_1 and will demonstrate how the higher contiguity of the assembly puts the annotation, NCBI *Bubalus bubalis* Annotation Release (AR) 101, on par with cattle AR 105 on assembly ARS-UCD1.2 and goat AR 102 on assembly ARS1. Gene models predicted on the new assembly are more complete, are better supported by experimental evidence and have higher-coverage hits to UniProtKB/SwissProt proteins than models annotated on the old assembly. In addition, the 15 billion RNA-Seq reads used in the annotation of UOA_WB_1 compared to 1.6 billion for UMD_CASPUR_WB_2.0 provide a richer annotation by increasing the number of genes with alternative splice variants by 40%.

NCBI *Bubalus bubalis* AR 101 is available for download and is in NCBI's Gene resource, BLAST databases, and Genome Data Viewer (GDV). Further information about NCBI's annotation resources and GDV is available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/ and https://www.ncbi.nlm.nih.gov/genome/gdv/.

This research was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health

W133: Buffalo Genomics

Distribution of Runs of Homozygosity in River and Swamp Buffalo

Nicolo P. P. Macciotta[1], Alberto Cesarani[1], Lloyd Low[2], Rick Tearle[2] and John Williams[2], (1)University of Sassari, Sassari, Italy, (2)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia

Runs of homozygosity are stretches of DNA that harbor sequences of consecutive homozygous genotypes. They are considered indicators of inbreeding events but also of signatures of selection. In the present work the ROH distribution of a sample of Riverine (n=185) and Swamp (n=153) Buffaloes is investigated. Animals were genotyped with the 90K Affymetrix Axiom Buffalo genotyping array. The SNP editing was based on call rate (>0.95) , significant deviation from the Hardy Weimberg equilibrium (P<0.0001) and Minor Allele frequency (0.01). After edits the retained SNPs were 22,368. A total of 31,084 different ROHs were detected, many of which (70%) in Swamp buffaloes. The most represented length class of ROH was 2-4Mb, about 50% and 42% in swamp and river respectively. The mean number of ROH per animal was 57 (sd 34) and 219 (47) in River and Swamp, respectively. The most frequent ROH was detected on chromosome 2, it was about 2.2 Mb long and it occurred in 39 SWAMP buffaloes of different breeds. Usually the most frequent ROHs were shared by SWAMP buffaloes of different breeds, but of interest is a ROH located on chromosome 3 that was shared by 19 SWAMP and 1 RIVER buffalo. The largest average number of ROH per animal was found in the Nusa Tenggar Indonesian (288) and in the Italian (113) breeds for the SWAMP and RIVER species respectively. SNPs the had the larger occurrence in different ROHs were located on chromosome 2.

W134: Buffalo Genomics

Computational Detection and Experimental Validation of Segmental Duplications and Associated Copy Number Variations in Water Buffalo (*Bubalus bubalis*)

Shuli Liu[1], Xiaolong Kang[2], Claudia R. Catacchio[3], Mei Liu[1], Lingzhao Fang[4], Steven G. Schroeder[1], Wenli Li[5], Benjamin D. Rosen[1], Daniela Lamartino[6], Leopoldo Iannuzzi[7], Tad S. Sonstegard[8], Curtis P. VanTassell[1], Mario Ventura[3], Lloyd Low[9], John Williams[9], Derek M. Bickhart[10] and George E. Liu[1], (1)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (2)USDA-ARS, Beltsville, MD, (3)Department of Biology, University of Bari, Bari, Italy, (4)ARS-USDA, Beltsville, MD, (5)U.S. Dairy Forage Research Center, Agricultural Research Service, USDA, Madison, WI, (6)Parco Tecnologico Padano, Via Einstein, Lodi 26900, Italy, Lodi, Italy, (7)Nationa Research Council (CNR), ISPAAM, Laboratory of Animal Cytogenetics and Gene Mapping, Naples, Italy, (8)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN, (9)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia, (10)Dairy Forage Research Center, USDA-ARS, Madison, WI

Background

Duplicated sequences are an important source of gene evolution and structural variation within mammalian genomes. Using a read depth approach based on next-generation sequencing, we performed a genome-wide analysis of segmental duplications (SDs) and associated copy number variations (CNVs) in the water buffalo (*Bubalus bubalis*).

Results

By aligning short reads of Olimpia (the reference water buffalo) to the UMD3.1 cattle genome, we identified 1,038 segmental duplications comprising 44.6 Mb (equivalent to ~1.73% of the cattle genome) of the autosomal and X chromosomal sequence in the buffalo genome. We experimentally validated 70.3% (71/101) of these duplications using fluorescent *in situ* hybridization. We also detected a total of 1344 CNV regions across 14 additional water buffalo individuals, amounting to 59.8 Mb of variable sequence or the equivalent of 2.2% of the cattle genome. The CNV regions overlap with 1245 genes and are significantly enriched for specific biological functions including immune response, oxygen transport, sensory system and signal transduction. Additionally, we performed array Comparative Genomic Hybridization (aCGH) experiments using the 14 water buffaloes as test samples and Olimpia as the reference. Using a linear regression model, a high Pearson correlation ($r = 0.781$) was observed between the $\log_2$ ratios between copy number estimates and the $\log_2$ ratios of aCGH probes. We further

designed Quantitative PCR assays to confirm CNV regions within or near annotated genes and found 74.2% agreement with our CNV predictions.

Conclusions

Our study provides a highly valuable resource of variable genome for future evolutionary and phenotypic studies in water buffalo. These results confirm sub-chromosome-scale structural rearrangements present in the cattle and water buffalo that may be useful in future selective breeding of both species.

W135: Buffalo Genomics

Local Ancestry Analysis Suggests Adaptive Introgression of River-Buffalo Derived Regions in the Genome of Brazilian Carabao Swamp Buffaloes

Licia Colli[1,2], Mario Barbato[1], Marco Milanesi[3,4], Yuri Tani Utsunomiya[3,4], Jesus Rommel V. Herrera[5], Libertado Cruz[6], Pietro S. Baruselli[7], Elisabete Amaral[8], Marcela Gonçalves Drummond[9], Jose Fernando Garcia[3,4], John Williams[10], Paolo Ajmone Marsan[1,2] and International Water Buffalo Consortium, (1)Dept. of Animal Science, Food and Nutrition - DIANA, Università Cattolica del S. Cuore, Piacenza, Italy, (2)Research Center on Biodiversity and Ancient DNA – BioDNA, Piacenza, Italy, (3)Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ. Estadual Paulista, Sao Paulo, Brazil, (4)International Atomic Energy Agency - IAEA, Sao Paulo, Brazil, (5)Philippine Carabao Center, College, Laguna, Philippines, (6)Philippine Carabaou Centre, Nueva Ecija, Philippines, (7)Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, Brazil, (8)UNESP Sao Paulo State University, Sao Jose Rio Preto, Brazil, (9)Myleus Biotecnologia, Belo Horizonte, Brazil, (10)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia

The domestication of water buffalo (*Bubalus bubalis*) happened ca. 5,000-6,000 years ago and has profoundly impacted human societies as the main livestock resource in tropical and subtropical region. Currently, two sub-species have been identified, namely river and swamp buffalo, which show different morphology, behaviour, production purposes and geographical distributions. Previous genome-wide analyses performed on 31 river and swamp buffalo populations from 16 regions worldwide found evidence of composite river/swamp genomic makeup in Carabao, a swamp type water buffalo breed from Brazil. To better understand the nature and putative adaptive value of the river type-derived introgressed genomic regions, we applied local ancestry investigations on 18 Carabao individuals genotyped for 90k SNPs. For comparison, the same analysis was performed on a recently established crossbred river *x* swamp population from the Philippines. Local ancestry results showed larger haplotypes of river-derived ancestry in the Philippine population with respect to the Carabao breed, confirming the more recent admixture in the Philippine population. Conversely, we identified shorter river type-derived haplotypes in Carabao, suggesting a more ancient event of introgression. Several admixed genomic regions appeared conserved among Carabao individuals, possibly due to their putative adaptive nature. Among them, a region located on chromosome 4 intercepted a QTL affecting somatic cell score in cattle. Further, a highly conserved region on chromosome 29 intercepted genes related to fertility in cattle. Our preliminary results suggest that adaptive introgression occurred in the Carabao population, provide better tools for future improvement of the species, and a better understanding of the underlying demography.

W136: Buffalo Genomics

Interferon Omega-1 Gene as a Highly Polymorphic Immune Genes in Pakistani Buffalo Breeds

Masroor Ellahi Babar, Tanveer Hussain, Abdul Wajid, Akhtar Ali, Kamran Abbas and Fahad Rafiq, Virtual University of Pakistan, Lahore, Pakistan

Buffaloes are considered as major dairy animal in Pakistan with five breeds providing milk, meat, hides and power. The genomics of buffalo need to be explored to better understand its potentials. Like other mammals the virus-infected buffalo cells release interferons to trigger the protective defenses of the immune system for pathogens eradication. Among type I interferons the IFNW1 gene encodes for Interferon omega-1 protein having role in immune response. We aimed to have insight about this gene in Pakistani buffalo breeds to identify potential molecular markers for better resistant animals. We genetically characterized the IFNW1 gene fragment (436 bp) in Pakistani River buffaloes (*Bubalus bubalis*) i.e. Nili (NI n=5), Ravi (RA n=4), Nili-Ravi (NR n=4), Kundi (KU n=2) and AzaKheli (AZ n=5) by PCR amplification followed by direct sequencing. Sequences were aligned and edited through CodonCode Aligner for polymorphism identification. The phylogenetic analysis was inferred by using the maximum likelihood method based on the Tamura-Nei's model, with standard error based on 1000 bootstrap replicates through MEGA 6.1. The sequencing analysis showed the IFNW1 gene was highly polymorphic in nature in all five studied buffalo breeds. We identified this gene highly polymorphic with 32 variants in all five breeds with 18 non-synonymous polymorphisms while 14 were synonymous substitutions. It was indicated that the IFNW1 sequences from four breeds NI, RA, NR and AZ with other buffalo sequences form a branch independent of the *Bos taurus* and *Bos mutus* species, with exclusion of KU. To the best of our knowledge, this is the first report of its kind in water buffaloes based on sequence diversity of immunity genes. The identified polymorphisms may be considered as baseline towards identification of molecular markers to be included in breeding and selection programs for selection of individuals with better immunity in days to come.

W137: Cacao Genomics Workshop

Transcriptomic Profiling of Cacao Genotypes Resistant and Susceptible to Phytophtora

Désiré N. Pokou, CNRA-LCB, Abidjan, Cote d'Ivoire and Andrew S. Fister2*, Noah Winters2, Mathias Tahi1, Coulibaly Klotiolonan1, Aswathy Sebastian2, Siela N. Maximova2,, Mark J. Guiltinan2, (1)Centre National de Recherche Agronomique, Laboratoire Central de Biotechnologies, 01BP 1740 Abidjan 01, Côte-d'Ivoir

Phytophthora megakarya, is an oomycete which constitute the of most aggressive back pod disease pathogen of cacao (*Theobroma cacao* L.). This species occurs only in Africa where 74% of the world cacao is produced. Considerable effort had been done in the past years to develop resistant varieties from the existing germplasm collection. However, deep understanding of the host-pathogen interaction seems important to direct the scientific effort of developing new resistant varieties for this region. Here we analyzed the transcriptome of the well-known resistant genotype SCA6 and the susceptible NA32 in order to characterized basal expression and polymorphism of defense genes. Samples from the two

genotypes split in separate group on the PCA graph showing different expression of the whole transcriptome. We detected induction of several Pathogenesis-Related genes, Pattern Recognition Receptors, and Resistance genes which could be critical for SCA6's ability to combat infection.

W138: Cacao Genomics Workshop

Dissecting the *Moniliophthora Perniciosa* x Host Interaction Using 'micro-Tom' Tomato As a Proxy: A Functional Genomic Approach

Antonio Figueira[1], Daniele Paschoal[2], Eder Marques Silva[3], Juliana Leles Costa[2], Danielle C. Scotton[4], Fabia Barbosa da Silva[4], Diogo Capeli[4], Ricardo F. Oliveira[4], Paulo J Teixeira[5], Gonçalo Amarante Guimarães Pereira[6] and Lazaro EP Peres[7], (1)University of Sao Paulo, Piracicaba, Brazil, (2)Plant Breeding Laboratory, Center for Nuclear Energy in Agriculture, University of Sao Paulo, Piracicaba, Brazil, (3)ESALQ/USP, Piracicaba, Brazil, (4)Universidade de Sao Paulo, Piracicaba, Brazil, (5)University North Carolina, Chapel Hill, NC, (6)Genomics and Expression Laboratory - University of Campinas (UNICAMP), Campinas, Brazil, (7)ESALQ, Piracicaba, Brazil

The basidiomycete *Moniliophthora perniciosa* is the causal agent of the witches' broom disease of cacao, the major limiting factor for production in South America. This unusual pathogen presents a hemibiotrophic lifestyle, with a long and peculiar biotrophic stage. The biology and life style of cacao restrict genetic and functional studies. We have shown that the miniature tomato cultivar 'Micro-Tom' (MT) is a suitable genetic model to study the pathogenic interaction with isolates from the S-biotype, exhibiting typical symptoms of the infection. The differential MT response against the compatible S-biotype or the incompatible cacao C-biotype offers a potential to identify presumed defense responses. Ectopic expression of cacao genes in MT has indicated potential resistance mechanisms. Mutants and transgenic lines altered for the synthesis or perception of hormones in the MT background have demonstrated the role of cytokinin (CK) and auxin during pathogenesis. The role of CK imbalance on sink manipulation of the host by the pathogen and the consequent physiological and metabolic effects on MT have been established. The major impact of shoot infection on fruit and biomass yield could be definitely quantified in tomato. Mutants for genes associated with development (eg. transition from vegetative to reproductive stage; senescence) are helping to shed some light on the physiological effects of infection. The vast tomato genetic resources warrant progress in the elucidation of gene functions during the *M. perniciosa* x host interaction.
Funding: FAPESP (2016/10498-4), CAPES and CNPq.

W139: Cacao Genomics Workshop

Long-Read Sequencing Provides Insights into Large Scale Genomic Expansions in the *Phytophthora palmivora* and *megakarya* Genomes

Abraham Morales-Cruz[1], Shahin Ali[1,2], Rosa Figueroa-Balderas[1], Andrea Minio[1], Jean-Philippe Marelli[3], Bryan Bailey[2] and Dario Cantu[1], (1)University of California Davis, Davis, CA, (2)Sustainable Perennial Crops Lab, USDA-ARS, Beltsville, MD, (3)MARS/USDA Cacao Laboratory, Miami, FL

Phytophthora palmivora and *megakarya* are oomycete plant pathogens that cause black pod rot in cacao (*Theobroma cacao*). These pathogens typically contribute to annual crop yield losses between 20-30%, as well as the death of 10% of the trees annually. *P. megakarya* is the more aggressive of the two species causing up to 100% of crop loss in non-managed plantations. In order to access the highly repetitive intergenic space and recently expanded regions, which are likely missing in the previous genome assemblies of the two organisms, generated using short reads sequencing assemblies, we applied Single Molecule Real Time (SMRT) sequencing at high coverage (>150x). The long reads were assembled with the diploid-aware software FALCON-Unzip and polished with Arrow. Primary contigs were then scaffolded with SSPACE and gap-filled with PBJelly. The *P. megakarya* primary genome was assembled in 502 scaffolds (N50 = 780 kbp) for a total size of the primary assembly of 214 Mbp. For *P. palmivora*, primary assembly size was of 115 Mbp (N50 = 877 kbp). Alternative haplotypes (haplotigs) accounted for 114 Mb in *P. megakarya* and 39 Mb in *P. palmivora*. Overall, the repeat content was estimated at 37% and 30% for the *P. megakarya* and *P.palmivora* genome assemblies, respectively. Gene space completeness was estimated with the BUSCO Eukaryota dataset detecting 86% and 85% of the conserved ortholog genes in *P. megakarya and P. palmivora*, respectively. RNA-Seq data from mycelia, zoospores and two *in planta* infection-time points were *de novo* assembled and used as evidence for gene prediction with BRAKER. Gene models were finally refined with the PASA gene annotator. In the *P. megakarya* genome, a total of 53,476 protein-coding genes were predicted in the primary assembly and 30,209 in the haplotigs. In the *P. palmivora* genome, 30,283 genes were predicted in the primary assembly and 11,978 in the haplotigs. The two highly contiguous genome assemblies will allow to study how segmental duplications and transposable element proliferation have shaped the evolution of these destructive pathogens.

W140: Cacao Genomics Workshop

New Insights for an Ancient Plant Disease: Undescribed Species of *Phytophthora* Causing Black Pod on Cacao in Brazil

Jean-Philippe Marelli[1], Jennifer Decloquement[2], Roberto Ramos Sobrinho[2], Dahyana S. Britto[3], Adilson Barreto[4], Jaime Honorato Junior[5] and Danilo Pinho[2], (1)Mars Wrigley Confectionery, Miami, FL, (2)University of Brasilia, Brasilia, Brazil, (3)Mars Center for Cocoa Science, Itajuipe, Brazil, (4)Embrapa, Cruz das Almas, Brazil, (5)Federal University of West of Bahia, Barra, Brazil

Cacao (Theobroma cacao) production is drastically reduced by the occurrence of diseases. In Brazil, increased incidence levels of black pod have been observed in the main cacao-growing areas, with Phytophthora capsici, P. citrophthora, P. heveae and P. palmivora being the main species identified by morphological comparisons. Molecular identification is a necessary tool in Phytophthora taxonomy, revealing additional species previously unreported in different crops. In this context, the present work aimed to morphologically and molecularly characterize Phytophthora species causing black pod disease on cacao in Brazil. A total of 40 hyphal tip isolates were obtained from symptomatic cacao pods collected in four distinct areas in the state of Bahia in 2017. Partial nucleotide sequences of the β-tubulin gene were used for preliminary identification of Phytophthora species, and then representative isolates from each species and sampling location were selected for sequencing of additional regions (COXII, EF1α, ITS and HSP90), morphological characterization, and evaluation of aggressiveness and in vitro mycelial growth. The combination of multigenic analysis with morphological comparisons confirmed the presence of P. palmivora. Interestingly, many isolates grouped in the Phytophthora Clade 2 and clustered apart from previously reported species, being considered a putative new species that must be proposed following the International Code of Nomenclature for Algae, Fungi and Plants. The new species produced papillate, persistent

sporangia on simple sympodially branched sporangiophores. While P. palmivora isolates were identified as A1 mating type by pairing each isolate with known A1 and A2 tester strains of P. capsici, oogonia/antheridia were not observed in isolates of Phytophthora sp. The mycelial growth rate on potato-dextrose-agar (PDA) was lower than on carrot-agar (CA), V8 juice-agar (V8A) and malt extract-agar (MA) when isolates were grown at 20°C. Phytophthora sp. grew faster than P. palmivora at 10, 15, 20 and 25°C, but they showed similar mycelial growth on clarified V8A at 30°C. Pathogenicity tests conducted on pods of four cacao clones (CCN51, PS1319, Cepec2004 and CP49) showed that Phytophthora sp. is more aggressive than P. palmivora. Finally, a ratio of 1:0.5 (Phytophthora sp./P. palmivora) was found. In this study, only two species of Phytophthora were found, and the putative new species seems to be more frequent and aggressive than P. palmivora.

W141: Cacao Genomics Workshop

Genomic Characterization of the Cacao Swollen Shoot Badnavirus 'Species Complex' in West Africa

Judith K. Brown, School of Plant Sciences, The University of Arizona, Tucson, AZ and Osman A. Gutierrez, USDA ARS SHRS, Miami, FL

In West Africa, cacao swollen shoot disease (CSSD) badnaviruses are economically important pathogens of the cocoa tree (*Theobroma cacao* L.) causing extreme yield loss and ultimately tree death. The disease is caused by a group of evolving badnavirus species [genus, *Badnavirus;* family, *Caulimoviridae*] that are transmitted by at least 14 species of mealybugs (Pseudococcidae: Hemiptera) are, with the best studied vectors being *Formicococcus njalensis* (Laing)*, Planococcus citri* (Risso), *Pseudococcus viburni* (Signoret), *and Ps.longispinus* (Targioni Tozzetti). The mode of transmission is considered to be semi-persistent, with a 72-hr acquisition access period (IAP) and retention time of 3-6 days. Quantitative PCR analysis has shown that juvenile mealybugs retain at least some viral load until they undergo ecdysis, during which time the lining of the foregut to which the virus is expected to attach, is replaced after molting. The virus can be experimentally transmitted by grafting, and mechanically from sap with great difficulty. With the recent discovery that multiple species of closely related badnaviruses are associated with the disease, the CSSD-associated badnaviruses are considered a 'species-group'. The CSSD badnaviruses have a circular, gapped dsDNA genome, ranging in size from ~6.8-7.3 kbp, encapsidated in a non-enveloped bacilliform particle ~128 × 28 nm. Phylogeographic (Bayesian, ML) analyses of CSSD-West African badnaviruses (82 genomes) and of two genomes sequenced from the cacao GeneBank in Trinidad indicated that the origins of the two groups differ, with African-CSSD badnavirus isolates showing strong regional affinity e.g. West Africa, while those from Trinidad cacao are most like each another and certain badnaviruses from Asia. Pairwise distance analysis of the RT-RNase H region and of the whole genome sequence indicated ~70-100% and 61-100% shared nt identity, respectively, indicative of ~ten species. Among them, the genome organization and predicted conserved protein domains (CPDs) are variable, sharing some conserved architecture. However, several CPDs not previously seen for cacao-infecting badnaviral genomes have been identified: DWNN, DUF3552, SEEEED, SMC, AIR1, DUF4939, and peptidase A3. Closer analysis of the CPDs are expected to provide important functional clues about virulence and/or differences in biological characteristics. Phylogeographic analyses strongly indicate that the CSSD-badnaviruses occurred in endemic non-cacao hosts when cacao was introduced to West Africa as a crop plant >100 years ago. The observations further suggest that extant CSSD-associated badnaviral species causing swollen shoot disease shared a common ancestor(s), and infect cacao following independent host-jumps from their wild host to cacao shortly after its arrival and/or recently. Collectively, these recent analyses support a West African endemism for CSSD-badnavirus species, and implicate the western region of Africa as the major center of CSSD-badnavirus diversification on the African continent. Recent results from Illumina sequencing have implicated non-endemic cacao-infecting badnaviruses in cacao samples in Ghana. Given their closest shared sequence identity with the three Caribbean (2) and Sri Lanka (1) badnaviruses, it seems plausible that the 'putative exotic' badnaviruses have been introduced into West Africa through movement of infected germplasm, despite established quarantine practices. Their distribution and potential contributions to the ongoing regional pandemic are not known.

W142: Cacao Genomics Workshop

Genome and Transcriptome Analysis of the Latent Pathogen *Lasiodiplodia theobromae*, an Emerging Threat to Cacao Industry

Shahin Ali[1], Jonathan Shao[2], Asman Adi[3], Mary Strem[4], Johnny Balidion[5], Lyndel Meinhardt[1] and Bryan Bailey[1], (1)Sustainable Perennial Crops Lab, USDA-ARS, Beltsville, MD, (2)ARS-USDA BARC, BELTSVILLE, MD, (3)Hasanuddin University, Sulawesi Selatan, Indonesia, (4)SPCL BARC ARS-USDA, BELTSVILLE, MD, (5)University of the Philippines, Los Baños, Philippines

Lasiodiplodia theobromae, a member of the *Botryosphaeriaceae* family, is becoming a significant threat to crops and woody plants in many parts of the world, including South East Asia, a major cacao growing areas. A phylogenetic analysis of *Lasiodiplodia* isolates indicates that at least three different species are associated with the cacao in South East Asia. The current study reports on a 43.75 Mb de novo genome of a *L. theobromae* isolate from cacao. *Ab initio* gene prediction identified 13,061 protein-coding genes. Transcriptome analysis by RNA-Seq of infected cacao leaves revealed that 11,860 gene models were transcriptionally active and 1,255 were induced *in planta* compared to mycelia. Many genes involved in carbohydrate, pectin and lignin catabolism, as well as cytochrome P450s, necrosis-inducing proteins and putative effectors were among the genes induced during cacao infection. To assess its abilities to alter/suppress cacao defense, we also studied the cacao genes that were differentially expressed during the *L. theobromae* infection. These findings significantly expand our knowledge of the *L. theobromae* genome and genes potentially involved in virulence and pathogenicity in association with the infection process in cacao.

W143: Cacao Genomics Workshop

Functional Genomics of Somatic Embryogenesis of *Theobroma cacao*

Siela N. Maximova, The Pennsylvania State University, University Park, PA

Siela N. Maximova, Adriana M. Gallego, Lucía Atehortúa, Andrew S. Fister, Natalia Pabón-Mora, Lena Landherr, Melanie Perryman, Yufan Zhang and Mark J. Guiltinan

Somatic embryogenesis (SE) can be applied to speed up development and deployment of improved genotypes of cacao. Genomics research reveals molecular mechanisms controlling the SE process that could be utilized for optimization of the SE protocols or alternatively could be applied for development of other commercial application of SE cultures. One of the key transcription factors known to regulate embryogenesis in

other species is LEC2. To increase the efficiency of cacao SE, we conducted functional genomics study of a transgene encoding a fusion of *Leafy Cotyledon 2* (*TcLEC2)* to a glucocorticoid receptor domain (*GR*) to control nuclear localization of the protein and we demonstrated for the first time that we can grow to maturity plants regenerated from cacao leaf tissue. Thus regulating TcLEC2 activity offers a powerful new strategy for optimizing somatic embryogenesis pipelines for cacao. Additionally we studied the influence of different light conditions on regulation of genes involved in the flavonoid biosynthesis pathway during SE. We incubated cacao cell suspensions under white-blue light and darkness for 14 days and conducted RNA-Seq analysis. Exposure to different light conditions resulted in faster accumulation of phenolic compounds and shifted the ratios of catechin/epicatechin as a response to switching from white to blue light. We demonstrated that genes encoding *HY5, MYB12*, *ANR* and *LAR* were differentially regulated under light/dark conditions. In conclusion, our RNA-Seq analysis of cacao cells cultured under different light conditions provides a platform to dissect key aspects into the genetic regulatory network of flavonoids. The results can be applied to increase SE efficiency or as an incubator for commercial production of catechins.

W1**44: Camelids**

Improving the North African Dromedary Genome with Hi-C and PacBio Reads

Jean P. Elbers[1], Mark Rogers[2], Polina Perelman[3], Anastasia A. Proskuryakova[3], Natalia A. Serdyukova[3], Warren E. Johnson[4,5], Petr Horin[6], Jukka Corander[7,8], David Murphy[2] and Pamela A. Burger[1], (1)Research Institute of Wildlife Ecology, Vetmeduni, Vienna, Austria, (2)University of Bristol, Bristol, England, (3)Institute of Molecular and Cellular Biology and Novosibirsk State University, Novosibirsk, Russian Federation, (4)The Walter Reed Biosystematics Unit, Museum Support Center MRC-534, Smithsonian Institution, Suitland, MD, (5)Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA, (6)Inst. of Genetics, Dept of Pathobiology, Faculty of Vet Med, CZ-612 42 BRNO, Czech Republic, (7)Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland, (8)Department of Biostatistics, University of Oslo, Oslo, Norway

Thousands of eukaryotic genomes have been assembled with Illumina sequencing; however, bacterial artificial chromosomes (BACs) are needed to span long repetitive elements in eukaryotic genomes. Newer chromosome conformation capture techniques such as Hi-C and Dovetail Genomics Chicago libraries are an alternative to BAC sequencing. Long-read sequencing from Pacific Biosciences and Oxford Nanopore is advantageous to Illumina sequencing, but long reads have more insertions and deletions. One of the last livestock species that does not have a high-quality genome is the dromedary (*Camelus dromedarius*). Draft genomes from North African and Arabian dromedaries exist, but both are highly fragmented. Here we describe our efforts to improve the North African dromedary genome assembly (i.e., CamDro1). We first used Chicago and Hi-C sequencing libraries from Dovetail Genomics to resolve the placement and order of previously assembled CamDro1 contigs, producing almost chromosome-level scaffolds. We then filled in scaffold gaps with Pacific Biosciences (PacBio) long reads, and finally these scaffolds were mapped to chromosomes. This generated the CamDro2 assembly for which we split these scaffolds at gaps, repeated scaffolding of the resulting contigs with Hi-C and Chicago libraries, and then filled in scaffold gaps with PacBio reads producing the CamDro3 assembly- a slight improvement over the CamDro2 assembly. PacBio reads added approximately 100Mbp to the CamDro2 assembly. Dovetail Chicago and Hi-C libraries increased the longest scaffold over 12-fold, and the scaffold N50 over 50-fold between the CamDro1 and CamDro2 assemblies. We demonstrate that Illumina-only assemblies can be substantially upgraded by adding chromosome conformation capture and long-read sequencing.

W1**45: Camelids**

Preliminary Sequence Assembly of the Alpaca (*Vicugna pacos*) Y Chromosome

Matthew Jevit[1], Andrew Hillhouse[1], Mark F. Richardson[2], Brian W. Davis[1], Rytis Juras[1], Malcom Ferguson-Smith[3], Ahmed Tibary[4], Vladimir Trifonov[5] and Terje Raudsepp[1], (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

We recently initiated a systematic study of the alpaca (Vicuna pacos) Y chromosome to better understand male development and fertility. Here we present the preliminary Y sequence assembly. The flow-sorted Y DNA was directly sequenced on PacBio Sequel and Illumina MiSeq (2x300bp reads) platforms. Long reads were assembled with CANU and short reads were incorporated with PILON. The assembly is 20,050,105 bp in 652 contigs. The longest contig was 1,433,978 bp and the shortest was 1,142 bp. The mean and median contig lengths were 30,751 bp and 5,852 bp, respectively. The N50 was 288,720 bp and the GC content was 39.8%. Assembly quality control and annotation of single copy and multicopy genes are in progress. We have concatenated Y assembly with the latest alpaca reference genome and have mapped the short reads, cDNA sequences, and Trinity assembled transcripts back to the genome. Variable sites will be called using SAMtools. Dimorphic sites will be considered diploid loci, considered autosomal contamination, and removed from the assembly after being inspected for possible autosomal and X transposition. Additionally, polymorphic sites may be indicative of multicopy genes, extracted from the assembly and identified via BLAST. Analysis of cDNA sequencing of Y mRNA has revealed transcripts for 31 putative Y genes, 12 of which were confirmed to be male specific by PCR. Gene annotation and identification of repeats will be completed with MAKER. We believe that this first Y assembly in camelids will have important implications for the management of these species.

W1**46: Camelids**

Dromedary Camel Genome Sequence

Samantha A. Brooks[1], Heather M. Holl[1], Donald Miller[2], Salma Abdalla[3], Benjamin M. Shykind[3], Joel Malek[3], Yasmin A. Mohamoud[4], Ayeda Ahmed[5], Kamaal Pasha[6] and Douglas Antczak[2], (1)University of Florida, Gainesville, FL, (2)Cornell University, Ithaca, NY, (3)Weill Cornell Medical College, Doha, Qatar, (4)WCMC-Qatar, Doha, Qatar, (5)Weill Cornell Medical College in Qatar, Doha, Qatar, (6)Tharb Camel Hospital, Doha, Qatar

Known as “the ship of the desert”, the dromedary camel (*camelus dromedaries*) plays an important role in agriculture as a source of milk, meat, and transportation. Their success as livestock is in part due to their remarkable physiological adaptations for survival in arid environments. To provide a powerful new tool for study of the dromedary we have generated a hybrid genome assembly utilizing 19x PacBio long reads, 96x Illumina paired end reads, and 38x 10X Genomics linked reads. Following repeatmasking (34.67% of the genome) the scaffolds were further

assembled into predicted chromosome fragments (PCFs) using RACA (Kim et al. 2013). The cow genome (bosTau8/UMD3.1.1) was selected to serve as the reference and the human genome (hg38) as the outgroup. Reference-assisted chromosome assembly (RACA) joined the assembled 3,165 scaffolds (N50 5,258,276bp) in to just, 100 predicted chromosome fragments (PCFs, N50 46,255,429bp). Future work will include annotation of the assembly using transcriptome assemblies, as well as examination of population structure using polymorphic SNPs derived from full genome Illumina sequence of from 17 camels and reduced representation sequencing of 238 additional camels.

W1**47: Camelids**

Chromosome-Level Alpaca Reference Genome VicPac 3.1 Improves Genomic Insight into the Biology of New World Camelids

Mark F. Richardson[1], Kylie Munyard[2], Lawrence J Croft[1], Theodore R Allnutt[1], Polina Perelman[3], Matthew Jevit[4], Fahad Alshanbari[4], Rhys Cransberg[2], Ahmed Tibary[5], Belinda Appleton[1] and **Terje Raudsepp**[4], (1)Deakin University, Geelong, Australia, (2)Curtin University, Perth, Australia, (3)Institute of Molecular and Cellular Biology and Novosibirsk State University, Novosibirsk, Russian Federation, (4)Texas A&M University, College Station, TX, (5)Washington State University, Pullman, WA

The development of high-quality chromosomally assigned reference genomes constitutes a key feature for understanding genome architecture of a species and is critical for the discovery of genetic blueprints of traits of biological significance. South American camelids serve people in extreme environments and are important wool and companion animals worldwide. Despite this, alpaca reference genome lags far behind of those available for other domestic species. Here we produced a chromosome-level improved reference assembly for the alpaca genome using the DNA of same female Huacaya alpaca as in previous assemblies. We generated 122X Illumina short-read, 5X Pacific Biosciences long-read and Dovetail Genomics HiRise scaffolding data for the assembly; testis and skin RNAseq data for annotation, and used cytogenetic map data for chromosomal assignments. The new assembly VicPac3.1 contains 90% of the alpaca genome in just 103 scaffolds and 88% of all scaffolds are mapped to the 36 pairs of the alpaca autosomes and the X chromosome. Altogether, the assembly shows 18,958 genes and over 12,000 novel isoforms. Comparative analysis of selected regions of the alpaca genome, such as the major histocompatibility complex and candidate genes for high-altitude adaptations, reveal unique features of the alpaca genome. The alpaca reference genome VicPac3.1 presents a significant improvement in contiguity and accuracy over VicPac2 and is an important tool for the advancement of genomics research in all New World camelids.

W1**48: Camelids**

Current Status and Potential of Genetic Improvement in Dromedary Camels

Mohammed Al Abri, Sultan Qaboos University, Muscat, Oman

Dromedary camels have developed through millennia the ability to produce quality meat, milk and fiber and transport goods in some of the most hostile environments in the globe. As impacts of climate change worsen, it is evident that camels could be a livestock of choice for sustainable protein production especially in hot and dry countries. However, to harness the potential of camels, an improved understanding of the genetics underlying their unique biology is needed. Nevertheless, less attention has been paid to camels genetics compared to other livestock species. As a results, there are relatively few published studies in the area of camel genetics and genomics. This is due, in part, to the lack of genomics tools to conduct such studies. For instance, no commercial genotyping platform has yet been developed for camels. Thus, whilst many QTLs have been reported in sheep, cattle and horses, only few of QTLs were reported in camels. Today, with the exception of racing camels, very limited genetic selection is applied in camels. Various studies indicated that camels have a high genetic variability due to the lack of selection and the current and historical movements of camels between countries for trade and sometimes war. This variability was reflected in the high estimates of heritabilities of various dairy traits, indicative of the potential for ample genetic gain if systematic selection is to be implemented. Currently, the demand for milk and red meat in many MENA countries is met either by importation or local production using exotic livestock that are not adapted neither to the local climatic conditions nor the low input systems dominating the region. However, during the last decade, camel milk and meat products have been widely accepted both locally (in arid regions) as well as internationally. As a result of this, a few camel intensive dairy farms have been founded and are currently successfully operating. Here, we discuss the potentials and challenges of genetic improvement in camels. We argue that genomic selection could boost the productivity of camels particularly because they are not traditionally pedigreed. Genomic selection is also specifically recommended for camels because of their unique physiology (long gestation period and generation interval).

W1**49: Camelids**

Molecular Mechanism of Camel Fat Metabolism and Deposition in Nutritional Restriction

Fucheng Guo[1], Liang Ming[2], Li Yi[2], Yu C. Xiao[1] and Rimutu Ji[3], (1)Inner Mongolia Agricultural University, Hohhot,China, Hohhot, China, (2)Inner Mongolia Agricultural University, Hohhot, China, (3)Inner Mongolia Agricultural University, Hohhot, Inner Mongolia, China

Adipose tissue is important for maintaining energy and glucose homeostasis at both the organ and systemic level. It effectively stores energy in the form of lipids and monitors the distribution of fat throughout the body. Adipose tissue also performs endocrine functions via the production of adipokines, through which it coordinates with other organs to monitor several metabolic functions. Domestic (Camelus bactrianus) and wild (Camelus ferus) Bactrian camels are one of the few large livestock in the world that can survive in the Gobi-desert, and famous for its unique tolerance to environmental stresses. The Bactrian camels have developed several mechanisms that help them adapt to the harsh environments of deserts. One such adaptation is the storage of energy as fat in their humps and abdomen; this helps them survive long durations of time in which they do not get food or water. Camels can survive for 67 days under fasting and water-free conditions. Here, We selected 6 male Bactrian camels (40 months old) with a nutritional limit of 15 days, then returned to a normal diet. we used RNA sequencing and liquid Chromatograph Mass Spectrometer examine the change of transcriptome profiles and metabolite in the Bactrian camel: fore hump, hind hump, subcutaneous, muscle (Before the nutritional restriction, the 15th day of the nutritional restriction, and the 23rd day after the normal diet, fore hump, hind hump, subcutaneous, muscle were collected respectively .) When hunger, fat and muscle are important energy-supplement sites for hunger. By

comparing their changes under nutrient limitation and recovery conditions, we look forward to discovering the special fat metabolism and deposition pathways of camels and further verifying them.

Key Words: Nutrient limitation; fat metabolism and deposition; *Camelus dromedarius*

This work was supported by the National International Scientific and Technological Cooperation Project [grant numbers 2015DFR30680, ky201401002], International and Domestic Science and Technology Cooperation Project [grant number 20140438], the National Natural Science Foundation of China [grant number 31360397] and Major Projects of Inner Mongolia Autonomous Region，and Alashan Desert Industrial Research Institute Project.

W150: **Cannabis Genomics and Breeding**

Current *Cannabis* Labelling Practices Poorly Capture Genetic Reality

Sean Myles[1], Michel McElroy[1], Jason Sawler[2], Arno Hazekamp[3], Jonathan Page[4] and Hugo Maassen[5], (1)Dalhousie University, Truro, NS, Canada, (2)Anandia Labs, Vancouver, BC, Canada, (3)Hazekamp Herbal Consulting, Leiden, Netherlands, (4)Aurora Cannabis, Vancouver, BC, Canada, (5)Bedrocan International, Veendam, Netherlands

The Cannabis industry tends to classify varieties using the terms 'Indica' or 'Sativa' as a way of informing consumers about their effects, aromas or purported pedigree. They also make use of variety names that are often shared among producers. The degree to which these labels correspond to any meaningful biology remains largely unexplored. To investigate the genetic basis of current ancestry and naming practices, I present analyses of genome-wide SNP data from over 200 Cannabis samples. We find only a moderate correlation between the genetic structure of Cannabis strains and their reported ancestries. Moreover, we show that Cannabis strain names often do not reflect any meaningful genetic identity. However, by assessing 35 metabolites across 150 samples, we reveal a strong relationship between ancestry labels and specific terpenes: Sativa ancestry was associated with farnesene and bergamotene, while Indica ancestry was associated with myrcene, elemene, and sesquiterpene alcohols. Our results suggest that, rather than reflecting two independent ancestral populations converging through hybridization, the Indica-Sativa spectrum in Cannabis may instead be largely the result of two ideotypes having been selected for during breeding from a relatively unstructured gene pool.

W151: **Cannabis Genomics and Breeding**

Unraveling the Mysteries of CBD and THC Content with a Chromosome Resolved *Cannabis* Genome

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

Human demand for non-psychoactive Cannabis with high concentrations of cannabidiol-acid (CBDA) and low delta-9-tetrahydrocannabinol-acid (THCA) has resulted in admixed populations. The enzymes *THCA* and *CBDA synthase* compete for a common precursor and copy number as well as sequence variation have been offered as alternative explanations for THCA/CBDA ratio. However, our understanding of the underlying genomic architecture of the CBDA/THCA synthase loci has been limited by the repetitive nature of the Cannabis genome. We assembled a high-CBD strain with Oxford Nanopore reads and resolved the genome into chromosomes with a genetic map, which enabled the resolution of 12 CBDA and THCA synthase cassettes in two linked tandem arrays deep in a low recombination island on chromosome 9. The CBDA/THCA tandem arrays are comprised of 30-80 kb cassettes of specific LTR retrotransposons, suggesting a putative mechanism for copy number variation across cultivars, and shedding light on the storied history of Cannabis breeding.

W152: **Cannabis Genomics and Breeding**

Clemon

Clemon J. Dabney III, Weiblen Laboratory, Plant and Microbial Biology, University of Minnesota, St. Paul, MN

W153: **Cannabis Genomics and Breeding**

Kevin

Kevin McKernan, Medicinal Genomics, Woburn, MA

W154: **Cannabis Genomics and Breeding**

Alisha

Alisha Holloway, Phylos, Portland, OR

W155: **Cannabis Genomics and Breeding**

Reference Genome and Whole Genome Resequencing in Medicinal Cannabis for Genomic Selection and Genome Editing, Enabling Precision Breeding

Shivi Braich, Lennon Matchett-Oates, Bec Baillie, Ehab Mohamaden, Karen Fulgueras, Simone Rochfort, German C. Spangenberg and **Noel O.I. Cogan**, Agriculture Victoria, Bundoora, VIC, Australia

The legal use of medicinal cannabis has increased rapidly, however cannabis is the only crop of significant dollar value that has largely evaded recent technology advances in agriculture in terms of accelerated precision breeding. We have generated three genome assemblies of *Cannabis sativa* strains, using long-read sequencing technology, augmented with Hi-C based sequencing. These genomes have provided the reference for whole genome resequencing of over 530 strains at a range of sequencing depths that has delivered c. 2.7 million SNP variants across the genome. These genome sequences are enabling detailed comparative genomics as well as pan-genome analysis. This work supports genomic selection and predictive breeding through the development of a reference population, complemented by detailed chemotypic data (Rochfort – PAG XXVII). The discovered sequence variants also enable an informed design of genome editing tools. A tailored set of *in vitro* tools and technologies for genome editing enabling routine transformation and regeneration is also being developed in concert. The rapid implementation of genomic selection approaches as well as genome editing for the generation of novel designer strains of cannabis will deliver significant benefit to the industry and enable tailored chemotypic profiles to be realised.

W156: **Cannabis Genomics and Breeding**

Molecular Phenotyping of Medicinal Cannabis Strains - Enabling Genomic Selection and Gene Annotation

Simone Rochfort, Aaron Elinkins, Christian Krill, Delphine Vincent, Vilnis Ezernieks and German C. Spangenberg, Agriculture Victoria, Bundoora, VIC, Australia

Cannabis is an herbaceous flowering plant of the Cannabis genus (Rosale) that has been used for its fibre and medicinal properties for thousands of years. In recent decades medicinal cannabis has become legal in several jurisdictions and the possibility of legalisation is being explored in many more. This legalisation has opened the field of medicinal cannabis research. The biochemistry of cannabis is rich and varied including phytocannabinoids, terpenes and phenolics. Each of these metabolite classes contains individual compounds with biological activity. This chemical diversity and the interaction between molecules may underpin the 'entourage effect' that is believed to contribute to the medical efficacy of cannabis. In order to fully explore this biochemistry we have undertaken both targeted and untargeted metabolomic, volatolomic and proteomic analysis of diverse cannabis strains. 70 strains have been characterised using liquid chromatography mass spectrometry (LCMS) and gas chromatography mass spectrometry (GCMS). When coupled with genomics data this comprehensive molecular phenomics data enables genomic selection strategies to breed for chemotypic specific strains. The data is also being used to annotate genes and gene pathways for the different molecular species.

W157: **Cassava Genomics**

Milestones Attained Towards Optimizing Cassava Breeding Programme of Uganda

Robert Kawuki, National Root Crops Resources Research Institute, Namulonge, Uganda

W158: **Cassava Genomics**

West African Virus Epidemiology (WAVE) for Root and Tuber Crops

Justin S. Pita, Université Félix Houphouët-Boigny, Abidjan, Cote d'Ivoire and **Joseph Onyeka**, National Root Crops Research Institute (NRCRI) Umudike, Umuahia, Nigeria

W159: **Cassava Genomics**

Leveraging Integrated Data Management and Informatics Platform to Accelerate Breeding in a Modern Cassava Breeding Program

Alex C. Ogbonna[1,2], Guillaume J. Bauchet[2], Isaak Y. Tecle[2], Naama Menda[2], Bryan Ellerbrock[2], Nicolas Morales[2], Titima Tantikanjana[2], David A. Lyon[2] and Lukas A. Mueller[1,2], (1)Cornell University, Ithaca, NY, (2)Boyce Thompson Institute, Ithaca, NY

Cassava is the staple food in many tropical countries for over 800 million people worldwide. Breeding activities and efforts to improve cassava crop, now depends on robust computational and integrated infrastructure for data management, sharing and analysis in a modern cassava breeding community. Cassavabase represents an ecosystem that promotes and enhances data sharing and communication within cassava breeding research communities in the world. The database stores crosses, phenotypes, sequencing data and houses tools for phenotyping, breeding management and molecular breeding resources.

Cassavabase currently has over 3000 experimental trials, about 11,000,000 phenotypes, and over 270,000 accessions of which 20,000 are genotyped from active communities of about 200 users across several breeding programs. This presentation will highlight advances made in cassavabase and its seamless integration with cassava breeding programs and research communities worldwide.

W160: **Cassava Genomics**

Advancement of Cassava Molecular Breeding in East-Asia

Yoshinori Utsumi, RIKEN Center for Sustainable Resource Science,, Yokohama, Japan

W161: **Cassava Genomics**

QTLs Mapping in Cassava for Whitefly Resistance

Adriana Bohorquez-Chaux, International Center for Tropical Agriculture, CIAT, Cali, Valle del Cauca, Colombia

W162: **Cassava Genomics**

Unraveling the Genetic Basis of Whitefly Resistance in Cassava

Danielle Garceau[1], Maria Irigoyen[1], Adriana Bohorquez-Chaux[2], L. Augusto Becerra Lopez-Lavalle[2] and Linda L. Walling[1], (1)University of California, Riverside, Riverside, CA, (2)International Center for Tropical Agriculture, CIAT, Cali, Valle del Cauca, Colombia

Outbreaks of superabundant whitefly populations throughout Eastern and Central Africa, beginning in the 1990s, have dramatically increased the pressures of whitefly feeding and virus transmission on their host, cassava. Considerable cassava yield losses due to this phenomenon evidence the need for both virus- and whitefly-resistant cassava lines for distribution to African small shareholder farmers. To achieve this, an international team, the African Cassava Whitefly Project (ACWP), was formed in 2014. Within the ACWP, a collaborative effort between CIAT and UCR aims to understand the genetic basis of whitefly resistance in cassava. Genetic mapping at CIAT using a resistant (ECU72) and susceptible (COL2246) cassava genotype identified multiple whitefly-resistance QTLs. To characterize genes in these regions, transcriptome profiling of ECU72 and COL2246 in response to whitefly infestation and in response to the defense hormones salicylic acid (SA) and jasmonic acid (JA) has been performed at UCR. From this analysis, we unexpectedly found that comparison of the SA- and JA-responsive transcriptomes of cassava to those of *Arabidopsis* revealed marked differences. Altogether, our analyses suggest that overall responsiveness to SA as well as expression of cell wall-related and hormone-responsive genes is likely associated with whitefly resistance in ECU72. Ultimately, our efforts to

characterize identified whitefly-resistance loci will inform their introduction into widely cultivated African lines for deployment by the mid 2020s.

W163: **Cattle/Sheep/Goat 1**

AdaptMap: Exploring Goat Diversity and Adaptation

Alessandra Stella, National Research Council of Italy, Lodi, Italy and the Adaptmap consortium

W164: **Cattle/Sheep/Goat 1**

Just Grazing the Surface Towards Understanding Meiotic Recombination in Livestock

Brenda M. Murdoch[1], Kimberly M. Davenport[1], Clare A. Gill[2] and Stephanie McKay[3], (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)Texas A&M University, College Station, TX, (3)University of Vermont, Burlington, VT

The production of viable gametes is integral to reproduction and critical for the sustainability of the livestock industry. Homologous recombination or crossovers (CO) are an extremely important process in gametogenesis that contributes to genetic variation and ensures proper chromosome segregation. It is clear from previous studies that at least one CO per chromosome arm is necessary to avoid mis-segregation, and improper placements of CO contribute to fetal loss and decreased fertility. Despite the importance of this process, we know very little about the factors that control and/or influence global meiotic recombination/CO in livestock. This research uses a direct cytological approach to quantify and characterize the number of CO in the spermatocytes of sheep and cattle. Testicular tissue samples were taken from sexually mature males and spermatocytes were spread and fixed on slides. Immunofluorescent staining was performed to identify the synaptonemal complexes (SYCP3) and CO (MLH1) of pachytene stage prophase cells. In total, we examined over 7,000 spermatocytes and approximately 450,000 recombination events in five breeds of sheep. We report that Targhee rams have significantly ($p < 2.00 \times 10^{-16}$) greater number of CO than rams of other breeds. Interestingly, cattle exhibit significantly fewer CO per spermatocyte compared to sheep, despite having the same number of chromosome arms. We found a positive correlation between the number of CO and the length of a chromosome, implying the mechanism(s) responsible for CO location preference is conserved in both species. This research provides essential data regarding meiotic recombination in spermatocytes and contributes important information towards understanding the mechanism that controls the composition of inherited chromosomes. These research will enhance the accuracy of genetic predictions and contribute towards the sustainability of the livestock industry.

W165: **Cattle/Sheep/Goat 1**

Bovine Haplotype Reference Consortium, 1000 Bulls and GTEx: International Projects to Advance Bovine Genomic Research

Robert D. Schnabel[1], Hans D. Daetwyler[2,3] and Amanda J. Chamberlain[3], (1)Division of Animal Sciences, Informatics Institute, Columbia, MO, (2)La Trobe University, Bundoora, Australia, (3)Agriculture Victoria, Bundoora, Australia

W166: **Cattle/Sheep/Goat 1**

Angus GS: Development of a Breed Specific Genotyping Platform to Facilitate Genomic Selection

Rich Tait Jr.[1], Stephen P. Miller[2], Xiao-Lin Wu[1], E. Barry Simpson[1], Ryan Ferretti[1], Mehdi Sargolzaei[3,4], Dan Moser[2] and Stewart Bauck[1], (1)Neogen GeneSeek Operations, Lincoln, NE, (2)Angus Genetics Inc., Saint Joseph, MO, (3)University of Guelph, Guelph, ON, Canada, (4)HiggsGene Solutions Inc., Guelph, ON, Canada

Development of Angus GS was undertaken through a collaboration between Neogen GeneSeek Operations and Angus Genetics, Inc. to accomplish near-term and future objectives of Angus breeders. These objectives were to: 1) deliver genomic selection relevant SNP to support the transition of the American Angus Association's routine genetic evaluation from a 2-step (calculate molecular breeding values (MBVs) from genomic information and include the MBV as a correlated trait in each genetic evaluation) to a 1-step genetic evaluation (adjustment of the additive genetic relationship matrix away from pedigree based expectations based on genomic relationships among genotyped individuals) (n = 37,955); 2) provide increased SNP density in genomic regions of interest from prior studies on important traits; 3) monitor for genetic traits previously reported in other breeds; 4) include SNPs purported to be important for novel traits, which need more direct genotype and phenotype data available to support those inferences; 5) augment with good genome coverage and informative SNP content to facilitate characterization of future inheritance within the Angus breed. Key trait complexes planned for future investigation specifically with Angus GS include: meat quality, fertility, and environmental adaptability. The Angus GS is a 50K SNP (n = 49,364) genotyping platform which is being used by Angus breeders around the world to better characterize the genetic merit of their seedstock animals through genomic selection methods.

W167: **Cattle/Sheep/Goat 1**

Molecular Characterization of a Congenital Overgrowth Syndrome Induced by Assisted Reproduction

Rocio Melissa Rivera, University of Missouri, Columbia, MO

W168: **Cattle/Sheep/Goat 1**

Identification of Long Non-Coding RNAs Linked to Parasite Host Immune Response in Sheep

Aroa Suarez-Vega[1], Samantha Dixon[1], Niel A Karrow[1], Paula Menzies[2], Bonnie A. Mallard[3], Andrew S. Peregrine[3], Emma Borkowski[3], Juan J. Arranz[4], Gwenola Tosser[5], Christophe Klopp[6] and Angela Cánovas[1], (1)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (2)Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada, (3)Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada, (4)Universidad de Leon, Leon, Spain, (5)INRA GenPhySE, CASTANET-TOLOSAN, France, (6)Genotoul Bioinformatics Platform, UR875, INRA, Auzeville, France

Haemonchus contortus is a blood-feeding gastrointestinal nematode considered as one of the most important parasites in small ruminant production worldwide. It is a highly pathogenic parasite that causes anemia, hypergastrianemia and alters abomasal secretion, resulting in reduced growth, a decrease in reproduction performance and elevated mortality in sheep and goat flocks. Recently, the emergence of anthelmintic-resistant *H. contortus* strains have encouraged research on the alternative parasite control approaches, such as the development of vaccines, optimized nutrition and genetic selection. The development of massive parallel RNA sequencing technology (RNA-Seq) has allowed the discovery of thousands of previously unannotated noncoding functional elements. Among these non-coding transcripts, there are increasing evidences that suggest that long non-coding RNAs (lncRNAs) play major roles in the interaction host-pathogen through regulating gene expression. The liver is an important participant in the host defense against gastro-intestinal nematodes, due to its role in inflammation, pathogen clearance and stress. To predict lncRNAs implicated in sheep immune response to parasites, we have used RNA-Seq samples from liver from 3 control sheep and 9 sheep naturally exposed to *H. contortus*. These latest sheep were previously categorized as high and (n=4) and low immune (n=5) responders based on their immunoglobulin G levels after vaccination using Hen Egg White (HEW) Lysozyme. Using the files from the alignment, we performed an assembly that allowed us to generate a reconstructed transcript model file. Then FEELnc software will be used for lncRNA prediction. Both the original DE gene list and the additional DE lncRNA-associated genes were combined to perform a comprehensive overrepresentation analysis. In this study, 17,529 lncRNAs were identified in the liver transcriptome. Of them, 13,153 were overlapping with a gene annotated in the *Ovis aries* genome (Oar_v3.1, release 94) and 4,376 were intergenic. In the preliminary differentially expression analysis, more than 1,000 lncRNAs were differentially expressed, in both comparisons, high immune responder sheep compared to controls and low immune responder sheep compared to controls. Among the enriched functional processes, we found Gene Ontology terms such as catabolic processes. Mammals are known to take advantage of amino acid catabolism to control pathogen invasion. The identification of an elevated number of differentially expressed lncRNA in the liver of sheep infested with *H. contortus* together with the detection of biological mechanisms related to the interaction host-pathogen suggests that lncRNAs could be modulating host response to parasites.

W1**69: Cattle/Sheep/Goat 1**

Incorporating Different Environmental and Phenotypic Information into Genomic Predictions for Dairy Cattle

Francesco Tiezzi, NCSU, raleigh, NC

Precision farming is becoming reality in livestock husbandry and is showing a tight connection with the use of big data. Appropriate statistical methods are needed to handle and make fruitful use of all the information available. Animal breeders are used to consider genomic data as major source of information, but environmental descriptors will turn useful in a precision livestock husbandry framework. In particular, environmental descriptors will be useful to breeders in presence of genotype by environment interactions.

Kernel regression has been successfully employed in the prediction of complex traits in crops and dairy cattle. On a previous study, we have shown that multiple regressors (genomic and environmental) can be successfully employed in prediction. Kernel regression was also able to account for the interaction between the genomic and environmental parts, which increased dramatically the number of (potential) predictors but also improved prediction accuracy when the right set of environmental descriptors was used.

A promising source of information seems to be the data coming from Fourier-transformed mid-infrared spectroscopy analysis performed on milk samples. Here, thousands of predictors are available, recorded at the cow or herd level. Literature reports several cases of successful prediction of milk quality and cow health status. Our group has also shown the use of phenotypic information as predicted with this methodology can improve the performance of breeding schemes. Furthermore, using predicted measures as correlated traits in selection seemed to improve the accuracy of prediction, with an advantage over the most popular genotyping strategies.

Future work should be focused on the use of this large amount of information in prediction, in the form of environmental descriptors. This could help benchmarking the herds, and the sires within the herds, in order to improve the allocation of genetic material over different farming conditions.

W1**70: Cattle/Sheep/Goat 1**

Pan-RNA Editing Analysis in Bovine Genome

Wentao Cai[1], Shengli Zhang[2] and Jiuzhou Song[1], (1)University of Maryland, College Park, MD, (2)Department of Animal Science, China Agricultural University, Beijing, China

RNA editing is an important phenomenon to modify nucleotides at specific sites of RNA molecules during post-transcription in many species, but its genomic landscape and characters have not been systematically explored in bovine. Here we deeply characterize the global RNA editing profiles from more than 50 tissues of dairy cattle and reveal noticeable spectrums of RNA editing in different tissues.A total of 1,117,831 RNA editing sites were identified using cluster strategy with high specificity (A-to-G rate >92%), high accuracy (>98%) and low FDR (< 3.3), which remarkably expand the number of the known RNA editing in bovine. We found most of editing sites were significantly enriched in specific non-LTR retrotransposons. The Bov-tAs, one of dispersed repeatBov-drived SINEsand like Alu in primates, likely form a dsRNA structure. By hybridizing, the ADAR oppositely oriented Bov-tAs. In the study, we first found ADARB1 (ADAR2), not ADAR, is predominant in determining global editing in bovine. The recurrence in editing patterns within tissues and across species deserve careful attentions since they are associated to tissue-specific and varieties of biological functions. Some editing sites were successfully validated in additional Chinese dairy cattle. Taken together, the widespread RNA editing clusters and their specificities in different tissues further provide evidence that RNA editing likely involve in regulatory mechanisms of gene expression and suggest that the RNA editome in bovine should be further explored in detail. Finally, the bovine editing data is available at: https://github.com/RNA-editing/Bovine.

W1**71: Cattle/Sheep/Goat 1**

USDA-NIFA Program Updates

Lakshmi Matukumalli, USDA-NIFA, Washington, DC

W1**72: Cattle/Sheep/Goat 2**

Selecting Microbial Drivers of the Ruminant Engine - New Forces for Change

Suzanne Rowe and Shannon Clarke, AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

W173: **Cattle/Sheep/Goat 2**
Maternal Influences on the Calf Rumen Microbiome and Subsequent Host Performance
Hannah C. Cunningham-Hollinger, Department of Animal Science, University of Wyoming, Laramie, WY

W174: **Cattle/Sheep/Goat 2**
A Look at the Galectin Gene Expression and Modulation in Ruminants
Mulumebet Worku[1], Sarah Adjei-Fremah[1], Kingsley Ekwemalor[1], Emmanuel K Asiamah[1], Eboghoye O. Eluka-Okoludoh[1], Bharath Kumar Mulakala[1,2] and Hamid D Ismail Mukhtar[1], (1)North Carolina Agricultural and Technical State University, Greensboro, NC, (2)NCAT, Greensboro, NC
Inflammatory diseases such as mastitis and gastrointestinal parasite infections cost the animal industry tens of billions of dollars. Concerns about food safety, drug residues and the rise in antibiotic and anthelmintic resistance is increasing. Galectins are an evolutionary conserved family of animal lectins important in disease and homeostasis. Our group has evaluated the expression, secretion and modulation of Galectins in blood from pasturing, non-pregnant and peripartutrient cattle, sheep and goats. New knowledge has been gained regarding systemic expression and modulation of Galectins and their ligands in relation to periparturient immune suppression, parasite egg shedding and gut health. This research is the first effort to define Galectin expression and its modulation in sheep, cow and goat blood. Galectin signatures may be useful in breeding of naturally resistant and resilient livestock, production of better diagnostics, preventives and targeted treatment for improved animal management. Novel Galectin based strategies to control animal diseases with consideration of genetics, stage of production, diet and infectious status can be designed for sustainable animal production, global food safety and security. Identification of evolutionary conserved Galectin signatures has translational implications for public health.

W175: **Cattle/Sheep/Goat 2**
Recent Advances in Medical Subject Heading (MeSH) Analysis
Gota Morota, Virginia Polytechnic Institute and State University, Blacksburg, VA

W176: **Cattle/Sheep/Goat 2**
Alternative Methods Improve the Accuracy of Genomic Prediction using the Information of a Pleiotropic Point Mutation in a Dairy Sheep Model
Claire Oget[1], Marc Teissier[1], Jean-Michel Astruc[2], Gwenola Tosser[1] and Rachel Rupp[1], (1)INRA GenPhySE, CASTANET-TOLOSAN, France, (2)Institut de l'Elevage, CASTANET-TOLOSAN, France
We evaluated genomic prediction accuracy using different methods to include the effect of a major gene pleiotropic point mutation. We studied as an example the *SOCS2* gene mutation influencing the host's inflammatory response to mammary infections, body growth and milk production. We performed genetic evaluation for Somatic Cell Scores as well as for four milk production and three mammary type traits. The data was composed of 14,729 genotyped animals (Illumina OvineSNP50 BeadChip). Pedigree-based BLUP was compared with single-step GBLUP (ssGBLUP) and four weighted ssGBLUP methods (WssGBLUP), which give weights to SNPs according to their effect on the evaluated trait. In order to include the *Socs2* genotype information in the evaluation, we previously imputed the point mutation for all the genotyped individuals. We then also tested pedigree and genomic-based Gene Content methods, a multiple-trait genomic model that includes *Socs2* genotypes as a trait, and all the ssGBLUP methods previously described when adding the point mutation as a SNP among markers. We computed accuracies of genomic estimated breeding values (GEBV).
We gained +12.4% in GEBV average accuracy between pedigree-based BLUP and ssGBLUP models and +3.4% between ssGBLUP and WssGBLUP methods. The inclusion of the *Socs2* genotypes with the genomic Gene Content method gave similar results as WssGBLUP. The addition of the *Socs2* SNP in ssGBLUP methods only brought an average gain of +0.23%. Weighted ssGBLUP methods were therefore more efficient for detecting SNPs with strong effects and for better predicting GEBV than ssGBLUP. When no chip data is available for genomic selection, the Gene Content method can efficiently account for partial genotyping information on major genes (+4.35%).

W177: **Cattle/Sheep/Goat 2**
Chromosome-Length Haplotigs for Cattle and Yak from Trio-Binning Assembly of an F1 Hybrid
Edward S. Rice[1], Sergey Koren[2], Michael P. Heaton[3], Jessica L. Petersen[4], Theodore S. Kalbfleisch[5], Timothy Hardy[6], Peter H Hackett[7], Brian L. Vander Ley[8], Derek M. Bickhart[9], Benjamin D. Rosen[10], Adam M. Phillippy[2] and Timothy P.L. Smith[11], (1)University of Nebraska/Department of Animal Science, Lincoln, NE, (2)NHGRI, NIH, Bethesda, MD, (3)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, (4)University of Nebraska, Lincoln, NE, (5)University of Louisville, Louisville, KY, (6)USYAKS, Chadron, NE, (7)Institute for Altitude Medicine, University of Colorado, Ridgway, CO, (8)University of Nebraska, Lincoln, Clay Center, NE, (9)Dairy Forage Research Center, USDA-ARS, Madison, WI, (10)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (11)USDA, ARS, USMARC, Clay Center, NE
Heterozygosity presents a challenge to the assembly of diploid genomes. Establishing phase of heterozygous segments between regions of homozygosity can reduce continuity, introduce switches between parental haplotypes, and add false duplications. The final product is a collapsed pseudo-haploid representation that does not correctly represent either haplotype. Trio-binning is a new approach that uses short reads from both parents to classify long reads from the offspring into bins according to maternal or paternal haplotype origin. The two are then assembled as separate haploid genomes. This process is helped rather than impeded by increasing heterozygosity. We used the high heterozygosity inherent in interspecies F1 hybrids to assemble both haplotypes of the offspring of a yak (*Bos grunniens*) cow and a Highland cattle (*Bos taurus*) bull with >1.2% heterozygosity. We sequenced both parents with high-accuracy short reads to find 21-mers unique to each parent, used these unique k-mers to sort long reads from the offspring into parent-of-origin bins, and then assembled the long reads from each haplotype independently. Both

the maternal (yak) and paternal (cattle) assemblies assemble half of the acrocentric chromosomes into single haplotigs, with the largest chromosomes (1 and 2) in single haplotigs for both haplotypes. These represent the first vertebrate chromosomes to be de novo assembled without gaps and completely phased into two distinct haplotypes. These results indicate that reducing the diploid assembly problem to a haploid one creates the most continuous and accurate assemblies currently possible.

W178: **Cattle/Sheep/Goat 2**
Editing for Genetic Improvement by Targeting Variants for Adaptive Traits
Tad S. Sonstegard, Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN

W179: **Cattle/Sheep/Goat 2**
Strong Resistance to Classical Scrapie in Goats
Mehmet Ulas Cinar[1,2], David A. Schneider[1,3], Daniel F. Waldron[4], Katherine I. O'Rourke[1,3] and **Stephen N. White**[1,3], (1)Washington State University, Pullman, WA, (2)Erciyes University, Kayseri, Turkey, (3)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (4)Texas Agrilife Research, San Angelo, TX
Scrapie is a fatal, infectious brain disease of sheep and goats. While sheep have benefited from strong genetic resistance that has contributed greatly to scrapie eradication programs, goats have not had widely recognized genetic resistance as a tool. Furthermore, scrapie in goats has become a bigger problem than scrapie in sheep for some countries. Early epidemiological studies showed that goats with prion protein substitutions of either S146 or K222 amino acids were present in herds with scrapie but absent in scrapie cases. To test the degree of resistance offered by these alleles, we used an oral scrapie challenge at birth to leverage the window for heightened gut permeability to protein. The control genotype consisted of goats homozygous for the most common haplotype, and all of these showed clinical scrapie at an average of 2 years. Two experimental genotypes each possessed one copy of the most common goat haplotype plus one copy of the allele of interest – either S146 or K222. No S146 or K222 heterozygotes have become scrapie positive by clinical signs, live animal tests, or post-mortem brain assays despite living longer (average 7 years) than the commercial lifespans of most goats. These data highlight the strong classical scrapie resistance provided separately by S146 and K222 in goats. Our study was the first published oral scrapie challenge study for K222 and the first challenge study of any kind for S146, and other investigators have used a variety of approaches to generate data that have also supported resistance. The U.S. Animal Health Association has recommended exploration of goat genetics for use in the U.S. scrapie eradication program, and the European Food Safety Authority released a detailed review article and statement supporting the use of S146 and K222 to enhance scrapie eradication.

W180: **Cattle/Sheep/Goat 2**
Reaction Norm Model of Body Temperature Response to Heat Stress in Beef Cattle
Raluca Mateescu, Department of Animal Sciences, University of Florida, Gainesville, FL

W181: **Cattle/Sheep/Goat 2**
Loci Associated with Cattle Conception Rates
Holly L. Neibergs, Department of Animal Sciences, Washington State University, Pullman, WA

W182: **Cattle/Sheep/Goat 2**
Multi-Tissue Transcriptome Enhances GWAS Interpretation and Genomic Prediction in Cattle
Lingzhao Fang, ARS-USDA, Beltsville, MD

W183: **Cattle/Sheep/Goat 2**
Selection of Cattle SNP Markers using Selective-Sequencing Experimental Design and Statistical Learning
Kiranmayee Bakshy, USDA ARS; Dairy Forage Research Center, Madison, WI, Robert D. Schnabel, Division of Animal Sciences, Informatics Institute, Columbia, MO and Derek M. Bickhart, Dairy Forage Research Center, USDA-ARS, Madison, WI

W184: **Cattle/Swine**
Genome-Enhanced Approaches to Improve High-Altitude Adaptability of Angus Cattle
Milt Thomas[1], Scott Speidel[1], Kelli Retallick[2], Stephen P. Miller[2], Dan Moser[2], Mark Enns[1] and Tim Holt[1], (1)Colorado State University, Fort Collins, CO, (2)Angus Genetics Inc., Saint Joseph, MO
Health of cattle is challenged in oxygen-limited mountainous production systems. An indicator trait for tolerance to this environment is pulmonary arterial pressure (PAP). This measurement is physiologically described as an indicator of hypoxia-induced pulmonary hypertension. Approximately 10,000 cattle in the Western United States are PAP-tested annually by veterinarians. Most of these cattle are seedstock-herd replacement yearling bulls and heifers. Yearling PAP has a moderate (0.2 to 0.4) heritability estimate and approximately 10% of the Angus cattle measured (> 49 mmHg) in a research herd (2,200 m) are considered high risk for suffering from pulmonary hypertension. This herd, the Colorado Beef Improvement Center (i.e., Rouse Angus herd), has calculated EPD for PAP for over 20 years and has shared these predictions with companies interested in progeny testing AI sires. The trait of PAP was determined to be very polygenic with GWAS; therefore, accuracy of EPD was improved with a single-step genetic evaluation using SNP data from BovineSNP50. Additionally, genetic parameters were estimated for PAP from Angus seedstock at moderate (1,200-1,620 m) and high elevations (>1,620 m). These PAP measures were genetically correlated (0.83) with similar heritability estimates of 0.29 and 0.34, respectively. Using this model, the American Angus Association will launch research EPD using single step GBLUP for PAP at high altitude (1 trait) in January 2019 using data from over 10,000 cattle. Whole-genome and RNA sequence of transcripts from cardiopulmonary tissues are also being used for SNP discovery to develop additional tools for genetic improvement of PAP in Angus cattle.

W185: **Cattle/Swine**
Benefits from Adoption of a New Reference Genome Assembly and Use of a Larger SNP Set in Genomic Predictions for US Dairy Cattle
Daniel J. Null[1], Paul M. VanRaden[1], Derek M. Bickhart[2], Jeffrey R. O'Connell[3], Benjamin D. Rosen[1] and **John B. Cole**[1], (1)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (2)Dairy Forage Research Center, USDA-ARS, Madison, WI, (3)University of Maryland School of Medicine, Baltimore, MD

W186: **Cattle/Swine**
New Bioinformatics Resources for Livestock Genome Assemblies and Annotations
Christine G. Elsik, Deborah A. Triant, Md Shamimuzzaman, Justin J. Le Tourneau and Amy T. Walsh, Division of Animal Sciences, University of Missouri, Columbia, MO
We report updates of the Bovine Genome Database Bovine (BGD; http://BovineGenome.org), and introduce two new livestock bioinformatics resources. First, BGD has been updated to support researchers as they transition their work from the UMD3.1.1 bovine genome assembly to the new bovine reference genome assembly, ARS-UCD1.2. The BGD data mining warehouse, BovineMine, provides database cross references between gene identifiers of the alternate assemblies, and the JBrowse/Apollo genome browser provides tracks for alternate assembly liftOver alignments. Second, we report a new resource, AgAnimalGenomes.org, which provides browsers and annotation tools for the genomes of bovine (ARS-UCD1.2), sheep (Oar_v4.0), goat (ARS1) and pig (Sscrofa11.1). Tools include JBrowse and Apollo for genome visualization and annotation, as well as a new Apollo plugin, called LSAA (locus specific alternate assembly), which allows users to view and submit potential genome assembly errors and structural variants. Finally, we present a new project to develop FAANGMine, a genomic data mining resource for domesticated animal species that are of interest to the FAANG (Functional Annotation of Animal Genomes) consortium. Species planned for incorporation into FAANGMine include bovine, chicken, goat, horse, pig, sheep, water buffalo, dog and cat. The incorporation of FAANG datasets will enable fine-grained data mining of functional elements in combination with gene annotations and additional biological data.

W187: **Cattle/Swine**
Artificial Intelligence and Feature Selection for Genomic Prediction
Cedric Gondro, Michigan State University, East Lansing, MI

W188: **Cattle/Swine**
Influence of Breed of Origin of Genomic Regions on Cow Productivity in a Crossbred Population
Clare A. Gill, Texas A&M University, College Station, TX

W189: **Cattle/Swine**
Comparative Analysis of DNA Methylome Provides Insights into Complex Traits and Epigenomic Evolution
George E. Liu, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

W190: **Cattle/Swine**
Integration of Structural and Functional Genomic Data to Identify Novel Haplotypes affecting Reproduction and its Pleiotropic Effects in Cattle
Angela Cánovas, University of California-Davis, Davis, CA

W191: **Cattle/Swine**
Genetic Improvement of Disease Resilience in Pigs
Jack C.M. Dekkers, Department of Animal Science, Iowa State University, Ames, IA

W192: **Challenges and Opportunities in Plant Science Data Management - an International Workshop**
Natural History Collections: Can You Breathe Life Back into Dead Dried Plants?
Matthew Clark, The Natural History Museum, London, United Kingdom
Nearly 3000 herbaria across the world are listed in Index Herbariorum and they are estimated to house more than 380 million specimens (Thiers 2017), adding animal and microbial samples to this there are probably over 1 billion natural history samples. The Natural History Museum (London) house over 80 million samples spanning from at least 1672, and to which we are still adding with new expeditions. To increase its worth, we are increasingly digitalising and making a virtual collection available online. It is possible to recover DNA from samples 100,000s of years old, and by adding sequencing data to our samples, it's possible answer questions of taxonomy but also see genetic change over time, and loss of diversity. Genome editing technology gives the possibility of reintroducing lost alleles into germplasm, but can we regain advantageous traits in this manner?
Thiers, B. 2017. “The World’s Herbaria 2016: A Summary Report Based on Data from Index Herbariorum.”

W193: **Challenges and Opportunities in Plant Science Data Management - an International Workshop**
Automatically Assigning GO Terms to Plant Gene Models
Dennis Psaroudakis, Kokulapalan Wimalanathan, Ha Vu, Parnal A. Joshi, Iddo Friedberg and Carolyn J. Lawrence-Dill, Iowa State University, Ames, IA
We constructed a high-coverage and reproducible functional annotation dataset for wheat based on Gene Ontology (GO) term assignments, covering 100% of the 107,891 high-confidence gene models in IWGSC’s RefSeq 1.1 genome with a median of 10 annotations per gene model.

To derive this annotation set, we used the GOMAP pipeline, which includes sequence similarity and protein domain presence methods as well as mixed-method pipelines that were developed for the Critical Assessment of Function Annotation (CAFA) challenge (https://biofunctionprediction.org/cafa/). Whereas application of the pipeline to maize was quality checked based on hand-curated functional annotations (Wimalanathan et al. 2018, DOI 10.1002/pld3.52), no such gold-standard exists for wheat so these predictions lack quality assessment. To build gold-standard datasets that would enable quality assessment for GOMAP annotations in wheat (as well as other genomes including rice, cotton, and soybean), we are building a platform to enable crowd-sourced review of computationally extracted gene-functional associations drawn from primary literature. The GOMAP pipeline is publicly available at https://github.com/Dill-PICL/GOMAP-singularity and the derived annotation datasets can be found at https://dill-picl.org/projects/gomap/gomap-datasets/.

W194: Challenges and Opportunities in Plant Science Data Management - an International Workshop
Adapting the ELIXIR Beacon Protocol to the Sharing of Plant Data

Gary Saunders, Elixir, Hinxton, United Kingdom

In this presentation I will describe the Beacon protocol and its planned extensions in order to serve the plants community. The Beacon protocol defines an open standard for genomics data discovery, developed by members of ELIXIR and the Global Alliance for Genomics & Health.
The initial version of the Beacon protocol had been developed to test the willingness and ability of international genome resources to share genomic data in a highly simplified context. The service was designed to accept specific queries in the form "Do you have any genomes with an 'A' at position 100735 on chromosome 3" and responds with "Yes" or "No."
The Beacon protocol has been designed to be:

- Simple: focus on robustness and easy implementation
- Federated: maintained by individual organizations and assembled into a network
- General-purpose: used to report on any variant collection
- Aggregative: provide a boolean (or quantitative) answer about the observation of a variant
- Privacy protecting: queries do not return information about single individuals

Recent and future versions of the Beacon protocol expand the original concept by providing a framework for querying other types of genome variation data (i.e. range queries and structural variants since v0.4) and also options for quantitative responses.
The primary Use Case for the Beacon protocol has been sensitive human data. However, ELIXIR had funded work to extend the Use Cases for the Beacon protocol into plant data. The Use Cases for human sensitive data and proprietary plant data overlap. A reference implementation of Beacon technology for plant data will be generated, extending it to integrate apricot variants data and thereafter extendable to plant data in general. Considered data standards for plant data will include Bioschemas.org and the BreedingAPI (www.brapi.org) alongside with their complementarity with GA4GH.

W195: Challenges and Opportunities in Plant Science Data Management - an International Workshop
Big Meets FAIR - Concepts, Implementations and Applications for Plant Genomics and Phenomics Data at the IPK Gatersleben

Matthias Lange, Daniel Arend, **Astrid Junker**, Patrick König, Nils Stein and Uwe Scholz, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Stadt Seeland, Germany

As the host of the largest European gene bank for cultivated plants, the IPK Gatersleben offers a rich basis for research of plant biodiversity at the structural and functional level. Using high-throughput technologies such as sequencing and phenotyping, millions of data points were collected and stored in the IPK research data infrastructure. We present methods, applications and new perspectives to harness these digital resources and demonstrate the impact of a FAIR-aware data management by the BRIDGE biodiversity informatics project for a molecular characterization of more than 22.000 barley accessions from IPK and partner Genebanks [1].
The data acquisition is featured by a universal laboratory information management system (LIMS), which stores structured, unstructured, binary data and metadata from the experiments in a central database and a scalable file storage backend [2]. The minimal metadata are adopted from the MIAPPE, the MCPD or DataCite standards. We support SQL as generic queries language by expose condensed, read-only views to the LIMS data schema. In order to support real-time enabled query response, we combined traditional relational tables and indexes, query caches, partitioning technology and in-memory columnar backend storage technologies. Canned queries that stored most common queries are implemented as RESTful API endpoints. In particular, the IPK implements parts of the Breeding API as well as proprietary APIs endpoints, e.g. to generate ISA-TAB/MIAPPE compliant exports or to feed web applications [3]. This enables a novel analytical access to the biodiversity of Genbanks. The BRIDGE web portal [4] is an exemplar that provides an explorative access to genotyping, phenotyping and data as well as well as comprehensive visualizations and exploration of analysis results of the mentioned BRIDGE project.
Raw and analysed data is published as DOI-citable data records, which are stored in the e!DAL-PGP repository [5]. For each dataset a landing page is generated that is annotated by a machine readable, JSON-LD encoded, DataCite and BioSchema compliant set of technical metadata. For some data sets, MIAPPE-compliant metadata is provided as ISA-TAB formatted files. This infrastructure implements the infrastructure-to-data (I2D) approach. It keeps the data at the point of origin, but ensures its FAIR aware publication.
The presented case studies show the enormous potential of FAIRification of IPK Gatersleben, but also the great demands on resources, organizational policies for the implementation of sustainable digital resources and the necessity of global authorities to homogenize and control individual actors. In particular, IPKs role as collaborator in national and international infrastructure programs, like de.NBI, ELIXIR, Emphasis, DivSeek, IPPN etc. have brought enormous progress in supplying a sustainable research data management toward to a global data re-use and value creation workflows.
[1] S. Milner et al. Genebank genomics highlights the diversity of a global barley collection. Nature Genetics, 2018.
[2] D. Arend et al. Data management experiences and best practices from the perspective of a plant research institute. Data integration in the life sciences: 10th international conference, Lisbon, Portugal, 2014.
[3] https://fair-ipk.ipk-gatersleben.de/
[4] https://bridge.ipk-gatersleben.de/
[5] D. Arend et al. PGP repository: a plant phenomics and genomics data publication infrastructure. Database, 2016

W196: **Challenges and Opportunities in Plant Science Data Management - an International Workshop**
Coordinating Communities Around Problems in Plant Phenomics
Jennifer Clarke, University of Nebraska Lincoln, Lincoln, NE
Challenges associated with plant phenomics for increasing agricultural productivity are complex – and interrelated. Addressing these challenges requires collaborations among scientists from diverse backgrounds because they will be able to pool a variety of skills and perspectives. Advances in plant phenomics will require these teams to work across numerous scientific communities, and to broaden the existing phenomics community by including new members from closely related fields. We will discuss some of these activities in the context of the Midwest Big Data Hub, the North American Plant Phenotyping Network, and the International Plant Phenotyping Network. As a result, we will also present what we have learned about multidisciplinary research. In particular, we highlight how these activities show plant phenomics to be a *convergent* field, i.e., one based on a combining expertise from diverse fields such as life sciences, physical sciences, engineering.

W197: **Challenges and Opportunities in Plant Science Data Management - an International Workshop**
The AgBioData Consortium: Working Together to Work Smarter!
Lisa Harper, United States Department of Agriculture, Ames, IA
Scientists are now generating and analyzing larger, and more complex genomic, genetic and breeding (GGB) datasets. The value of these rich data sets significantly increases when they are curated, organized, annotated, integrated with other data, and shared, usually in the framework of an online relational database. To keep up with the tidal wave of data, curators have to change the way they work. The AgBioData Consortium is a group of scientists from 30 different plant and animal genetic, genomic and breeding (GGB) databases. Our mission is to work together to ensure standards and best practices for acquisition, interoperability, display and retrieval of genomic, genetic and breeding data, and to promote the FAIR data principles (https://www.force11.org/group/fairgroup/fairprinciples). After two years of collaboration, we have published an extensive article on standards and best practices for GBB databases in 7 areas: biocuration, ontologies, metadata and persistence, database platforms, programmatic (machine) access to data, communication and sustainability (https://doi.org/10.1093/database/bay088). This talk will describe the process and results, our next steps and our thoughts on what is working, and what is not.

W198: **Citrus Genome**
Citrus Genome Database Resources for *Citrus* Genomics, Genetics, and Breeding Research
Jodi L. Humann[1], Morgan Frank[1], Sook Jung[1], Chun-Huai Cheng[2], Taein Lee[2], Heidi Hough[1], Katheryn Buble[3], Ping Zheng[1], Kristin Scott[1], Julia Piaskowski[1], Mirella Flores-Gonzalez[4], Surya Saha[4], Lukas A. Mueller[4], Jill L. Wegrzyn[5], Margaret Staton[6], Stephen P. Ficklin[7], Albert G. Abbott[8] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Washington State University, Pulmman, WA, (4)Boyce Thompson Institute, Ithaca, NY, (5)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (6)University of Tennessee, Knoxville, Knoxville, TN, (7)Dept of Horticulture, Washington State University, Pullman, WA, (8)Forest Health Research and Education Center, University of Kentucky, Lexington, KY
The Citrus Genome Database (CGD, www.citrusgenomedb.org) is a curated, integrated genomics, genetics and breeding database resource for basic, translational and applied citrus research. CGD currently contains 6 annotated genome sequences; an annotated and mapped *Citrus sinensis* reference transcriptome analyzed from published RNA-Seq and EST data; 70 genetic maps; 48,154 markers; 579 QTL; 1426 germplasm; metabolic pathways for *C. clementina* and *C. sinensis* genomes; 23,070 GRIN phenotype measurements; and synteny data for six citrus genomes with links to genes, mRNA, orthologs and function. Tools include the genome browser JBrowse, Synteny Viewer, MapViewer, CitrusCyc, 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. Genes and transcripts, maps, markers, germplasm, QTL, sequences and publications can be queried through the search interfaces on CGD with results available to download. 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. In addition to the *Citrus* genomes, the complete genomes of *Ca*. Liberibacter sp. and *Liberibacter crescens* are available to explore in BLAST, JBrowse, and the Synteny Viewer. CGD is built using the Tripal database platform and is supported by USDA-NRSP10, NSF-PGRP, USDA-SCRI and US Land Grant Universities.

W199: **Citrus Genome**
Frequency and Characteristics of Large Apparent Deletions in Citrus Germplasm
Mikeal L. Roose, University of California, Riverside, CA, Yoko Eck, University of California Riverside, Riverside, CA and Tulsi Dey, Kalyani Mahavidyalaya, Kalyani, India
Many citrus cultivars originated by selection of natural or induced somatic mutations. The nature of the mutational events involved has remained unclear and in general it has been difficult to identify molecular markers that distinguish such cultivars. Using Affymetrix SNP arrays that genotype about 51000 SNPs we studied patterns of loss-of-heterozygosity (LOH) in about 900 citrus cultivars. LOH can be reliably detected between cultivars within groups that have diverged only by mutation such as sweet oranges, lemons, grapefruit, and satsumas. We found more than 57 unique deletions ranging in size from about 50 kb to more than 10 Mb and which converted 5 to 817 contiguous heterozygous markers to apparently hemizygous state. Among the cultivars with such mutations are Rio Red grapefruit, Midknight Valencia, Miyamoto satsuma, Cukurova navel, Lapithiotiki lemon, several blood oranges, Pera sweet orange, and Arrufatina clementine. Some groups of cultivars documented as having shared ancestry also shared deletion mutations including haplotype on the remaining chromosome. Shared heterozygous deletions also suggest shared ancestry among some cultivars not previously proposed as being closely related.

W200: **Citrus Genome**
Citrus Breeding 2.0: A Novel Approach Integrating Deciphered Parentage and Genomics-Assisted Selection

Tokurou Shimizu[1], Mai Minamikawa[2], Keisuke Nonaka[1], Terutaka Yoshioka[1], Eli Kaminuma[3] and Hiroyoshi Iwata[2], (1)Institute of Fruit Tree and Tea Science, NARO, Shizuoka, Japan, (2)The University of Tokyo, Tokyo, Japan, (3)Genome Informatics Laboratory, NIG, ROIS, Mishima, Japan
Current citrus breeding programs have three major objectives: enlarging diversity of fruit characteristics to satisfy broad commercial needs, improving overall fruit quality, and releasing new varieties in a short time frame. However, these objectives are often in conflict with each other owing to the limited available orchard space, long juvenile period, and small chance of selecting a seedling of premium quality. This constraint, or breeding trilemma, is hard to avoid when using conventional breeding systems. Of these objectives, parent selection is the key to expanding diversity. Recent parentage analysis of citrus varieties revealed the involvement of various unpredicted varieties for expanding citrus diversity. Furthermore, it was suggested that one or several generations would be sufficient for the selection of a promising variety. Although the use of a non-elite indigenous variety for breeding that was referred from parentage analysis will contribute to extending the diversity of various fruit traits, it will reduce the opportunity of selecting a promising offspring from these crosses. 'Citrus breeding 2.0' is a novel approach for avoiding the trilemma of citrus breeding when using non-elite citrus varieties for breeding by integrating genomic selection (GS) and genome-wide association study (GWAS) techniques. This approach increases the chances of selecting more promising seedlings through composite selection by GS and GWAS and contributes to fast-breeding by achieving selection from a single cross, thereby minimizing the total cost of breeding. This approach will also be valuable for rebreeding and improving existing varieties via cross-breeding.

W201: Citrus Genome

Genomic Selection for Fruit Quality in Citrus

Itamara B. Goes[1], Marcos D.V. Resende[2], Maiara Curtolo[3,4], **Mariangela Cristofani-Yaly**[3], Marcos A. Machado[3] and Aluízio Borém[1], (1)Department of Plant Science, Viçosa, MG, Brazil, (2)Embrapa, Vicosa, Brazil, (3)Sylvio Moreira Citrus Center, Agronomy Institute, Cordeiropolis, Brazil, (4)State University of Campinas - Unicamp, Campinas, Brazil
Genome Wide Selection (GWS) is a selection based on markers and has been proposed as an alternative to increase the efficiency of breeding programs. This work was carried out with the objective of evaluating the applicability of GWS in the genetic improvement of Citrus and to evaluate its efficiency when compared to selection based on phenotypes. A Citrus population from the cross between Pera sweet orange (*Citrus sinensis* Osbeck) and Murcott tangor (*C. sinensis* Osbeck x *C. reticulata* Blanco) was evaluated for ten quantitative characteristics related to fruit quality, and genotyped with 16,618 SNPs. The genomic genetic values of the individuals were predicted using the Bayesian RKHS method. The predictive capacity (r and g) of the GWS ranged from 0.33 to 0.48 and the accuracy of the prediction ranged from 0.33 to 0.66 for the additive-dominant model. The fraction of heritability captured by the markers (ω) was greater than 48%. The accuracy estimates based on the phenotypes were higher than those obtained through the GWS, however, it was possible to observe the superiority of the GWS, in terms of genetic gain per unit of time. Therefore, GWS presents potential to be used as a tool in the genetic improvement for fruit quality in Citrus, due to the possibility of reduction of the selective cycle.

W202: Citrus Genome

A MADS-Box Transcription Factor Involved in the Regulation of Earliness during the Ripening of Citrus Fruits

Javier Terol[1], Maria Jose Nueda[2], Daniel Ventimilla[1], Francisco R Tadeo[1] and Manuel Talon[1], (1)Centre de Genòmica, Institut Valencià de Investigacions Agràries, IVIA, Moncada, Valencia, Spain, (2)Facultad de Ciencias, Universidad de Alicante, Alicante, Spain
Background: Harvest time is a relevant economic trait in citrus, and selection of varieties with different fruit maturity periods has a remarkable impact in the market share. Generation of early- and late-maturing varieties is an important target for citrus breeders, therefore, generation of knowledge regarding the genetic mechanisms controlling the ripening process and causing the early and late phenotypes is crucial. In this work we analyze the evolution of the transcriptome during fruit ripening in 3 sport mutations derived from the Fina clementine (*Citrus clementina*) mandarin: Clemenules (CLE), Arrufatina (ARR) and Hernandina (HER) that differ in their harvesting periods. CLE is considered a mid-season variety while ARR and HER are early- and late-ripening mutants, respectively.
Results: We used RNA-Seq technology to carry out a time course analysis of the transcriptome of the 3 mutations along the ripening period. The results indicated that in these mutants, earliness and lateness during fruit ripening correlated with the advancement or delay in the expression of a set of genes that may be implicated in the maturation process. A detailed analysis of the transcription factors known to be involved in the regulation of fruit ripening identified a member of the MADS box family whose expression was lower in ARR, the early-ripening mutant, and higher in HER, the late-ripening mutant. The pattern of expression of this gene during the maturation period was basically contrary to those of the ethylene biosynthetic genes, SAM and ACC synthases and ACC oxidase. The gene was present in hemizygous dose in the early-ripening mutant
Conclusions: Our analysis provides new clues about the genetic control of fruit ripening in citrus and allowed the identification of a transcription factor that appears to function as a negative regulator of the maturation process.

W203: Citrus Genome

The Transcription Factor FcWRKY40 of *Fortunella crassifolia* Functions Positively in Salt Tolerance through Modulation of Ion Homeostasis and Proline Biosynthesis by Directly Regulating SOS2 and P5CS1

Ji-Hong Liu, Huazhong Agricultural University, Wuhan, China
WRKY proteins represent one of the largest families of transcription factors and play a pivotal role in plant response to environmental stresses. Although several WRKYs have been functionally characterized, regulatory role and physiological function of most WRKYs, particularly those from non-model plants, remain largely unknown. In this study, we report that overexpression of *FcWRKY40* from kumquat (*Fortunella crassifolia*) led to enhanced salt tolerance in both tobacco (*Nicotiana nudicaulis*) and lemon (*Citrus lemon*). The transgenic lemon lines displayed substantially lower Na^+contents and higher proline levels in comparison with wild type (WT). Furthermore, treatment of the transgenic lemon plants with 24-*epi*-brassinolide reduced endogenous proline levels and thus compromised salt tolerance. In parallel, mRNA abundances of *SOS2*(*Salt Overly Sensitive 2*) and *P5CS1*(*D-1-pyrroline-5-carboxylate synthetase 1*), two genes involved in ion homeostasis and proline

synthesis, respectively, were dramatically enhanced in transgenic lemon. Protein-DNA interaction assays revealed that FcWRKY40 binds directly to W-box elements in the promoters of *FcSOS2*and *FcP5CS1*and functions as a transcriptional activator. Moreover, *FcWRKY40*was up-regulated by both ABA and salt, while salt-induced up-regulation was ABA-dependent. FcABF2 (ABA-responsive element binding factor 2) was shown toregulate*FcWRKY40* by interacting with the ABRE (Abscisic acid response element) element in the promoter. Collectively, these results demonstrate that FcWRKY40 plays a positive role in salt tolerance, which may be ascribed to modulation of ion homeostasis and proline biosynthesis by regulating *SOS2*and *P5CS1*. Our findings reveal a transcriptional pathway composed of ABF2-WRKY40-SOS2/P5CS1 that orchestrates salt stress response in plants.

W204: Citrus Genome

Developing Transgenic Citrus using Tissue-Specific Promoters from Citrus Small Cyclic Amphipathic Peptides (SCAmpPs) Genes

Ed Stover[1], Shujian Zhang[2], Joseph Krystel[3], Guixia Hao[1], William Belknap[4], Goutam Gupta[5], James Thomson[6] and Qingchun Shi[1], (1)U. S. Horticultural Research Laboratory, USDA-ARS, Fort Pierce, FL, (2)New Mexico Consortium, Los Alamos, NM, (3)USDA/ARS, Fort Pierce, FL, (4)USDA/ARS, Eugene, OR, (5)NMC Inc., Los Alamos, NM, (6)USDA/ARS, Albany, CA

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.

W205: Climate Change and ICRCGC 1

Global Collaboration for Addressing Global Climate Change

Chittaranjan Kole, Department of Atomic Energy, New Delhi, India

W206: Climate Change and ICRCGC 1

Moving Wheat Yield Frontiers and Climate Resilience through Conventional and Genomic Assisted Breeding

Ravi Prakash Singh[1,2], Fnu Philomin Juliana[3], Jesse Poland[4], Suchismita Mondal[5], Govindan Velu[1], Leonardo Crespo_Herreara[1], Julio Huerta-Espino[6] and Uttam Kumar[7], (1)CIMMYT, Texcoco, Mexico, (2)CIMMYT, Mexico City, Mexico, (3)CIMMYT, El Batan, Mexico, (4)Department of Plant Pathology, Kansas State University, Manhattan, KS, (5)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (6)INIFAP, Chapingo, Mexico, (7)Borlaug Institute for South Asia, Ludhiana, India

Climate change is already causing wheat yield fluctuations in various CIMMYT target countries of Africa, Middle East, Africa and Latin America where over 70% varieties released during the last decade trace to CGIAR germplasm, either released directly or used as a parent. A slight increase of 1^0C night temperature during critical growth periods of wheat crop results in 8-10% grain yield reductions in various spring wheat-growing areas. Because plants require more water under heat stress, breeding for climate resilience needs a simultaneous improvement for grain yield potential, water-use efficiency/drought tolerance and heat tolerance. All these traits are quantitative and under polygenic control. Therefore, a large-scale breeding, maintaining high genetic diversity in breeding germplasm, utilizing conventional and modern strategies and tools used by CIMMYT are enabling the genetic yield gains together with tolerance to drought and heat, and other necessary traits. Diverse field sites in Mexico and Kenya permit two generations per year shuttle breeding for selection, and managed phenotyping for grain yield under a range of high yield and stress environments. Combined with international testing of elite germplasm through international trials and nurseries by hundreds of partners in numerous wheat growing countries further assists in rapid identification of superior, climate resilient germplasm as well as locally adapted varieties by the partners. In an attempt to further increase genetic gain, we have genotyped >45,000 wheat breeding lines that were included in the 1st year yield phenotying during the past 5 years. Selected subsets of >5,000 lines (>1,000 each year) were phenotyped for grain yield extensively in Mexico under 5-6 managed environments each year, and further subsets of 2,700 lines (540 each year) phenotyped in South Asia at 5-6 sites. The genotypic and phenotypic meta-data are currently being used in conducting meta-GWAS (Genome Wide Association Studies) to identify consistent genomic regions and linked markers contributing to grain yield, heat and drought tolerance, agronomic, disease resistance and quality traits. These studies could aid in determining parental combinations that have higher probabilities of giving superior progenies. We also used GEBV (Genomic Estimated Breeding Value) determined through various models however so far the predicted values cannot be used reliably in selecting superior performing lines but can be used in discarding 25-30% lower performing lines with high confidence. The superior wheat germplasm distributed annually by CIMMYT continues to be released as numerous new popular varieties; which aid in enhancing productivity and profitability in targeted regions.

W207: Climate Change and ICRCGC 1

Accelerated Development of Climate Resilient Maize for the Tropics through Mainstreamed use of Genomic Tools and Precision Phenotyping

Michael S. Olsen[1], Manje Gowda[2], Yoseph Beyene Aydagn[3], Dan Makumbi[2], Jill Cairns[4], Mainassara Zaman-Abdou[4], Pervez H. Zaidi[5], Sudha Nair[6], Xuecai Zhang[7], Félix San Vincente[8], Elizabeth Jones[9], Star Yanxin Gao[9], Kelly Robbins[10], Paulino Perez-Rodriguez[8], Gregorio Alvarado[11], José Crossa[12], Kate A. Dreher[13], Claudio Ayala[8] and Prasanna Boddupalli[3], (1)CIMMYT, MINNEAPOLIS, MN, Kenya, (2)CIMMYT, Nairobi, Kenya, (3)CIMMYT, NAIROBI, Kenya, (4)CIMMYT, Harare, Zimbabwe, (5)CIMMYT-Asia, Hyderabad, India, (6)CIMMYT, Hyderabad, India, (7)CIMMYT, Mexico City, Mexico, (8)CIMMYT, Texcoco, Mexico, (9)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (10)Cornell University, Ithaca, NY, (11)International Maize and Wheat Improvement Center, Texcoco, Edo. de Mexico, Mexico, (12)International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico, (13)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Climate change is expected to increase the severity and variability of abiotic stress imposition on crops. Climate change models also project that insect and pathogen populations will move into new agro-ecological zones. The rapid development and deployment of improved maize varieties with tolerance to multiple abiotic and biotic stresses is critical for building resilience and adaptive capacity of farming communities as climate changes. CIMMYT is focused on developing maize varieties with tolerance to drought, heat, and low soil nitrogen together with resistance to pests and diseases by incorporating modern breeding tools and innovations. Optimization and scaling of managed stress sites has focused recently on heat stress (vapor pressure deficit stress) and the combination of heat + drought stress. Improvement in phenotyping precision for abiotic stress traits has enabled consistent breeding progress, and is the basis for ongoing implementation of genomic selection strategies to improve grain yield in high stress conditions. Results of genomic selection applied to large scale breeding programs to predict the best performing lines under optimum and drought conditions in Africa and Latin America will be presented demonstrating progress toward reduction of breeding cycle time. A brief review of progress in deploying forward breeding strategies and targeted marker-assisted backcrossing to rapidly increase allele frequency of important genes for different diseases in response to emerging threats will also be presented. By utilizing precision phenotyping in combination with genomic tools an integrated breeding pipeline focused on developing maize varieties needed for future climate and pest/pathogen fluctuations is enabling greater responsiveness to the challenges of maize agriculture which continue to emerge.

W208: Climate Change and ICRCGC 1

Genomics-Assisted Breeding for Development of Climate-Smart Rice

Jauhar Ali, International Rice Research Institute, Metro Manila, Philippines

Global climatic alterations (GCA) such as drought, flood, salinity, submergence, and high/low temperature pose serious threats to crop productivity across the world. Developing and deploying climate-smart rice (inbred and hybrid) varieties (CSRVs) resilient to biotic and abiotic stresses could be as a viable mitigation option to GCA.

Genomics-assisted breeding (GAB) was carried out by integrating advanced genomics tools, and an improved green super rice (GSR) breeding strategy (GSR-BT) led to the development of CSRV's. Under the GSR program at the International Rice Research Institute (IRRI), first, BC_1F_2 populations derived from Huanghuazhan (HHZ), Weed Tolerant Rice 1 (WTR1), and TME80518 (recipient parents) and 16 donors were developed. Later, these introgression lines (ILs) underwent simultaneous screenings over three rounds for different abiotic and biotic stresses as well as evaluation under normal irrigated conditions, resulting in the identification of trait-specific ILs. 1333 (HHZ-ILs) + 2232 (WTR1-ILs) + 1408 (TME80518-ILs) ILs were developed, which were further used either for varietal improvement or as parental lines for the development of pyramided lines (PDLs). Stacking traits/genes derived from the generated trait-specific ILs was carried out by a designed QTL pyramiding (DQ) approach. It resulted in the development of 1280 (HHZ-PDLs) + 850 (WTR1-PDLs) PDLs that showed significantly superior performance over the tolerant checks. For a successful GAB approach, requires innovative cross-tolerance screening and skillful selection techniques. It resulted in the development of 240 CSRVs with multiple abiotic and biotic stress tolerance that were distributed to Asian and African countries without compromising on grain yield and quality.

These efforts led to release of a total of 24 IRRI-bred CSRVs, and nomination of 85 cultivars into national cooperative yield trials derived from three recipient parents within a short span of seven years. Currently, the released CSRVs cover 2.7 million ha on a cumulative basis in Asia and Africa.

W209: Climate Change and ICRCGC 1

Exploiting Genomic Resources of Non-Crop Models for Development of Climate Resilient Rice

Niranjan Baisakh, School of Plant, Environmental, and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA

Rice is known as a profligate user of water, which uses nearly 40% of the total irrigation water used for crops and about 30% of the total fresh water. It is estimated that ~1,500 liters of water is required to produce 1 kilogram of paddy rice. Unfortunately, water resource is becoming quantitatively limited and/or qualitatively deteriorated under the perpetual climatic uncertainties. Thus, the focus of rice breeding has encompassed recently on the development of varieties to cope with climate-induced environmental stresses, such as drought and salinity. Conventional breeding has resulted in limited success with a few reports of improved varieties that are able to complete life cycle under drought or salinity. This is due in part to the complex, multigenic inheritance of the abiotic stress resistance traits and in part to the non-availability of superior stress-resistance alleles in the primary gene pools. Wild relatives in the secondary/tertiary gene pools of cultivated crops, on the other hand, have the ability to adapt under the environmental stressors through physiological and biochemical adjustments at the plant, tissue, cellular and molecular levels. *Spartina alterniflora* and *Portersia coarctata* are two halophyte grasses that are being studied as non-crop models for exploitation of their genomic resources to improve abiotic stress resistance in rice. To this end, we have developed comprehensive transcriptome maps of these two monocot halophytes. Further, the transcriptome resources have been exploited through genetic engineering to validate the functional superiority of their stress-resistance alleles in rice. Allelic polymorphisms are being investigated as targets for manipulation via genetic editing/genome engineering to develop non-transgenic improved rice varieties. The results from our studies thus far will be discussed during the presentation.

W210: Climate Change and ICRCGC 1

Genomic Prediction in the Face of Increasing Seasonal Variability: Predicting Barley Yields on the UK Recommended List
Keith A. Gardner, NIAB, Cambridge, United Kingdom, Ian J Mackay, IMplant Consultancy, Chelmsford, United Kingdom, Bill Thomas, The James Hutton Institute, Dundee, United Kingdom and IMPROMALT Consortium
Genomic prediction is a powerful tool for accelerating genetic improvement in modern breeding programmes. However, increasing climate variability has the potential to reduce this predictive ability. Here, we investigated the genomic prediction of barley yield using 27 years of National List and Recommended List trials at multiple sites across the UK. Overall genomic predictions were good. Using the optimum training set of 6 previous years of yield data, the average correlation between genomically predicted and measured yield of newly released lines of spring barley was 0.40, for winter barley 0.30. This compares well to a 0.57 phenotypic correlation between the first (NL1) and second (NL2) years of national list trials of new varieties of spring barley. However, there was high variability in prediction ability between years in the data set; for spring barley c 50% of years had correlations of ≥0.5 but 20% of years had a correlation of ≤0. We investigated the underlying causes of this variability, including differences in mean yield between years, in within-year yield variability and in genetic relatedness within the training and test sets. These phenomena all varied but did not account for the inter-annual variation either singly or in combination. In fact the single strongest explanatory variable of genomic prediction ability (correlation 0.75) was the phenotypic correlation between NL2 lines and the previous year's NL1 data. In other words, the genomic prediction is largely constrained by the phenotypic correlation of lines between years. Adding climate data into this scenario revealed a clear pattern; the years where predictive ability was low were all years of extreme climatic conditions, although the particular kind of extreme conditions differed between these years. The implications of these results are discussed in the context of climate change scenarios, together with potential ways to ameliorate the problem.

W211: **Climate Change and ICRCGC 1**
GM and Gene Edited Sorghum: Larger Grain, Higher Protein and Altered Plant Development
Ian D Godwin, School of Agriculture and Food Sciences, The University of Queensland, Brisbane, Australia
Sorghum is a major staple cereal with over 500 million people worldwide dependent on it every day in sub-Saharan Africa and India. We have used genetic engineering and CRISPR/Cas9 gene editing approaches in parallel to improve key quality parameters of sorghum: grain size, protein content and protein quality. We have targetted the kafirin seed storage proteins, which affect digestibility in monogastric animals. Using a sorghum inbred line, Tx430, we have manipulated the kafirin seed storage proteins, signalling proteins involved in grain size, and foldase enzymes which are key components in packaging the endosperm protein:starch matrix. This has led to altered grain size, digestibility and end-use processing qualities. Selected lined were grown in Australia's first GM sorghum field trial in 2018, with multi-environment trials currently underway. Sorghums produced include lines with increased grain number, increased grain size, and protein contents of 15-16% compared to the parent at 11% protein. A number of transgenic lines have both larger grain and more grain compared to Tx430, with thousand kernel weights up to 75% higher than the check lines. Reduced grain size is a common outcome resulting from heat/drought stress. We have also produced plants with altered panicle morphology, root and tiller morphology and productivity traits. We have optimised CRISPR/Cas9 gene editing to target key genes involved in sorghum grain quality, and will field trial these outcomes as soon as OGTR legislation becomes clear by early 2019.

W212: **Climate Change and ICRCGC 2**
Genetics, Physiology, and Applications of Heat Tolerance in Common Bean
Timothy Porch, USDA-ARS Tropical Agriculture Research Station, Mayaguez, PR
The sensitivity of common bean (*Phaseolus vulgaris*) to high temperatures (> 25^{o}C), primarily evident in failed reproductive development, threatens reductions in yield and production area due to a warming climate. Continued advances in the genetic improvement of heat tolerance in common bean is critical for maintaining current production areas and for increasing yields of this crop critical for food security. This work evaluates the physiological response of bean in contrast to its abiotic stress tolerant sister species, tepary bean (*Phaseolus acutifolius*). The genetics of heat response in bean is presented based on evaluations of diversity panels, and of recombinant inbred line populations genotyped using genotyping-by-sequencing (GBS), and then evaluated using GWAS or QTL mapping. Traits associated with vegetative and reproductive development, including pollen shed, and high-throughput methods for measuring NDVI, canopy temperature and canopy height, were evaluated from field experiments conducted under high ambient temperatures. Novel approaches developed are being applied to the collaborative breeding of resilient Andean beans for Sub-Saharan Africa.

W213: **Climate Change and ICRCGC 2**
Field Crops Breeding for Resistance to Biotic and Abiotic Stresses at ICARDA: Achievements and Prospects
Michael Baum[1], Wuletaw Tadesse[1], Filippo Maria Bassi[2], Miguel Sanchez-Garcia[3], Ramesh Verma[1], Andrea Visioni[1], Shiv Kumar Agrawal[1], Mustapha El-Bouhssini[1], Seid Ahmed[1], Sripada Udupa[4] and Aladdin Hamwieh[5], (1)ICARDA, Rabat, Morocco, (2)International Center for Agricutural Research in the Dry Areas, Rabat, Morocco, (3)ICARDA, Al Irfane, 10112 Rabat,, Morocco, (4)ICARDA, Rabat,, Morocco, (5)ICARDA, Cairo, Egypt
Crop production is highly affected by biotic and abiotic stresses at global level in general and in the Central and West Asia and North Africa (CWANA) and Sub Saharan Africa (SSA) regions in particular. Associated to climate change, heat and drought stresses are increasingly important resulting in reduction of photosynthesis, pollen viability, grain number and weight, and hence lowering yield and quality of major cereals and legumes crops. The crop breeding program at ICARDA uses conventional and molecular approaches such as FIGS, mega environments, shuttle breeding, doubled haploids, marker assisted selection, speed breeding and key location phenotyping to identify sources of resistance, develop elite genotypes with high yield potential and resistance to the major biotic and abiotic stresses. ICARDA distribute yearly more than 1000 of such genotypes to its partners through international nurseries. In the last 5 years alone, a total of more than 100 heat and drought tolerant varieties of ICARDA origin have been released by NARS in the CWANA and SSA regions. Marker Trait Association (MTA) have been identified for different crops. MTA studies for heat stress in wheat at Wad Medani, Sudan has led to the identification of *Ex_c12812_20324622* (4A) and *wsnp_Ex_c2526_4715978* (5A) SNP marker which are significantly associated with yield under heat stress. Wheat genotypes carrying the cytosine base at the *wsnp_Ex_c12812_20324622* and *wsnp_Ex_c2526_4715978* markers out-yielded the ones

carrying the alternative bases by 15% while genotypes carrying the cytosine base at only one of the two markers increased their yield by 7.9-10%, suggesting the importance of using these markers for MAS in breeding programs to increase yield under heat stress.
Key words: Crop breeding, Climate change, drought, heat

W214: Climate Change and ICRCGC 2

Evolution in a Changing Environment: The Genetic Architecture of Adaptation Outside Centers of Domestication

Roberto Papa, Marche Polytechnic University, Ancona, Italy

Domesticated species spread outside their native range and faced new environmental challenges. This evolutionary scenario makes crops excellent models to deeply understand adaptation to new ecological conditions. The BEAN_ADAPT project, funded through the 2ndERA-CAPS call, ERA-NET for Coordinating Action in Plant Sciences, aims to dissect out the genetic basis and phenotypic consequences of the adaptation to new environments of the common bean (Phaseolus vulgaris L.), through the study of the introduction, from the centres of domestication in the Americas, and the expansion through Europe, as a recent and historically well-defined event of rapid adaptation.
We re-sequenced 220 American and European common bean accessions at an average 7X coverage per accessions. This sample represents a subsample of a nested collection of 500 common bean accessions which was phenotyped both in controlled conditions and in multi-field trials at two different latitudes (Northern Germany and Southern Italy).
Here we present the results of genomic data based on population genomics approaches and of integration of genomic, transcriptomic, metabolomic and phenotypic data to identify genes and/or genomic regions associated to important traits related to adaptation of P. vulgaris to the European agroecosystems.

W215: Climate Change and ICRCGC 2

Quantifying Transpiration from Seedling to Harvest: Revisiting Old Methods with New Technology

Andreas Stahl, Justus Liebig University, Giessen, Germany and Rod Snowdon, Department of Plant Breeding, Justus Liebig University, Giessen, Germany

Agricultural plant production in major production areas repeatedly suffers from periods of low water availability caused by climatic changes. Although there is a wide consensus that plant breeding for more adapted varieties is a sustainable strategy, the inheritance of responsible traits is less clear. Although genomic tools have developed tremendously in recent decades, future breeding progress to secure productivity under drought relies critically on accurate plant phenotyping techniques that allow accurate quantification of transpiration.
An ideal system should unite i) the ability to closely simulate field growth conditions, ii) control of soil moisture, ii) automated quantification of transpiration dynamics and iv) sufficient sensitivity to detect gradual and temporal differences between genotypes.
Together with Phenospex (Heerlen, NL) we have developed a unique *Drought Spotter XXL* system that enables us to constantly track mini-plot evapotranspiration in 240 large-scale containers with a deep soil profile (90cm, 180 kg), at 5 minute intervals across an entire vegetation period from sowing until maturity. Depending on the experiment the system can monitor drought responses either by maintenance of a constant low water capacity based on container weights, or by application of a fixed quantity of water per day, regardless of the genotype-specific water consumption.
Coefficients of determination comparing seed yield to field conditions ranged between R^2=0.342 and R^2=0.544, giving reasonable confidence that the data collected in the containers can be translated to field performance. The system gives deep insight into differences in genotype responses throughout different developmental stages depending on the specific drought scenario. Based on extremely detailed datasets, the *Drought Spotter XXL* reveals diurnal patterns of genotype-specific transpiration behavior which provide highly valuable data for breeders.

W216: Climate Change and ICRCGC 2

Digging into Cowpea Root Traits Under Drought Stress

Philip A. Roberts, University of California, Riverside, Riverside, CA, Bao-Lam Huynh, Department of Nematology, University of California - Riverside, Riverside, CA and James D. Burridge, Department of Plant Science, The Pennsylvania State University, University Park, PA

Efforts to advance resilience of cowpea (*Vigna unguiculata* L. Walp.) to water stress during drought events under rainfed production systems in sub-Saharan Africa and other developing regions is of vital importance to boost yield of this protein-rich grain legume. Our focus has been to develop genomic and genetic resources to advance knowledge of and breeding for drought tolerance traits. To better understand root traits associated with enhanced biomass and grain yield productivity, we have used a cowpea Multi-parent Advanced Generation Inter-Cross (MAGIC) population of 305 recombinant inbred lines generated from eight founder parents that vary in drought tolerance phenotypes. Analyses of root architecture traits of field-grown plants that correlate with growth and yield under water-restricted field conditions provide targets for genome-based selection, using both high-density SNP fixed array and flexible SNP-KASP assays. Progress in utilizing these resources to develop high-yielding drought tolerant cowpea genotypes will be presented and discussed.

W217: Climate Change and ICRCGC 2

Can We Use Wild Ancestors to Improve Drought Tolerance? A Case Study in Common Bean

Paul Gepts[1], Jorge C. Berny Mier y Teran[1], Eneas R. Konzen[2], Antonia Palkovic[1], Viviana Medina[3], Andrea Ariani[4], Siu M. Tsai[5] and Matthew E. Gilbert[1], (1)University of California, Davis, CA, (2)Universidade do Estado de Santa Catarina, Lages, SC, Brazil, (3)World Coffee Research, San Juan, PR, (4)Bayer CropScience N.V., Gent, Belgium, (5)University of São Paulo - CENA, Piracicaba, Brazil

Domestication generally reduces genetic diversity in crop plants compared to their wild ancestors as a consequence of selection, demographic events like genetic drift, and human contingencies. The domesticated gene pools are therefore believed to contain only part of genetic diversity of the ancestral gene pools. This raises the question as to which traits might have been left behind during the domestication process. Genomics, together with other approaches such as Geographic Information Systems and phenotypic analyses, provide a way to elucidate the genetic control

of traits that may be of agronomic interest and potential tools for validation and introgression. We are investigating whether wild bean can be a source of traits conferring tolerance to drought. This raises several questions, including the nature of the tolerance mechanism, if any, the distribution of the mechanism(s) among wild bean populations, potential genomic regions controlling the drought tolerance mechanism(s), and the closeness of introgressed progenies to domesticated / market class phenotypes. We will discuss these issues using common bean (*Phaseolus vulgaris*) as a case study.

W218: Climate Change and ICRCGC 3

Using Genomics to Help Understand, Adapt and Mitigate Climate Change

Catalina Lopez-Correa, Genome BC, Vancouver, BC, Canada

The use and application of genomic technologies can support the shift towards a low carbon economy and minimal greenhouse gas emissions. When applied to solving real life problems, genomics becomes a tool that helps scientists and policy makers decide on the best course of action to address many challenges facing our ecosystems and our communities by helping us to 1) ***understand***, 2) ***adapt*** and 3) ***mitigate*** climate change. At Genome British Columbia we are funding projects and initiatives that use "omic" technologies in the areas of human health, agriculture, fisheries, aquaculture, forestry, mining and environment. Through these investments we have developed large scale projects that address climate change issues from different perspectives. During the presentation I will provide a series of concrete examples in the areas of fisheries, forestry and agriculture where Genome BC uses genomics to help understand, adapt and mitigate climate change. In particular, the analysis of genetic adaptation to climate change in Chinook Salmon, the study of genetic variation in hypoxia tolerance in Atlantic Salmon as well as the use of genomic tools for testing resistance to heat, cold, drought stress and disease in Douglas fir and lodge pole pine western larch and jack pine to help them, grow healthy in western Canada's new climates. The presentation will also discuss how the results of some of the studies are helping develop new policies around climate change in British Columbia.

W219: Climate Change and ICRCGC 3

Introgression of genes for climate resilience from wild rice into modern high yielding cultivars

Nagendra K. Singh, ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

One of the biggest challenges for 21st century is to produce sufficient food to meet the demands of growing human population from diminishing crop acreage, deteriorating soil health and abiotic stresses induced by global climate change. Crop wild relatives adapted to wide geographical and climatic ranges are a rich source of genes and alleles that can be harnessed to develop climate-resilient cultivars. Although are many successful examples of introgression from wild crop relatives into cultivars, further exploration, evaluation and utilization of fast depleting crop wild relatives gene pool is required. We have evaluated a large pool of wild rice (*Oryza rufipoghon* Griff.) germplasm and identified accessions that withstand drought, flood and soil salinity stresses better than what is available in the cultivated rice germplasm pool. Precise introgression of novel QTLs and genes for drought, flood and salinity tolerance identified in the wild rice germplasm is in progress using modern efficient tools of genomics. Accessions were evaluated for their tolerance to salinity stress. After screening, out of 292 accessions, two were found highly salt tolerant, 11 tolerant, 29 moderately tolerant, 70 salt sensitive and 180 highly salt sensitive. Phenotyping of 202 genotypes, including wild rice accessions and check varieties was done in a rain-out shelter for drought stress tolerance at vegetative stage. Four parameters *i.e.* canopy temperature, chlorophyll content, leaf rolling and relative water content was taken for screening and 35 accessions were found tolerant, 115 susceptible, and 52 moderately tolerant to drought stress. Screening of 283 wild accessions was also done for anaerobic germination and submergence tolerance. Among them 13 and 26 genotypes were found highly and moderately tolerant to anaerobic germination, respectively. Similarly, for submergence tolerance 11 accessions were reported to be highly tolerant, while 40 lines were moderately tolerant. Total 11 crosses were made for generation of F1 seeds for different abiotic stress related tolerance traits including anaerobic germination, submergence tolerance, drought and salinity stress. BC1 progenies were developed for five crosses for traits including anaerobic germination, submergence tolerance and salinity tolerance.

W220: Climate Change and ICRCGC 3

Diversifying through Minor Crops for Climate Resilient Agriculture

Wai Kuan Ho[1,2], Festo Massawe[2], Luis Salazar Licea[3], Katie Mayes[3], Niraj Shah[3], Alberto Stefano Tanzi[1,2], Norain Jamalluddin[2], Rachael Symonds[4], Hui Hui Chai[1] and Sean Mayes[1,3], (1)Crops For the Future, Semenyih, Malaysia, (2)University of Nottingham Malaysia Campus, Semenyih, Malaysia, (3)University of Nottingham, Leicestershire, United Kingdom, (4)Liverpool John Moores University, Liverpool, United Kingdom

Attaining food and nutritional security for the projected growing World population is currently being challenged by climate variability. Improving minor crops that are locally adapted (and often socioeconomically or culturally associated) could be an option, particularly for extremely vulnerable smallholder farmers. While genetic gain in the grain yield of major crops has been consistent (although farm yields have stagnated in many), crop diversification by non-resource intensive minor crops could support the livelihood sustainability of smallholder farming systems. To improve these crops for stable higher yield and produce stress resilient cultivars and mixes, molecular breeding is crucial, given that limited research resources are available.

Some minor crops are good candidates for low input and suboptimal growing environments and these could complement major crops, for additional income generation or for more nutritionally balanced diets for farming families. Our work has been focusing on bambara groundnut, winged bean, amaranth and (more recently) foxtail millet, with the first two being legumes. Bambara groundnut (*Vigna subterranea*) is a drought tolerant African legume and winged bean (*Psophocarpus tetragonolobus*) is an herbaceous multipurpose legume grown in hot and humid climates as a pulse, a vegetable (leaves and pods) or a root tuber crop depending on local consumption preferences. Amaranth (*Amaranthus;* a pseudocereal) and foxtail millet (*Setaria italic*) are cultivated for grains) with some amaranth species also being used as leafy vegetables. With the scarcity of research resources, the utilisation of molecular tools helps in the understanding of the breeding system of these species so that a more structured pre-breeding germplasm can be designed. Information on the genetic relatedness, genetic diversity and population structure helps in the better use of genetic resources by screening accessions to form a mini-core collection at a manageable scale for phenotypic evaluation. Further exploration of the genomic variation present in these germplasm would provide a better insight into the genetic basis underlying complex traits for crop improvement.

W221: **Climate Change and ICRCGC 3**

Studies on Underutilised Legumes in West Africa

Michael T Abberton, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

Bambara groundnut (*Vigna subterranea* L.*)* and African yam bean (*Sphenostylis stenocarpa*) are two important but underutilised legumes of sub-Saharan Africa. At the Genetic Resources Centre of the International Institute of Tropical Agriculture we have carried out an analysis of genetic diversity in our accessions of these crops and evaluated the same accessions for key traits. Ths we havea laid a foundation for pre-breeding and breeding in these crops.

W223: **Climate Change and ICRCGC 3**

TBA

Rudi Appels, University of Melbourne, Melbourne, Australia

W224: **Coffee Genomics**

Release of the *Coffea arabica* Variety Caturra Genome and that of its Maternal Diploid Ancestor *C. eugenioides* to Provide a Strong Foundation for Breeding and Functional Genomics Studies in Coffee

Alvaro Gaitán[1], Marcela Yepes[2], Aleksey Zimin[3,4], Carlos Ernesto Maldonado[1], Lucio Navarro[1], Claudia Flórez[1], Carmenza E. Góngora[1], Pilar Moncada[1], James Yorke[4] and Herb Aldwinckle[2], (1)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (2)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (3)Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, (4)University of Maryland, College Park, MD

The world's most widely cultivated coffee species, representing 70% of the global market, is the allotetraploid, *Coffea arabica* (2n=4x=44; genome size ~1.1 Gb). *C. arabica* evolved through the interspecific hybridization of the ancestors of two diploid *Coffea* species: *C. eugenioides* (2n=2x=22, maternal donor, genome size ~0.67 Gb) and *C. canephora* (2n=2x=22, paternal donor, genome size ~0.71 Gb previously sequenced, Denoeud *et al.* 2014).

Due to extreme bottlenecks created by human migration in the 6th and 17th centuries, throughout the world cultivated *C. arabica* varieties benefit very little from genetic diversity. The very narrow gene pool keeps cultivated *C. arabica* constantly vulnerable to diseases and insect pests [coffee berry disease (*Colletotrichum kahawae)*, coffee leaf rust (*Hemileia vastatrix)*, wilt *(Gibberella xylarioides,* anamorph *Fusarium xylarioides),* coffee berry borer (*Hypothenemus hampei*), leaf miner (*Leucoptera coffeella*), stem borer (*Xylotrechus quadripex*), and nematodes (*Meloidogyne, Pratylenchus* sps.)]. These vulnerabilities are exacerbated in the context of environmental stress due to climate change: excess rain, drought, and increased temperature.

To stream-line coffee genome analysis and breeding efforts for climate change adaptation, we sequenced, assembled, and chromosome-scaffolded the genome of allotetraploid *C. arabica* variety Caturra, as well as the genome of its diploid maternal ancestor *C. eugenioides*, to generate high-quality reference genome assemblies. To celebrate the public release of our coffee reference genomes, we have several presentations describing the importance of the target genotypes to the coffee community, the status of their release, as well as our progress on gene annotation.

This work was co-funded by the US National Science Foundation (NSF Award#1444893), the InterAmerican Development Bank through FONTAGRO, the Colombian National Coffee Growers Federation (FNC) and its National Coffee Research Center CENICAFE.

W225: **Coffee Genomics**

***Coffea arabica* variety Caturra and *C. eugenioides* Genome Assembly and Annotation**

Aleksey Zimin[1,2], Marcela Yepes[3], Carlos Ernesto Maldonado[4], Lucio Navarro[4], Sam Kovaka[2], Mihaela Pertea[5], Claudia Flórez[4], Carmenza E. Góngora[4], Pilar Moncada[4], James Yorke[1], Alvaro Gaitán[4] and Herb Aldwinckle[3], (1)University of Maryland, College Park, MD, (2)Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, (3)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (4)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (5)McKusick-Nathans Institute of Genetic Medicine, JHU, Baltimore, MD

We sequenced, assembled, and chromosome scaffolded the genome of the allotetraploid *Coffea arabica* variety Caturra, as well as the genome of its diploid maternal ancestor *C. eugenioides*, to generate high-quality reference genome assemblies for these two species that have now been deposited to NCBI Genbank under accessions RHJT00000000.1 and RHJU00000000.1.

For both genomes, we generated high coverage PacBio long-reads, error corrected with high coverage Illumina paired end reads. We assembled the data with Falcon and MaSuRCA assemblers and combined the two assemblies to produce a highly contiguous *C. arabica* (contig N50 ~4 Mb) genome, which is a substantial improvement compared to other on-going efforts co-funded by private companies and their consortia. Assembled contigs were mid-range scaffolded using 10X Genomics data with fragscaff software, and long-range scaffolded into super-scaffolds using Hi-C data and HiRise scaffolder by Dovetail Genomics. Finally, the super-scaffolds were placed onto the chromosomes with the help of genetic and physical maps for the target genotypes (Moncada *et al*. 2009, 2016; López *et al*. 2013) and existing published pseudo-chromosomes of *C. canephora* (Denoeud *et al.* 2014).

We are using several RNASeq datasets generated with Illumina and 454 to annotate our *C. arabica* and *C. eugenioides* genomes, as well as, PACBio IsoSeq data to generate robust gene models. Genome guided transcript assembly (using StringTie v. 2) to integrate short read Illumina RNAseq and long read PacBio Iso-Seq data is being used to yield a near complete gene set (protein-coding genes), and isoforms covering 94% complete single-copy plantae BUSCOs. Assembled transcripts along with trained gene prediction methods within MAKER-P are being used to create our genome annotation. We will discuss tools that we used to create our assemblies and annotation.

This work is co-funded by the US National Science Foundation (NSF Award Award#1444893), the InterAmerican Development Bank through FONTAGRO, the Colombian National Coffee Growers Federation (FNC) and its National Coffee Research Center CENICAFE.

W226: **Coffee Genomics**

Adaptive Horizontal Transfer of a Bacterial Gene to an Invasive Insect Pest of Coffee

Ricardo Acuña[1], Beatriz Padilla[1], Claudia Flórez[1], Jose Rubio[1], Juan Carlos Herrera[1], Pablo Benavides[1], Jeffrey J. Doyle[2] and Jocelyn KC Rose[2], (1)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (2)Plant Biology Department, Cornell University, Ithaca, NY

Horizontal gene transfer (HGT) involves the nonsexual transmission of genetic material across species boundaries. Although often detected in prokaryotes, examples of HGT involving animals are relatively rare, and any evolutionary advantage conferred to the recipient is typically obscure. We identified a gene (HhMAN1) from the coffee berry borer beetle, *Hypothenemus hampei*, a devastating insect pest of coffee, which shows clear evidence of HGT from bacteria. HhMAN1 encodes a mannanase, representing a class of glycosyl hydrolases that has not previously been reported in insects. Recombinant HhMAN1 protein hydrolyzes coffee berry galactomannan, the major storage polysaccharide in this species and the presumed food of *H. hampei*. HhMAN1 was found to be widespread in a broad biogeographic survey of *H. hampei* accessions, indicating that the HGT event occurred before radiation of the insect from West Africa to Asia and South America. However, the gene was not detected in the closely related species *H. obscurus* (the tropical nut borer or "false berry borer"), which does not colonize coffee beans. Thus, HGT of HhMAN1 from bacteria represents a likely adaptation to a specific ecological niche.

W227: **Coffee Genomics**

Identification of Dof1 Transcription Factor in Coffee (*Coffea arabica* L.) and its Expression in Response to 2-Oxoglutarate

Ricardo Acuña[1], Jefferson Medina[1], Monica Quintero[1], Carlos Ernesto Maldonado[1], Marcela Yepes[2], Herb Aldwinckle[2] and Alvaro Gaitán[1], (1)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (2)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY

Coffea arabica L. is a worldwide economic crop and nitrogen is one of the most important mineral elements for its growth and production. 2-oxoglutarate (2-OG) is an important regulator of carbon and nitrogen metabolism in higher plants. Feeding experiments were designed to investigate the role of 2-OG in regulation of transcription and DNA binding with the one zinc finger (Dof1) transcription factors (TFs) involved in nitrogen metabolism. These TFs participate widely in plant responses to nitrogen assimilation, but there are no reports of their activity in coffee. Using bioinformatics tools a complete sequence for *Coffea arabica* transcription factor *Cardof1* was obtained. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis showed that expression levels of *CarDof1* were higher in coffee roots than in leaves. This work lays the foundation for further analysis of the function of *CarDof1* in *Coffea arabica,* which will be helpful for improving the metabolism of nitrogen assimilation in coffee.

W228: **Coffee Genomics**

Developing a US STEM Workforce that is Globally Competitive

Marcela Yepes[1], Stephanie Fuchs[2], Matthieu Fuchs[2], Aleksey Zimin[3,4], James Yorke[3] and Herb Aldwinckle[1], (1)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (2)Cornell University/ Biological Engineering, Ithaca, NY, (3)University of Maryland, College Park, MD, (4)Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

As part of our coffee genomics NSF funded project (NSF Award Award#1444893), we organize every year a STEM career exploration workshop. The participants are rising undergraduates and High School minority students as well as Biology/Science instructors and teachers. The program is designed to explore the intersection of science, technology, engineering and math (STEM) as it pertains to applications in biological research including big data analysis, gene editing, genomics, high through put sequencing, nanotechnology, among others. Our NSF funded STEM exploration workshops are designed for undergraduate students, high school students, and rising undergraduates, who may be interested in exploring applications of STEM paths to medicine and biological research with emphasis in genomics and new cutting-edge science. Discussions and workshop exercises allow students to gain insight into applications of STEM while they interact with faculty and guest speakers. Students have the opportunity to gain a broad overview of STEM applications while they learn about current biological and medical cutting-edge research topics that are at the intersection of medicine, biology, and engineering career paths that will equip students to lead in increasingly complex, interconnected, and diverse fields. The primary workshop objective is to give participants a greater understanding of interdisciplinary fields and motivate students to follow STEM paths to produce future leaders prepared to influence their communities and the world in positive ways. Exposure to STEM is critical for high school students and for inspiring future scientists in the US. The spark is the discovery of what science and technology have to offer them in the future. **In this presentation, we will highlight one of our outreach STEM modules on the origin of the Pacific BioSciences (PACBio) long-read sequencing technology at Cornell University, with one of the co-inventors Dr. Harold Craighead, co-inventor of the technology and former Director of the National Nano-fabrication Facility at Cornell University.**

Our STEM outreach workshops are funded by the US National Science Foundation (NSF Award Award#1444893).

W229: **Comparative Genomics**

Subgenome Dominance in Interspecific Hybrids and Allopolyploids

Patrick Edger, Michigan State University, EAST LANSING, MI

W230: **Comparative Genomics**

Comparison of the Genome of Allotetraploid Peanut with its Diploid Ancestors - Homeologous Exchange Drives Diversity

David Bertioli, Center for Applied Genetic Technologies, University of Georgia, Athens, GA

Like many other crop plants, cultivated peanut (*A. hypogaea* L.) is a polyploid of hybrid origin, it has essentially complete chromosome sets from two ancestral species (*A. duranensis* and *A. ipaensis*). Here we report the genome sequence of cultivated peanut and show that after its

polyploid origin, the genome has evolved through mobile element activity, deletions and, notably, by the flow of genetic information between corresponding chromosomes derived from the different ancestors (homeologous recombination). Uniformity of some of the patterns of recombination favors a single origin for cultivated peanut and its wild counterpart *A. monticola*. However, through much of the genome, homeologous recombination creates diversity. Using new polyploid hybrids made from the ancestral species, we show how this can generate phenotypic change: spontaneous changes in flower color. This flow of genetic information between ancestral genomes is influenced by chromosome structures and is biased in different ways in different genome regions. Homeologous recombination is ongoing and is orders of magnitude more frequent than spontaneous mutation. We suggest that this mechanism, which creates genetic diversity, may have helped favor the domestication of the polyploid *A. hypogaea*over other diploid *Arachis* species cultivated by humans.

W231: **Comparative Genomics**
Assembly and Comparative Genomic Analysis of the Maize NAM Founders
Matthew B. Hufford, Iowa State University, Ames, IA

W232: **Comparative Genomics**
Havesting 15 Million Year of Oryza Genome Evolution to Help Solve the 10-Billion People Question
Rod Wing, Arizona Genomics Institute, University of Arizona, Tucson, AZ

W233: **Comparative Genomics**
The Transcriptional Landscape of Polyploid Wheat
Ricardo Humberto Ramirez Gonzalez, John Innes Centre, Norwich, United Kingdom

W234: **Comparative Genomics**
The *Chara* Genome: Secondary Complexity and Implications for Plant Terrestrialization
Stefan A. Rensing, University of Marburg, Marburg, Germany and The Chara genome consortium
Land plants evolved from charophytic algae, among which Charophyceae possess the most complex body plans. I will present the genome of *Chara braunii*; comparison of the genome to those of land plants identified evolutionary novelties for plant terrestrialization and land plant heritage genes. *C. braunii* employs unique xylan synthases for cell wall biosynthesis, a phragmoplast (cell separation) mechanism similar to that of land plants, and many phytohormones. *C. braunii* plastids are controlled *via* land plant-like retrograde signaling, and transcriptional regulation is more elaborate than in other algae. The morphological complexity of this organism may result from expanded gene families, with three cases of particular note: genes effecting tolerance to reactive oxygen species (ROS), LysM receptor-like kinases, and transcription factors (TFs). Transcriptomic analysis of sexual reproductive structures reveals intricate control by TFs, activity of the ROS gene network, and the ancestral use of plant-like storage and stress protection proteins in the zygote.

W235: **Components of Apomixis**
Apomixis in *Zephyranthes* (Amaryllidaceae): Characterization, Mapping, and Unsolved Problems
Charles F. Crane, USDA-ARS and Dept. of Botany and Plant Pathology, Purdue University, West Lafayette, IN
Evolution has favored apomixis in *Zephyranthes* because its flowering four to five days after sporadic rains presents insect pollinators with feast or famine; the apomictic species can self-pollinate without inbreeding depression. Apomeiosis in *Zephyranthes* follows the *Antennaria* type (directly mitotic divisions of the megasporocyte) and hemigamy (failure of karyogamy in the egg cell, but plasmogamy with the egg cell is necessary for embryogenesis). Four F_1 mapping populations have been produced from sexual x obligately apomictic species: *Z. traubii* (4x = 24) x *Z. chlorosolen* (8x = 48)(N = 330), *Z. longituba* (8x) x *Z. chlorosolen* (8x) (N = 109), *Z. traubii* x *Z. jonesii* (8x) (N = 53), and *Z. candida* (heterobasic 6x = 38) x *Z. pulchella* (8x) (N = 168). Each population is sufficiently pollen-fertile for self-pollination, with the exception of a few possible androgenetic haploids within the first three populations. Seed set segregates from complete to none, and intermediate levels suggest the possibility of facultative apomixis in some individuals. Most individuals of the fourth population set plump seeds, while most individuals in the first three populations set few, mostly empty seeds. Several selfed progenies of the fourth population have been uniform in morphology, indicating duplicate dominant loci for apomixis. However, hemigamy and the *Antennaria* type have been separated in various advanced-generation horticultural hybrids of related species, which reach ploidies to 2n = 36x = 216. Reliably phenotyping the mapping populations requires distinction of B_{II} and B_{III} hybrids from each other and from maternals and possible maternal and paternal haploids. Morphological analysis of testcross progenies requires space and two to three years to grow out thousands of progeny to flowering. Flow cytometry of testcross progenies is labor-intensive and requires modification of the usual Galbraith protocol. Detection of SNP or similar markers in testcross progeny pools confounds different frequencies of B_{II} and B_{III} hybrids and requires numerous DNA extractions, depending on the depth of pooling. Automated image analysis of gametophytic nuclei and nearby nucellar nuclei requires software development, is complicated by mitotic cycling in the nucellus, must reliably distinguish 1.26 from 1.00, and says nothing about hemigamy. Reliably genotyping the populations is also problematic. The large genome sizes (5-6 Gb / 1x) make genotyping by targeted sequencing of large inserts relatively difficult but also attractive, since in principle multiple “superalleles” can be distinguished with unique combinations of linked SNPs or indels within a 1-2 kb locus. Translocation heterozygosity and deletion of NOR regions occur in apomictic *Zephyranthes*, but recently developed mapping software can handle such unconventional situations. Finally, the role if any of ploidy-dependent penetrance should be considered in interpreting mapping results. Two possible mechanisms of this will be outlined.

W236: **Components of Apomixis**
TBA
Olga Kirioukhova, Jacobs University, Bremen, Germany

W237: **Components of Apomixis**

Pharmacologically Induced Apomixis in *Boechera*, *Arabidopsis* and *Vigna*: Longstanding Theories of Apomixis Origins and Regulation are Contradicted

David A. Sherwood[1], Lei Gao[1], Mayelyn Mateo de Arias[2], Bo J Price[1] and John G. Carman[1], (1)Utah State University, Logan, UT, (2)Santo Domingo Institute of Technology, Santo Domingo, Dominican Republic

Apomixis in eukaryotes is nonreduction (meiotic restitution or avoidance) coupled with parthenogenetic egg formation. In angiosperms, apomixis also includes fertilization dependent or independent endosperm formation. Since its discovery, it has been assumed that apomixis is a highly derived, mutation-generated anomaly of sexual reproduction. To test this, we studied metabolic and molecular shifts responsible for apomixis sex switching in facultatively apomeiotic *Sorghum* and *Boechera*. Much higher levels of bioenergetic homeostasis were observed in ovules of apomeiotic taxa than in ovules of sexual controls. Expression profiling followed by pharmacological tests conducted with sexual and apomictic taxa indicated that glucose signaling induces apomeiosis by activating Target of Rapamycin (TOR). The elevated levels of bioenergetic homeostasis in apomeiotic taxa appeared to originate from elevated levels of fatty acid beta-oxidation and possibly elevated rates of photosynthesis and/or sucrose import. From these findings, we developed procedures that metabolically induced high frequency apomeiosis in sexual *Boechera*, *Arabidopsis* and *Vigna* and low frequency parthenogenesis in *Arabidopsis*. Three types of apomeiosis were observed, Antennaria and Taraxacum types of diplospory and Hieracium type apospory. Our findings suggest that differences in apomeiosis types reflect tightly regulated variations in ovule metabolism. Our apomeiosis inducing procedures produced acute increases in bioenergetic activity that were timing sensitive and needed to be repeated to induce parthenogenesis. In contrast, bioenergetic homeostasis in apomicts, compared to sexual controls, was chronically elevated. Collectively, our findings are consistent with a distinct inducible apomixis program that *i*) silences sex and replaces it with nonreduction and parthenogenesis, *ii*) evolved perhaps with sex during eukaryogenesis, and *iii*) has persisted as a polyphenic alternative to sex, whether expressed or silent, in many eukaryote lineages through extended periods of deep evolution.

W2**38: Components of Apomixis**

Apomixis and Heterochrony in *Boechera:* stressful Connections Revisited

Amal J. Johnston, Jacobs University Bremen, Bremen, Germany

W2**39: Components of Apomixis**

TBA

John G. Carman, Utah State University, Logan, UT

W2**40: Components of Apomixis**

TBA

Vladimir Brukhin, Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russian Federation

Vladimir Brukhin[1,2,*], Sergei Kliver[1], Mike Rayko[1], Evgeny Bakin[1], Catherine Rushworth[3], Thomas Mitchell-Olds[4], Ueli Grossniklaus[5]

[1] Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, 41 Sredniy Prospekt, Vasilievsky Island, St. Petersburg 199004, vbrukhin@gmail.com

[2] Department of Plant Embryology & Reproductive Biology, Komarov Botanical Institute Russian Academy of Sciences, St. Petersburg, 197376 Russia.

[3] University and Jepson Herbaria, University of California, Berkeley, USA

[4] Department of Biology, Duke University, Durham NC 27708-0338 NC, USA

[5] Department of Plant & Microbial Biology Zurich-Basel Plant Science Center, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

* presenting author

Closely related to the model plant *Arabidopsis thaliana*, the genus *Boechera* is known to contain both sexual and apomictic species or accessions that grow in agamic complexes. The genome of *Boechera retrofracta* a diploid sexually reproducing species, which is thought to be an ancestral parent species of apomictic accessions after WGS study revealed a low level of heterozygosity and the presence of detectable duplications and triplications vs genomes of the apomictic accessions *B. divaricarpa*. Genes of *B. retrofracta* and 6 other Brassicaceae species were used for phylogenetic tree reconstruction. Also, we explored the histidine exonuclease *APOLLO* locus, related to apomixis in *Boechera* sexual and apomictic species, and proposed model of its evolution through the series of duplications. *APOLLO* phylogenetic tree reconstruction based on the orthologs of this gene in other species showed that there are the three copies associated with clusters of orthologous genes. The branch leading to the apo-alleles is under the positive selection, which is typical for paralogues that are required to serve a novel function. Phylogeny of the other apomixis associated genes has been analyzed as well. Moreover, chloroplast (cp) DNA of *Boechera divaricarpa* was assembled and annotated, which allowed to identify the trnL intron + trnL/F IGS. Based on this system a cp phylogenetic tree was reconstructed. It was shown that *B. divaricarpa* falls in lineage III of the trnL intron + trnL/F IGS classification.

W2**41: Compositae**

Patterning of Head-like Inflorescences in Asteraceae

Teng Zhang[1], Mikolaj Cieslak[2], Feng Wang[1], Suvi K. Broholm[1], Teemu H. Teeri[1], Przemyslaw Prusinkiewicz[2] and **Paula Elomaa**[1], (1)University of Helsinki, Helsinki, Finland, (2)University of Calgary, Calgary, AB, Canada

The key question in biology is how organisms generate reproducible patterns in a highly precise manner. In plants, the regularity is visible in the architecture of inflorescences that may vary from a single flower to large flower clusters, and is thus one of the major determinants of crop yield and reproductive success of plants. Still, the development and evolution of distinct inflorescence forms remain poorly understood. The unique feature in the Asteraceae plant family is that their inflorescence forms a pseudanthium, or false flower. While it superficially mimics a solitary flower, it is actually a tightly packed flower head (capitulum) composed of morphologically and structurally distinct types of flowers. The transference of a flower-like identity into an inflorescence is considered as the key innovation for the diversification of this largest family of

flowering plants. Using gerbera (*Gerbera hybrida*) and sunflower (*Helianthus annuus*) as models, we are exploring the dynamic networks of cellular and molecular interactions that affect early patterning of the inflorescence meristem, establishment of spiral phyllotaxis and signaling the regulates determinacy of the capitulum and leads to differentiation of distinct flower types. The knowledge is used to refine the mathematical models to understand capitulum organization.

W242: **Compositae**
Comparative Genomics in *Lactuca* and *Taraxacum*
M. Eric Schranz, University of Amsterdam, Amsterdam, Netherlands

W243: **Compositae**
Phylogenomics of Chresta (Asteraceae) and Implications for Relationships within Tribe Vernonieae
Carolina Siniscalchi[1], Benoit Loeuille[2], José Rubens Pirani[3] and Jennifer Mandel[1], (1)University of Memphis, Memphis, TN, (2)Universidade Federal do Pernambuco, Recife, Brazil, (3)Universidade de São Paulo, São Paulo, Brazil

W244: **Compositae**
Ecophysiology, Ionomics, and Genomics of Salinity Tolerance in Cultivated Sunflower
Andries A. Temme, Kelly L. Kerr, Rishi R. Masalia, John M. Burke and Lisa A Donovan, University of Georgia, Athens, GA
Given the increased agricultural use of marginal lands and current irrigation practices, there is a need for high yielding, stress-tolerant, crop varieties. Developing more stress tolerant crops will require greater knowledge of both the physiological and genomic basis of stress tolerance. Small scale studies have revealed considerable growth tolerance to salinity stress in cultivated sunflower (*Helianthus annuus* L.), making this crop a good contender for growth on salinized lands. In order to identify the genomic basis of the traits that confer tolerance to salinity stress we grew 239 genotypes of a diversity panel at zero and 100mM NaCl. After four weeks of growth (late vegetative stage) plants were harvested and assessed for total biomass, biomass allocation, leaf morphology and leaf elemental composition. Inductively coupled plasma spectroscopy (ICAP) measurements were taken to determine leaf N, P, K, S, Na, Mg, Ca, Mn, Cu, Zn, Fe, and B. Genome wide associations of >600k SNPs showed several genomic regions underlying the expression of multiple physiological and morphological traits. Interestingly different regions associated with the plasticity of trait expression under stress were also identified, indicating a trade-off in performance such that vigorous genotypes, higher biomass at zero mM NaCl, had both a larger absolute as well as proportional decrease in biomass due to increased salinity. Contrary to expectation, genotypes with a lower increase in leaf Na and Na:K ratio were no better at maintaining biomass at high salinity. Rather, genotypes with a greater reduction in leaf S and K content were better at maintaining biomass in the face of increased salinity. While we found a trade-off between vigor and tolerance, some genotypes were more tolerant than expected. This heightened tolerance was linked to a distinct suite of leaf elemental adjustments. Taken together, these results highlight the potential for breeding desirable traits conferring stress tolerance through the plants ionome.

W245: **Compositae**
Heliaphen, an Outdoor High-Throughput Phenotyping Platform for Genetic Studies and Crop Modeling
Florie Gosseau[1], Nicolas Blanchet[1], Didier Vares[1], Philippe Burger[2], Didier Campergue[3], Céline Colombet[3], Louise Gody[1], Jean-François Lievin[3], Brigitte Mangin[1], Gilles Tison[3], Patrick Vincourt[1], Pierre Casadebaig[2] and Nicolas Langlade[1], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France, (2)AGroécologie, Innovations et TeRritoires (AGIR) – INRA/CNRS, TOULOUSE, France, (3)Unité Expérimentale (UE) - INRA/CNRS, TOULOUSE, France
Heliaphen is an outdoor platform designed for high-throughput phenotyping. It allows the automated management of drought scenarios and monitoring of plants throughout their lifecycles. A robot moving between plants growing in 15-L pots monitors the plant water status and phenotypes the leaf or whole-plant morphology. From these measurements we can compute more complex traits, such as leaf expansion (LE) or transpiration rate (TR) in response to water deficit. Here, we illustrate the capabilities of the platform with two practical cases in sunflower (*Helianthus annuus*): a genetic and genomic study of the response of yield-related traits to drought, and a modeling study using measured parameters as inputs for a crop simulation. For the genetic study, classical measurements of thousand-kernel weight (TKW) were performed on a biparental population under automatically managed drought stress and control conditions. These data were used for an association study, which identified five genetic markers of the TKW drought response. A complementary transcriptomic analysis identified candidate genes associated with these markers that were differentially expressed in the parental backgrounds in drought conditions. For the simulation study, we used a crop simulation model to predict the impact on crop yield of two traits measured on the platform (LE and TR) for a large number of environments. We conducted simulations in 42 contrasting locations across Europe using 21 years of climate data. We defined the pattern of abiotic stresses occurring at the continental scale and identified ideotypes (i.e., genotypes with specific trait values) that are more adapted to specific environment types. This study exemplifies how phenotyping platforms can assist the identification of the genetic architecture controlling complex response traits and facilitate the estimation of ecophysiological model parameters to define ideotypes adapted to different environmental conditions.

W246: **Computational Gene Discovery**
Predicting the Disease Associations of Long Non-Coding RNAs using Genome-Wide Tissue Expression Profile
Xiaoyong Pan, Erasmus University Rotterdam, Rotterdam, Netherlands, Lars J. Jensen, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark and **Jan Gorodkin**, University of Copenhagen, Frederiksberg, Denmark
Long non-coding RNAs (lncRNAs) are being discovered at a fast growing rate in both animals and plants. However, far most of these discoveries are limited to non-coding transcripts without any functional characterization. And while these uncharacterized transcripts extracted from high-throughput data pile up in the databases, the functional studies remain low throughput, addressing typically individual ncRNAs. Here we aim to bridge this gap within the context of disease. In contrast to other studies that aim to associate lncRNAs with diseases from their

sequence composition, we here use their expression profiles. From a given RNA-seq data set, we thus extract the tissue expression profiles of lncRNAs (input) and associate them with diseases from a pre-generated list of tissue profiles for diseases (output). The association between input and output is obtained using a machine learning approach, here random forest models. The models are trained on expression profiles from disease-associated protein-coding genes (PCGs), as these are more abundant and better characterized than the lncRNAs. Our DislncRF method then employs models that each has learned to associate the patterns of the PCG tissue expression profiles to a specific disease. The PCG trained models are then used to make prediction on lncRNAs tissue profiles distributed over the same tissues as the PCGs. We find that DiclncRF outperforms other methods using a gold data set of lncRNAs with disease associations not used in the training.

W247: Computational Gene Discovery
Towards a Complete Genome Map of Long Noncoding RNAs in Human and Mouse
Julien Lagarde, Centre for Genomic Regulation (CRG), Barcelona, Spain

W248: Computational Gene Discovery
Lessons Learned from the Maize Pan-Genome Annotation
Nadia Kamal[1], Manuel Spannagl[1], Andreas Houben[2], Thomas Presterl[3], Milena Ouzunova[3], Frank Hochholdinger[4], Klaus F.X. Mayer[1], Eva Bauer[5], Chris-Carolin Schön[5] and Georg Haberer[1], (1)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, (3)KWS SAAT AG, Einbeck, Germany, (4)INRES Crop Functional Genomics, University of Bonn, Bonn, Germany, (5)Technical University Munich, Freising, Germany
Precise gene predictions are an essential requirement for any pan- and comparative genomics analysis because errors without correction may result in wrong or incomplete conclusions. In a maize pan-genomics project, we report the analysis of four novel reference genomes of European flint lines: three elite lines (F7, EP1 and DK105) that are important founders of European breeding programs; and a doubled haploid (DH) line derived from Petkuser, a European landrace. All lines were assembled to pseudo-chromosomes with scaffold N50 ranging from 6 to 10 Mb using the DeNovoMagic™ pipeline.
Detailed inspection of the gene content suggests that the genic presence/absence variation (PAV) in maize may have been overestimated by recent studies due to e.g. missing or low quality gene models. To compensate for gene calls missed in one or more lines we developed an automated protocol to identify, locate and re-annotate gene models. The routine utilizes whole genome alignments to identify syntenic regions, cross-map gene models of each of the lines and identifies gene structures missed by the initial annotation targeting one genome at a time only. In a second cross-mapping phase, structure of the gene models in each line are re-modelled by borrowing information from syntelogs of all lines. We show that this approach achieves significant improvements in both the number of pairwise and total syntenic relationships as well as the completeness and structural uniformity of syntelog clusters.
Funding acknowledgements: Bavarian State Ministry of the Environment and Consumer Protection (BayKlimaFit; http://www.bayklimafit.de/) and the German Federal Ministry of Education and Research (BMBF; MAZE; http://www.europeanmaize.net/)

W249: Computational Gene Discovery
Automatic Genome Annotation Looping over Species
Mario Stanke, University of Greifswald, Greifswald, Germany and Katharina J. Hoff, Ernst-Moritz-Arndt Universität Greifswald, Greifswald, Germany
RNA-Seq spliced alignments are needed both for training of statistical gene finders and for increasing the accuracy of their prediction. For a species that ought to be (re-)annotated, the sequence read archive (SRA) often contains large data from hundreds or thousands of sequencing runs. However, the read data is very heterogeneous with highly varying alignability and percentage of spliced reads. Further, experiments from different tissues or conditions cover different subsets of the set of all transcripts.
We present results from the tool VARUS that uses an online algorithm to decide from which run to download and align (further) data in order to achieve a high usefulness for annotation with a small amount of data. It prefers runs whose reads align with a high average number of splices and whose expression profiles complement each other. VARUS does not use meta data that indicates tissue or condition, but rather the coverage distribution of genome regions to infer, how well further data from a run would cover yet poorly covered regions. We thereby find the same number of true introns with a much smaller number of downloaded reads than when manually selecting a set of sequencing runs to download and align completely.
VARUS enables us further to have an annotation loop, where we need for each species only a genome assembly file and the binomial species name. VARUS creates a BAM file of spliced alignments, the BRAKER pipeline then trains GeneMark and AUGUSTUS, and the species-specifically trained AUGUSTUS uses the spliced alignments for annotating each genome.

W250: Computational Gene Discovery
Orthodb and BUSCO Evolutionary Perspective on Genes and Genomes
Evgeny M. Zdobnov, University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland

W251: Computational Gene Discovery
Whole-Genome Annotation at NCBI
Francoise Thibaud-Nissen, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD
The National Center for Biotechnology Information (NCBI) plays the dual role of an archival repository (e.g. GenBank, PubMed, SRA) and of a source for curated data and aggregated knowledge. An example for the latter is gene annotation in RefSeq.

The eukaryotic genome annotation pipeline is an actively maintained pipeline that has been used by RefSeq to annotate over 500 species. Organisms are selected for annotation based on their relevance to human health and their economic or biological importance, as well as the quality of the genomic assemblies, the public availability of suitable RNA-Seq data and user requests. The core component of the pipeline is Gnomon, a gene prediction algorithm highly dependent on experimental evidence. The prokaryotic genome annotation pipeline (PGAP) calculates structural and functional annotation using protein alignments, *ab initio* predictions with GeneMarkS-2, and annotation rules such as hidden Markov models. PGAP has been used for the annotation of over 130,000 RefSeq prokaryotic assemblies and is also run on GenBank-submitted assemblies upon request.
One common strength between the two pipelines is their ability to inform the annotation using curated datasets in addition to primary data. We will present how transcript curation for eukaryotes, and hidden Markov model construction and curation for prokaryotes are critical elements to the success of both annotation projects.
Finally, beyond the delivery of annotated products, we will show how NCBI is stepping up its effort to put annotation tools in the hands of outside users and is now providing a standalone, containerized, version of PGAP.
The results of the pipelines are publicly available on the NCBI FTP site ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/ and through Entrez Gene, Assembly, Protein and Nucleotide.

W252: Connecting Crop Phenotype and Genotype Data
G.E.M.S, Managing Data and Analysis for Genomes2Fields
Kevin A. T. Silverstein, Supercomputing Institute, University of Minnesota, Minneapolis, MN
G.E.M.S is a new agroinformatics platform designed to facilitate public-private data sharing and analysis. It has been developed in collaboration by the College of Food Agriculture and Natural Resources Sciences (CFANS) and the Minnesota Supercomputing Institute (MSI) at the University of Minnesota with strong input by the international agricultural community. Genomes2Fields (G2F) is a 23-state, 37-site coordinated corn genomics and predictive phenomics initiative formed in 2014. With over 63,000 plots planted from 2014-2017, this collaborative enterprise is designed to expand our understanding of the expression of corn genes across diverse growth environments. As of January 2018, GEMS now plays a major role in planning, coordinating, and implementing data management for G2F. Working together we are streamlining data collection and submission, enabling real-time data cleaning upon upload, and progressing toward real-time analysis/comparison to previous years' data.

W253: Connecting Crop Phenotype and Genotype Data
Information Management Tool to Support Accelerated Potato Breeding Scheme
Elisa Salas, Ivan Perez, Vilma Hualla, Thiago Mendes and Hannele Lindqvist-Kreuze, International Potato Center, Lima, Peru
The approach conventional breeding programs have followed for decades, start with population development and end with the selection of the "best" individuals in the genetic variation/breeding population and variety release. This conventional breeding scheme takes from 7 to 8 years in the International Potato Center (CIP). With the pressure to make faster improvements in farmer's lives, the "conventional" breeding scheme is no longer adequate. Now, with the Accelerated Breeding Scheme (ABS) the release of varieties since year 5 is feasible only if the major bottlenecks are broken with creative and innovative tech in the scheme of a breeding program.
To support the potato breeding process and make decisions in real time, **the Global Trial Data Management System (GTDMS)** https://research.cip.cgiar.org/gtdms/ was developed. This site contains protocols, catalogs, and novel breeding information management tools designed for clonally propagated crops. A key advance has been the development of data management structures to connect breeding material lists of families and clones with the institutional pedigree and the Root and Tuber Base (R&TBase). This connectivity permits verification and maintenance of the identity of clones across the different selection stages and to follow their pedigrees.
The data generated in the CIP potato breeding program is stored at "**Global Roots & Tubers Base" (R&TBase)** utilizing the free BioMart software https://research.cip.cgiar.org/gtdms/biomart/. The data have been structured for storage of phenotypic, genotypic, pedigree, geographical, and environmental data. Through the metadata and the search function using filters, the user can retrieve data from the experiments conducted by CIP scientists or NARS partners using CIP materials. The availability of the data is managed in conjunction with the Dataverse following CGIAR open access guidelines.

W254: Connecting Crop Phenotype and Genotype Data
Modernizing Breeding in Africa: Primed for More
Jean-Marcel Ribaut, Integrated Breeding Platform, Texcoco, Mexico
There is considerable potential to raise agricultural productivity in Africa through the development of improved cultivars that generate more yields and respond to market demands. Sustainable change will be made possible by the implementation of a demand-led breeding practice, based on modern technologies adapted to local realities, and on a strong capacity development component (human and infrastructure). Recent advances in genotyping technologies, combined with the development of local genetic resources, have had a significant impact on the capacity of African breeders to link phenotype to genotype through association studies, linkage mapping and diversity analysis. Various international initiatives have been working on re-sequencing and/or screening on high density SNP arrays a number of representative sets of African germplasm (e.g. reference sets, MAGIC populations or association panels), and the number of trait-linked markers for simple inherited traits keeps increasing. As of today, the capacity to generate reliable phenotyping data from segregating material, and to manage the different data sets in an effective and integrated way, remain major limitations that hinder the effective implementation of molecular breeding in Africa. Building on concrete examples, the Integrated Breeding Platform (IBP, https://www.integratedbreeding.net) is tackling these challenges head on. This talk will illustrate how to overcome some of the bottlenecks encountered in data management by using the Breeding Management System (BMS), underlining the importance of crop ontology and meta-data, a must-have to conduct analyses across locations, projects and institutes.The talk will also report on the increasing number of National Breeding Programmes in Africa digitalizing their breeding programmes and on, how they can now integrate crop information with phenotypic and genome data more easily thanks to increased interoperability between tools and platforms – hence enabling effective forward breeding, gene introgression and potentially genomic selection.

W255: Connecting Crop Phenotype and Genotype Data

Efficient Integration of Marker-Assisted and Genomic Selection in a Modern Rice-Breeding Program

Juan D. Arbelaez[1], Erwin S. Tandayu[1], Maria Ymber V. Reveche[1], Patrik Stolt[2], Eng Hwa Ng[3], Maria Stefanie Dwiyanti[4], Damien J. Platten[1], Tobias Kretzschmar[5], Jessica Rutkoski[1], Star Yanxin Gao[6] and Joshua N. Cobb[1], (1)International Rice Research Institute, Los Baños, Philippines, (2)Intertek, Alnarp, Sweden, (3)Excellence in Breeding, Los Baños, Philippines, (4)Hokkaido University, Sapporo, Japan, (5)Southern Cross University, East Lismore, Australia, (6)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY

Marker-assisted and genomic selection (MAS and GS) has great potential to increase breeding program efficiency and rate of genetic gain for multiple traits in public rice breeding programs. Validation and integration of MAS and GS in breeding programs is often hindered by the high cost of genotyping, as well as a lack of intuitive analytical tools, statistical expertise, and streamline protocols for scalability. To address some of these challenges we have taken an interdisciplinary approach to improve the return on investment of genotyping by developing a custom low-density genotyping platform (1k-RiCA) with 1K uniformly distributed genome-wide SNPs and more than 100 trait markers. This platform was specifically developed to enable the application of combined GS and MAS strategies for *indica* rice germplasm. In addition, we optimized a single-seed DNA based MAS forward-breeding protocol that decreases the cost and improves the efficiency of MAS by 4 fold compared to a conventional leaf DNA-based MAS protocol. To demonstrate the utility of these resources in breeding, we performed a large scale single-seed DNA based MAS and using the 1k-RiCA, GS models were validated for multiple relevant traits framed in the context of our breeding program. In collaboration with GOBii (http://www.gobiiproject.org/) a genomic prediction open-source Galaxy platform is being developed to enable the routine application of GS in breeding programs.

W256: **Connecting Crop Phenotype and Genotype Data**

Connecting Genotyping Data, Sample IDs, Breeding Data and Analysis Tools to Facilitate Breeding Decisions

Elizabeth Jones[1], Yaw A. Nti-Addae[1], Star Yanxin Gao[2], Philip Glaser[1], Pete Selby[2], Kate A. Dreher[3], Mariano Crimi[4], Abhishek Rathore[5], Marko Karkkainen[6], Petra Vanroggen[7], Eng Hwa Ng[8], Iain Milne[9], Qi Sun[2], Andrzej Kilian[10], Susan McCouch[2] and Kelly Robbins[2], (1)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (2)Cornell University, Ithaca, NY, (3)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (4)CIMMYT, Int., Texcoco, EM, Mexico, (5)International Crops Research Institute For The Semi-Arid Tropics, Patancheru, India, (6)International Rice Research Institute, Los Baños, Philippines, (7)Intertek, Alnarp, Sweden, (8)Excellence in Breeding, Los Baños, Philippines, (9)The James Hutton Institute, Dundee, United Kingdom, (10)Diversity Arrays Technology Pty Ltd, Bruce, ACT, Australia

For genomic selection to be effectively implemented, genomics data management systems must be fully integrated with breeding data management, sample tracking and downstream analysis tools. GOBii (Genomic Open-Source Breeding informatics initiative) is a Bill and Melinda Gates Foundation funded project with the mission to implement genomic and marker assisted selection as part of routine breeding programs for staple crops in the developing world. The first step in this process has been to develop a genomic database management system that is highly searchable and flexible enough to be used across different crops profiled with diverse marker and sequence-based platforms. We are now challenged with connecting this system with other data management systems in various stages of development and adoption. First we are focusing on connecting GOBii with the Sample Tracker tool developed at CIMMYT, that will push projects and samples to GOBii, and then coordinate automatic genotyping data loading from vendors to GOBii. Adoption of unique sample IDs are key to maintaining data integrity across systems and enabling genotypic data to be connected back to field plots. We have developed downstream marker analysis tools in Flapjack and Galaxy for marker-assisted backcrossing, forward breeding, pedigree verification, and genomic selection and are moving towards more seamless integration between GOBii and these applications. Our goal is to develop molecular breeding tools that minimize the need for manual data manipulation and improve the efficiency of breeding decisions to make genetic gains.

W257: **Connecting Crop Phenotype and Genotype Data**

Android Apps for Phenotyping

Trevor W. Rife and Jesse Poland, Department of Plant Pathology, Kansas State University, Manhattan, KS

Plant breeding and genetics research inherently relies on rapid and accurate data collection. A typical experiment or breeding nursery can contain thousands of unique entries and breeding programs will often evaluate tens of thousands of plots each year. To operate efficiently and effectively at this scale, electronic data management is essential. However, many research programs continue to operate by scribing and transcribing massive amounts of data on paper field books. This form of data collection and management places heavy burdens on human resources, decreases data integrity, and limits future utilization of data and the ability to expand the breeding program. To help address these constraints by efficiently collecting and organizing data throughout all aspects of the breeding cycle, we have developed several Android apps including Field Book, Inventory, Intercross, and Coordinate. With these apps, we are working to decrease both technological and cost barriers that hinder adoption of electronic data management in breeding programs by focusing on simple applications with intuitive and customized interfaces. Integrating low-cost hardware into breeding programs will provide an inexpensive platform for additional apps to be developed and distributed, creating accessible solutions for breeding programs around the world. The adoption of electronic data collection and management is a necessary and essential step to the realization of a contemporary green revolution.

W258: **Connecting Crop Phenotype and Genotype Data**

Excellence in Breeding Platform

Michael Quinn, Excellence in Breeding, Chapingo Issste, Mexico, Mexico

W259: **Connecting Crop Phenotype and Genotype Data**

Breeding Insight Platform

Jean-Luc Jannink, USDA-ARS / Cornell University, Ithaca, NY and **Moira J. Sheehan**, Cornell University, Ithaca, NY

W260: Cool Season Legumes

A Tale of Two Systems: Microbiome Diversity of Wild and Cultivated Chickpeas

Betsy Alford, University of California-Davis, Davis, CA

Soil is among the most complex and understudied of microbial habitats, with an estimated 10,000 to 50,000 different species in a single gram of soil. Plants depend on soil and the constituent microorganisms for a range of services, including nutrient acquisition. Most domesticated crops are cultivated in areas distant from their center of origin. Domestication typically involves reduction in plant genetic diversity, and it is vital to understand the extent to which the structure, diversity and functions of the plant microbiome have been conserved through domestication. Wild chickpea populations of *Cicer reticulatum* and *C. echinospermum*offer an attractive system to elucidate plant-microbe interactions associated with domestication. The ecology and genomics of wild chickpea, which is confined to a ~100 km^2region of SE Turkey, have been well characterized by our lab, with detailed surveys of host population, soil and climatic factors. Previous research in this project identified factors that drive microbial community structure in longstanding wild populations of *C. reticulatum* and *C. echinospermum*to understand the impact of domestication and agronomic factors on microbial communities associated with the derived cultivated species, *C. arietinum.*The fundamental analytical method of the work is culture-independent sequencing of bulk DNA with 16S rDNA amplicon sequencing, with the goal of assigning microbial taxa and enumerating their abundance. The study system includes the natural environments of SE Turkey, but also thousand years-old cultivated sites at sites of secondary crop diversification in India (~6KYA) and Ethiopia (~3KYA), and recent <100 year sites of cultivation in N. America. Correlations of microbial taxa frequencies with edaphic and environmental variation will estimate the impact of ecological factors on community structure in natural systems, as will a common garden reciprocal soil transplant experiment involving Turkish soils in an attempt to discern the relative importance of plant genotype and soil type. Finally, a greenhouse experiments in agricultural soils will be conducted to estimate the resiliency of microbial community structure to common US agronomic inputs, such as nitrogen supplementation. This project aims to elucidate the change in relationship between plant host, symbiont, microbiome, and environmental factors in the context of domestication of the plant host and soils to identify potential impacts on plant health.

W261: Cool Season Legumes

Towards Identification of Genes underlying Two Key Domestication Traits in Pea: Pod Dehiscence and Seed Dormancy

Petr Smykal, Palacky University Olomouc, Olomouc, Czech Republic

The process of crop domestication was based on selection driven by human cultivation practices and agricultural environments, as well as other population genetic processes such as a reduction in effective population size. These processes led to the so-called domestication syndrome, including changes in plant structure, plant defensive and palatability. However there are two traits considered crucial: reduced dispersal ability and eliminated seed dormancy.

We have used an integrated view combining comparative anatomy, metabolomics, genetic mapping and transcriptome profiling in order to identify genes associated with loss of seed dormancy and pod dehiscence in pea. Differences were found in texture of the testa surface, length of macrosclereids and seed coat thickness. The light line within the macrosclereid layer was identified as the major barrier to water, representing the interface between two distinct environments, the waxy subcuticular layer and the cellulose-rich secondary cell wall.

Liquid chromatography–mass spectrometry and mass spectrometry imaging identified significantly higher contents of proanthocyanidins and hydroxylated fatty acids in seed coats of dormant compared to non-dormant genotypes. RNA sequencing identified 770 and 148 differentially expressed genes between dormant and non-dormant seeds or dehiscent and indehiscent pods, respectively. The homolog of peptidoglycan-binding domain or proline-rich extensin-like protein was both differently expressed and mapped to predicted *Dpo1* locus on PsLGIII. Genes of the proanthocyanidin pathway were significantly enriched and upregulated in dormant genotypes. Genetic mapping identified 2 to 3 loci involved in seed dormancy (testa thickness and germination response).

Germination testing identified a subset of wild peas (*P. elatius*) with substantially reduced seed dormancy, which might be used during domestication process. There have been at least two domestication events in pea crop and it will be interesting to study the consequences of this once the respective underlying genes have been identified.

W262: Cool Season Legumes

Dissecting the Genetic Architecture of Aphanomyces Root Rot Resistance in Lentil by Combining QTL Mapping and Genome-Wide Association Analysis

Yu Ma[1], Clarice J Coyne[2], Dorrie Main[1], Lyndon Porter[3], Deus Mugabe[1] and Rebecca McGee[1], (1)Washington State University, Pullman, WA, (2)USDA ARS, Pullman, WA, (3)USDA-ARS, Prosser, WA

Lentil (*Lens culinaris* ssp. *culinaris* Medikus), an important grain legume crop, is widely grown throughout the world. With high protein, minerals, carbohydrates and fiber, it is an inexpensive food source to alleviate malnutrition in developing countries. Associated with nitrogen fixation, it significantly benefits cereal-based cropping system. Aphanomyces root rot (ARR), caused by *Aphanomyces euteiches* Drechs., is one of most devastating diseases in lentil production, and can cause yield loss up to 80%. Cultural practices, fungicides, biological control, and soil fumigants are either ineffective, uneconomic, or environmentally unfriendly approaches to manage ARR. The most effective, economical and sustainable management of ARR is through the development and utilization of cultivars with high levels of partial resistance. No lentil cultivars resistant to ARR are currently available. Genome-wide association studies (GWAS) and linkage analysis can be used to identify marker trait associations and develop breeder friendly markers to use in a breeding program. In this study, large-scale SNPs data was generated via genotyping by sequencing (GBS). We performed linkage analysis in a RIL population composed of 189 F_6-derived lines, and GWAS using the USDA lentil single plant-derived (LSP) collection (223 accessions) and an international collection (109 accessions) from ICARDA. Root rot index and percentage of shoot and root dried weight losses per plant were collected in controlled conditions, while above ground index was collected in an ARR field nursery. Through the combination of two QTL analysis approaches, putative QTLs associated with ARR resistance were identified to uncover the genetic basis underlying the complex trait.

W263: Cool Season Legumes

Pathogenicity Factors in the Fast Evolving Australian *Ascochyta rabiei* Population

Ido Bar, Grifith University, Brisbane, QLD, Australia
Australian chickpea production has increased dramatically in the past 5 years through expansion into new growing regions, driven by record-high prices and water availability. This resulted in the industry onboarding of multiple novice growers who sourced seed outside of the clean seed programs and did not adhere to best disease management practices. This in turn led to greater diversity in the *Ascochyta rabiei* fungal populations than in all other growing regions, with a higher frequency of highly aggressive isolates. The sub-population of high risk isolates appear to be well adapted to all growing regions and able to cause substantial disease on the broadly grown 'resistant' cultivars as well as the major source of resistance used in the Australian national chickpea breeding program, ICC3996 [1]. Identifying the 'pathogenicity factors' that differentiate these high risk isolates from those less able to infect and cause disease epidemics will inform on future disease management strategy and may aid in understanding the evolution and adaptive mechanisms that the pathogen is employing in the Australian environment.
Therefore, ongoing investigation is underway to determine association between specific genetic features and disease aggressiveness among collections of isolates that have previously been phenotyped on a consistent differential host set (ref). For this, Whole-Genome-Sequencing (WGS) of selected isolates was performed on isolates from a range of host cultivars, pathogenicity levels and regions.
Genomes of forty *A. rabiei* isolates were sequenced on an Illumina HiSeq2500 platform, producing high quality 100 bp paired end reads, which were aligned to a reference *A. rabiei* genome (strain *me14*, Curtin University), covering the 34.6 Mb genome at an average depth of x85 with high alignment rates averaging 97.5%.
Genomic variants were called from the mapped alignments, identifying 14,441 single nucleotide polymorphism sites (SNPs). High confidence SNPs were scanned for significant association to the isolate pathotype groups. Associated SNPs were further annotated to determine their effect on gene protein coding sequences. Notable SNPs were found at genes involved in plant-pathogen interactions and some contained a molecular signature characteristic of effector molecules, assumed to be the key factor determining host infection and colonization [2].
Functional validation of these associated genes as pathogenicity effector molecules will be performed using gene knockout methods and will help understanding the infection pathway to inform breeding programs, fungicide development and cultivar usage guidelines.
Mehmood Y, Sambasivam P, Kaur S, Davidson J, Leo AE, Hobson K, et al. Evidence and Consequence of a Highly Adapted Clonal Haplotype within the Australian *Ascochyta rabiei* Population. Front Plant Sci. 2017;8. doi:10.3389/fpls.2017.01029.
Selin C, de Kievit TR, Belmonte MF, Fernando WGD. Elucidating the Role of Effectors in Plant-Fungal Interactions: Progress and Challenges. Front Microbiol. 2016;7. doi:10.3389/fmicb.2016.00600.

W2**64: Cool Season Legumes**

Achieving Higher Genetic Gain by Enhancing Precision through Genomic Selection Breeding in Chickpea

Manish Roorkiwal[1], Nicholas Santantonio[2], Diego Jarquin[3], Bharadwaj Chellapilla[4], Muneendra Singh[1], Pooran M Gaur[1], Srinivasan Samineni[1], Reka Howard[3], Star Yanxin Gao[2], Elizabeth Jones[2], José Crossa[5], Kelly Robbins[2] and Rajeev K Varshney[1], (1)ICRISAT, Hyderabad, India, (2)Cornell University, Ithaca, NY, (3)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, (4)IARI, Delhi, India, (5)International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico
Genomic selection (GS) predicts breeding values of lines using genome-wide marker data and allows breeders to select lines prior to field phenotyping and shortening the breeding cycle. A set of 320 elite breeding lines were extensively phenotyped for different traits (e.g., plant height, days to maturity, and seed yield). These lines were genotyped using DArTseq (1.6K SNPs), Genotyping-by-Sequencing (GBS; 89K SNPs) and Axiom®*CicerSNP* Array (24K SNPs) platforms. Phenotypic data for eight traits for three seasons at two locations along with genotyping data were used to assess the impact of environment, lines, and other interactions for 13 different GS models. Three different cross-validation (CV) schemes that mimic real scenario that breeder encounters on field, were used to assess the prediction accuracies (CV2: incomplete field trials; CV1: newly developed lines; and CV0: new environments). These results suggested the potential of these GS models for chickpea cultivar improvement. In order to deploy the GS models in regular breeding program for selecting lines based on marker data, a pilot experiment was initiated as part of Genomic Open-source Breeding Informatics Initiative (GOBII) activities to deploy genomic information for crop improvement. GOBii focus on developing user friendly tools that have great potential in developing improved cultivars faster and more precisely by deploying modern breeding approaches. Chickpea breeding program from ICRISAT and IARI were targeted deploying the use of markers for selecting lines using GS. A total of 6000 F_5 lines form 12 different crosses from ICRISAT and IARI breeding program were selected and genotyped using DArTseq platform. Based on data quality genotyping data for 4,923 lines were used for genomic predictions using 10-fold consolidation scheme cross-validation. We would like to highlight that, many lines that we targeted were not related with training population, therefore accuracies achieved for these lines is lower than the cross-validation accuracies. All the traits showed significant genotype by environment (i.e. ICRISAT and IARI) interactions, therefore, we produced predictions within each of the two breeding programs. Univariate (single trait) and multivariate (multi-trait) cross-validation produced equivalent prediction accuracies within the population, therefore we focused on univariate predictions. Based on these univariate prediction top 200 lines each for ICRISAT and IARI breeding programs were selected. In parallel, 200 lines each were selected by ICRISAT and IARI breeders based on visual score. These two sets (selected based on genomic prediction and based on visual score) evaluated in field condition to validate the genomic prediction results. Based on the initial results, we plan to expand the current training population by including founder parents from different chickpea breeding program in India and initiate deployment of GS. Preliminary results suggest that GS models hold potential for breeder's applications on chickpea cultivar improvements.

W2**65: Cool Season Legumes**

Improved *Lens culinaris* Reference Assembly gives new Insight into Regions of Genomic Complexity

Larissa Ramsay[1], Chu Shin Koh[2], Taryn Heidecker[1], Dongying Gao[3], P.R. MacLachlan[1], Peter Chang[4], Jaroslav Doležel[5], Douglas R Cook[4], R. Varma Penmetsa[4], Andrew G. Sharpe[6], Clarice J Coyne[7], Dorrie Main[8], Sukhjiwan Kaur[9], Bert Vandenberg[1] and Kirstin Bett[10], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)University of Georgia, Athens, GA, (4)University of California-Davis, Davis, CA, (5)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (6)National Research Council Canada, Saskatoon, SK, Canada, (7)USDA ARS, Pullman, WA, (8)Washington State University, Pullman, WA,

(9)Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia, (10)Dept of Plant Sciences/University of Saskatchewan, Saskatoon, SK, Canada
Due to its large genome (4.2Gb) and high repeat content, cultivated lentil (*Lens culinaris*) has been a challenge for short-read assembly. However, with a combination of long read technologies (Pacific Biosciences SMRT sequencing and 10x Genomics linked reads) and optical map (BioNano), we generated an improved assembly of 4.06Gb of assembled sequence in 24,510 scaffolds (N_{50} size 1.9Mb, N_{50} value 532 scaffolds). Using an exome capture array on a single *L. culinaris* x *L. culinaris* recombinant inbred line (RIL) population, we developed a high-density genetic map containing 74,046 single nucleotide polymorphism (SNP) markers in 17,484 bins, anchoring a total of 3,154 scaffolds (3.4Gb). Assembly completeness was validated with BUSCO (93.8%), and whole genome accuracy evaluated by comparison to the previous short read assembly and assembled BAC sequences. We are able to use the improved scaffolds to better characterize long stretches of repetitive elements, examine repeat expansion relative to wild species, and investigate genomic breakpoints associated with translocations that have been identified in interspecific maps. Hi-C and additional long read data (Oxford Nanopore) is being added to disambiguate scaffold ordering in regions of low recombination and complete pseudomolecule assembly.

W266: Cool Season Legumes

Cool Season Food Legume Genome Database: A Resource for Pea, Lentil, Faba Bean and Chickpea Genetics, Genomics and Breeding

Jodi L. Humann[1], Sook Jung[1], **Chun-Huai Cheng**[2], Taein Lee[2], Ping Zheng[1], Morgan Frank[1], Deah McGaughey[1], Kristin Scott[1], Katheryn Buble[3], Jing Yu[1], Heidi Hough[1], Marwa N.M.E. Sanad[4], Clarice J Coyne[5], Rebecca McGee[6] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Washington State University, Pulmman, WA, (4)National Research Center- Egypt, Cairo, Egypt, (5)USDA ARS, Pullman, WA, (6)USDA-ARS, Pullman, WA
The Cool Season Food Legume Genome database (CSFL, www.coolseasonfoodlegume.org) is a curated and integrated database resource for genomics, genetics, and breeding research of chickpea, lentil, pea, and faba bean. CSFL currently contains three annotated genome sequences for *Cicer* sp.; annotated reference transcriptomes analyzed from published RNA-seq and EST data for all four crops; 167 genetic maps; 137,268 markers; 2,952 QTL; 2,429 germplasm; metabolic pathways for the *Cicer* sp. genomes; and synteny data for the *Cicer* sp. genomes with links to genes, mRNA, orthologs and function. Tools include the genome browser JBrowse, Synteny Viewer, MapViewer, PathwayCyc, 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. Public cool season legume phenotype data from the USDA-GRIN database is available to explore with BIMS by all CSFL users. Genes and transcripts, maps, markers, germplasm, QTL, sequences and publications can be queried through the search interfaces on CSFL with results available to download. CSFL is built using the Tripal database platform supported by USDA NRSP10, the USA Dry Pea and Lentil Council, Northern Pulse Growers Association, USDA-ARS and Washington State University.

W267: Cool Season Legumes

Knowpulse: An Evolving Breeder-Friendly Web-Portal for Chickpea, Common Bean, Field Pea and Lentil

Lacey-Anne Sanderson, Carolyn T. Caron, Reynold Tan, Crystal Chan and Kirstin Bett, Dept of Plant Sciences/University of Saskatchewan, Saskatoon, SK, Canada
KnowPulse (http://knowpulse.usask.ca) is a publicly-searchable web-based resource for plant breeders and geneticists interested in pulse crops. It is built using Tripal and developed at the University of Saskatchewan, providing project management for the pulse breeding program. KnowPulse also provides public access to various tools for utilizing genomic and diversity data for chickpea, common bean, field pea and lentil. Genomic data can be used via our crop-specific JBrowse instances and BLAST databases. Genotypic data is summarized on marker pages and can be queried, resulting in a marker-by-germplasm table for comparison among germplasm. To enhance accessibility, species-specific "launchpads" are now available which provide access to summary charts, publications and searches for germplasm and genetic markers. KnowPulse is constantly evolving with data and tools added as they become available. Full integration of phenotypic data via trait and germplasm pages, and visualizations of genetic maps is imminent.

W268: Crop Evolution Genomics & Future Agricultural Productivity

TBA

Edward S. Buckler, USDA-ARS-Cornell University, Ithaca, NY

W269: Crop Evolution Genomics & Future Agricultural Productivity

The Abyssinian Pea (*Pisum sativum* ssp *abyssinicum*) appears to have been Produced from a *P.s.* ssp *elatius* x *P.s.* ssp *sativum* Hybridization

Norman Weeden[1], Clarice J Coyne[2] and Kevin McPhee[1], (1)Montana State University, Bozeman, MT, (2)USDA-ARS-WRPIS, Pullman, WA
The origin of the Abyssinian pea (*Pisum sativum* ssp. *abyssinicum*) has been a topic of considerable discussion since the taxon's scientific recognition in the early 1900s. It is cultivated in the Ethiopian highlands and Yemen, is not known from the wild, possesses a karyotype distinct from other domesticated forms, and exhibits reduced sterility in crosses with both domesticated and most wild peas. The taxon displays very little genetic variability, suggesting that it is of recent (<4000 years ago) origin and leading some to propose that the Abyssinian pea is of hybrid origin. We tested this hypothesis by examining mitochondrial, plastid and nuclear DNA markers in a wide range of domesticated and wild *Pisum* germplasm. The results allowed a clear rejection of the possibility that this taxon was the immediate product of a *P. fulvum* x *P. sativum* hybridization, currently the most widely accepted hypothesis. Instead, the taxon appears to have been generated by the cross *P. sativum* ssp. *elatius* x *P. sativum* ssp. *sativum,* with the elatius germplasm being closely related to several accessions collected in Israel. These elatius accessions are some of the few with cytoplasmic markers identical to those characterizing *P. fulvum*. Hence, the previously observed similarity between certain markers in the Abyssinian pea and *P. fulvum* is most likely due to a distant relationship through the maternal lineage.

W270: **Crop Evolution Genomics & Future Agricultural Productivity**

Evolution of Novelty in Wheat Disease Resistance occurs at very different Time Scales

Beat Keller[1], Coraline R. Praz[1], Simrat Singh[1], Thomas Wicker[2] and Simon G. Krattinger[3], (1)University of Zurich, Zurich, Switzerland, (2)Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland, (3)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

In the bread wheat genome, several hundred resistance genes against fungal diseases have been genetically described. An increasing number of them has been isolated and characterized at the molecular level, although most of them remain to be studied. We have isolated several powdery mildew and leaf rust resistance genes from the cultivated wheat gene pool. Based on their sequence we have searched for their relatives in wheat progenitors and relatives to determine the evolutionary history of these genes. We found two very different time scales when resistance genes evolved: several genes such as *Lr1* and *Lr10* are present in identical sequence already in tetraploid or diploid wheat progenitors, demonstrating their evolution in these relatives. In contrast, other genes such as *Lr34* and the 17 functional alleles at the *Pm3* locus evolved only after domestication/formation of hexaploid wheat as the specific resistance alleles were never found in genotypes outside the bread wheat gene pool. The triticeae crop species (such as wheat, rye and barley) also allow to study evolution of orthologous genes in relation to functional resistance. We found that the two powdery mildew resistance genes *Pm8* and *Pm17* introgressed from rye to wheat on chromosomal translocations are homologs of the *Pm3* gene in wheat. *Pm17* is a chimeric gene between a close relative of *Pm3* in rye and an unknown further homolog of *Pm3* in the rye genome. Thus, our studies have revealed a highly complex evolutionary origin of active resistance genes in wheat and have a number of implications for developing strategies to improve resistance genes and their use in breeding: first, there is a large potential in wheat relatives (wild and cultivated) to identify novel alleles useful for breeding. For example, it seems very promising to examine the rye gene pool for additional functional alleles of the *Pm8/Pm17* gene. Second, the evolution of novel resistance genes after domestication shows that novel genes can evolve in short evolutionary time periods. It is likely that additional useful variants can be engineered as natural evolution in agricultural ecosystems did not occur long enough to evolve all possible, useful variants. Thus, artificial design of genes as well as random laboratory evolution are promising ways to expand the natural repertoire of resistance genes which could then be used as transgenes or by genome editing.

W271: **Crop Evolution Genomics & Future Agricultural Productivity**

TBA

Harkamal Walia, University of Nebraska Lincoln, Lincoln, NE

W272: **Crop Evolution Genomics & Future Agricultural Productivity**

An Ecological Look at Crop Evolution

Ruben Milla, Associate Professor, Madrid, Spain

Crop species are a subsample of angiosperm plants. That subsample is not random, neither from phylogenetic nor from phenotypic viewpoints. The fact that crop progenitors were non-randomly selected from the pool of wild species might imply relevant consequences for their performance under agriculture, their future adaptability, or the services they can provide. In this contribution we will show phylogenetic patterns of crop species to test whether crops are highly or scarcely phylogenetically-clustered, and whether domestication events tended to occur more intensely on certain angiosperm clades. Also, we will explore phenotypic traits of crops in comparison to those of wild species to test if crops have particular phenotypic profiles. Finally, we will ask if crops and wild species deliver ecosystem services differently, through different ways of impacting soil functioning or plant performance in crop populations and communities.

W273: **Crop Evolution Genomics & Future Agricultural Productivity**

Domestication of Cereals and Origins of Agriculture in Western Asia

Jan Dvorak, Department of Plant Sciences, University of California, Davis, Davis, CA

Domestication of three cereals, einkorn wheat, emmer wheat, and barley, initiated the transition from hunting gathering to agrarian economy in western Asia and ultimately led to the evolution of western civilization. Genetic evidence has implicated the Kartal-Karadag region in the northwestern part of the Fertile Crescent, the Karakadag region in the northeastern part of the Fertile crescent, and southern Levant as putative sites of cereal domestication. Our previous work showed that introgression from domesticated emmer to wild emmer took place in all areas of present-day wild emmer distribution. Likewise, introgression from wild emmer into common wheat took place in the areas of their sympatry in western Asia. These and other observations suggest that gene flow between wild and domesticated wheat has been of a common occurence and raise legitimate concerns about the authenticity of populations of wild cereals and validity of genetic inferences based on them. Multi-locus genetic analyses placed the domestication of emmer into two geographic areas, the Karakadag region in Turkey and Israel. We hypothesize that if wild emmer in the Karakadag region was introgressed or naturalized from domesticated emmer, it may have acquired alleles from domesticated wheat at some of the domestication loci. This could be detected by comparing segregation in wild x cultivar populations. We tested this hypothesis by a QTL analysis of wild emmer accession PI 428082 collected in the Karacadag region. We mapped 64 QTLs, most of them for domestication traits, in a durum cv Langdon x PI 428082 RIL population. All domestication QTLs reported to segregate in mapping populations of wild emmer collected in Israel also segregated in this mapping population. Our findings therefore failed to provide evidence that wild emmer in the Karacadag region is introgressed or naturalized population. Possible scenarios reconciling of the origins of domesticated wheat will be briefly discussed.

W274: **Crop Genomics for Global Food Security**

Broadening Genetic Resources through Mutation Breeding for Global Food Security

Ljupcho Jankuloski, IAEA, Joint FAO/IAEA Division, Vienna, Austria

W275: **Crop Genomics for Global Food Security**

Mega-Trends Influencing Global Food Security

C. Lynne McIntyre, CSIRO, Brisbane, QLD, Australia
Increasing food production to ensure adequate food supplies for a growing population is difficult given climate change and a changing economy. Increased levels of greenhouse gases are likely to lead to changes in temperature, precipitation, CO2 and other climatic variables while costs of water, fertiliser and labour continue to rise. Adapting agriculture to a changing climate requires understanding the likely climate changes in current cropping and non-cropping regions and the impacts of climate change on crop growth and productivity. Adapting agriculture to a changing economy requires understanding new technologies and their role in ameliorating rising costs and enhancing sustainability. In this presentation, wheat and sugarcane will be used as examples of research that is investigating global agriculture megatrends: the likely changes in climate in the wheat and sugarcane growing regions of Australia, the effect in changes in climate variables on wheat and sugarcane growth and production, and the new genomic, phenomic and modelling approaches available to manage wheat and sugarcane production in a variable climate and changing economy. Despite the difficult genetics of both crops, especially sugarcane, advances in big data in both genomics and phenomics is predicted to have a major impact in wheat and sugarcane breeding and gains in productivity and profitability.

W276: **Crop Genomics for Global Food Security**
High Quality Genomic Resources to Understand Independent Domestication Processes and Genetic Diversity in Lima Bean
Jorge Duitama, Universidad de los Andes, Bogotá, Colombia

W277: **Crop Genomics for Global Food Security**
Speed Breeding Resilient Crops
Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

W278: **Crop Genomics for Global Food Security**
Conservation and Use for Global Food Security: Genomic Highlights in the CGIAR Clonal Crop Genebanks
Nicolas Roux[1], Mathieu Rouard[1], Julie Sardos[1], Rachel Chase[1], Michael Abberton[2], Peter Wenzl[3] and David Ellis[4], (1)Bioversity International, Montpellier, France, (2)International Institute of Tropical Agriculture, Ibadan, Nigeria, (3)CIAT, Cali, Colombia, (4)International Potato Center, Lima, Peru
The biggest challenges for agriculture in the 21st century are the increasing effects of climate change and an expanding world population, both significant threats to global food security. To meet these challenges, CGIAR aims to improve food security through its varied research programmes and platforms, while contributing to the United Nation's Sustainable Development Goals and targets for the 2030 agenda.
The CGIAR Research Programme on Roots, Tubers and Bananas (RTB) focuses on important staple crops that provide 15% or more of the daily per capita calorie intake for the 763 million people living in the least developed countries. However, these crops have characteristics, such as clonal reproduction and high genetic complexity, which require innovative approaches, including those focused on genomics to unlock their full potential.
Although whole genome sequences have been available for many years for potato (2011), cassava (2012) and banana (2012), yam (2017) and sweetpotato (2018) sequences were only recently published. Therefore, developments in understanding these crops are taking place at multiple stages.
Since the banana *Musa acuminata* was sequenced in 2012, high-throughput genotyping technologies were performed to characterize material from the Bioversity's International genebank, enabling linking genomic regions to phenotypic traits such as parthenocarpy. The development of such markers also allows the application of new approaches to study genome rearrangements and shed light on the evolution and emergence of triploid cultivars. Further to the publication of the improved reference sequence (2016), other wild species were sequenced, allowing clarification of their evolutionary relationships.
In potato, while the reference genome has been of extreme value, the complexity of cultivated potato has led to the need for a more complete understanding of genome structure through additional reference genomes and several pangenome projects. For characterization of collections, the International Potato Center (CIP) has fingerprinted the global collections of potato and sweetpotato, allowing increased understanding of this collection as well as comparison across all *ex situ* collections to identify unique germplasm and ensure its safe long-term conservation.
A first whole genome sequence of white yam *Dioscorea rotundata* has been released and this is being followed by *D. alata* along with re-sequencing in both crops. Fingerprinting of the collection at the International Institute of Tropical Agriculture (IITA) is well underway using DArTseq markers and is shedding light on diversity, duplication and trueness to type within the collection. The cassava collections at the International Center for Tropical Agriculture (CIAT) and IITA are being analysed together on the DArTseq platform to identify cases of cross-genebank duplication and to investigate how well different genepools are represented in either of the two genebanks.
Findings from RTB genomic research can impact how we conserve and use our genetic resources. By better understanding the mechanisms that lead to important traits such as yield or resistance to biotic and abiotic stresses, breeders can select the best parent material according to the most pressing needs. Farmers can then benefit by using newly available varieties to bring better food to the market and to the table.

W279: **Crop Genomics for Global Food Security**
The use of RenSeq Technology in Crop Productivity
Gabriel S. C. Alves, Laboratório de Microbiologia: Interação Planta-Praga, Bloco I-1-35/8, Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Campus Universitário Darcy Ribeiro,, Brasília, Brazil

W280: **CSSA: Translational Genomics**
Translational Genomics Comes of the Age for Legume Improvement in Developing Countries
Rajeev K Varshney, ICRISAT, Hyderabad, India
Legume crops such as chickpea, pigeonpea and groundnut contribute to livelihood as well as human nutrition and are mostly grown in semi-arid regions in many Asian and African countries. Exposure of these crops to different biotic/abiotic stresses in marginal environments results in low

crop productivity. Recent advances in sequencing, automation/robotics and computational biology have started an era of –omics sciences in legumes. A number of –omics approaches have been deployed to understand the genome architecture, genome diversity and complexity of trait. Modern trait mapping approaches have been used to map a number of agronomic traits. Molecular markers and genes identified through various approaches have been used to enhance precision and efficiency of breeding programs. A number of legume lines with improved traits related to production constraints as well nutrition developed through molecular breeding are in advanced stage of field trials in India and Ethiopia. In summary, our efforts highlight the role of translational genomics for improving agriculture in marginal environments. In my opinion, accelerated and coordinated efforts in the area of translational genomics as well as fast-track release policy of molecular breeding products in combination of better agronomy will be helping delivering faster genetic gains in farmers fields.

W281: CSSA: Translational Genomics
Genetic Improvement of Climate Resilient Cowpea for Food and Nutritional Security
Bao Lam Huynh, Philip A. Roberts and Timothy J. Close, University of California, Riverside, Riverside, CA
Cowpea (*Vigna unguiculata* L. Walp.) is an important staple food crop in sub-Saharan Africa and also grown in other warm-to-hot regions worldwide. Like most legumes, cowpea is nutritious as a source of protein, energy and essential nutrients, making it a desirable target for enhancing food and nutritional security. Drought, heat, low fertility soils, pests and diseases are key factors causing traditional varieties in Africa to yield much lower than their potential. Collaborative breeding programs aiming to develop improved cowpea cultivars for Africa and the USA are enabled by the application of genomic resources including SNP genotyping arrays, genetic maps, QTLs, and the decision support tools of the Integrated Breeding Platform. These programs have been supported in large part by grants from Generation Challenge Programme, Feed the Future Innovation Labs for Collaborative Research on Grain Legumes and Climate Resilient Cowpea, and the California Dry Bean Board. Progress in modern breeding activities will be presented and discussed, with an emphasis on MAGIC breeding as part of the ongoing NSF-BREAD project “Advancing the Cowpea Genome for Food Security”.

W282: CSSA: Translational Genomics
Models and Methods for Increasing Genomic-Enabled Prediction in Plant Breeding
José Crossa[1], Osval A Montesinos-Lopez[2], Abelardo Montesinos-Lopez[3], Daniel Gianola[4], Paulino Perez-Rodriguez[5], Philomin Juliana[6], Ravi Prakash Singh[7], Yoseph Beyene Aydagn[8], Manje Gowda[9], Xuecai Zhang[6] and Michael Olsen[9], (1)International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico, (2)Universidad de Colima, Colima, Mexico, (3)University de Guadalajara, Guadalajara, Mexico, (4)Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison, Madison, WI, (5)Colegio de Postgraduado, Texcoco, Mexico, (6)CIMMYT, Mexico City, Mexico, (7)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (8)CIMMYT, NAIROBI, Kenya, (9)CIMMYT, Nairobi, Kenya
In the last years several genomic-enabled prediction models have been developed for the prediction of large number of unobserved phenotypes in different environments using dense molecular markers. Models include single and multi-traits, and single and multi-environments aiming to increase the prediction accuracy of the primary trait 'gran yield' on unobserved individuals. Increase in prediction accuracy over the genomic best linear unbiased estimator (GBLUP) are achieved by means of the Gaussian kernel model with genomic × environments interaction. A Bayesian multi-trait multi-environment model was described and used to efficiently exploit correlated traits and environments; this model has increased the prediction accuracy of about 10-20% over the single trait and single environment model. However, due to the nature of the estimation process this Bayesian model requires intense computing resources and running time is significantly extended. In an attempt to speed up the prediction of extensive number of unobserved phenotypes in large sets of multi-environments (big data) we have been developing Deep Machine Learner (DL) models and methods. DL models with densely connected network architecture were compared with GBLUP on nine real genomic data sets. The DL appeared to be competitive, since they were better than GBLUP in 4 of the 9 data sets under a scenario that ignored covariates capturing genomic x environment interaction. However, when genomic × environment interaction was included, DL was inferior in terms of prediction accuracy to the GBLUP model. Authors have extended the multi-environment DL to the multi-trait DL case and found that among models excluding the genomic × environment interaction, the multi-trait DL model was the best, while among models including genomic × environment interaction the Bayesian multi-trait multi-environment was superior. Although implementing the multi-trait DL models is feasible and practical in the genomic prediction context it is challenging due to the large number of hyper-parameters involved. Recent results on the application of genomic prediction vs phenotypic prediction in maize will be shown.

W283: CSSA: Translational Genomics
Large-Scale GWAS and CRISPR in Maize
Jianbing Yan, Huazhong Agricultural University, Wuhan, China

W284: CSSA: Translational Genomics
Activities and Specificities of CRISPR/Cas9 and Cas12a Nucleases for Targeted Mutagenesis in Maize
Kan Wang, Iowa State University, Ames, IA

W285: CSSA: Translational Genomics
Developing Improved Crops using Genome Editing
Doane Chilcoat, Corteva Agriscience, Agriculture Division of DowDuPont, Johnston, IA
Recent breakthroughs in reference genomics, plant regeneration, and CRISPR/Cas enable plant genome editing for crop improvement. We have developed a cutting edge toolkit, including methods for elite line plant regeneration, novel methods for characterizing off-site cutting and characterization of a number of novel Cas9 orthologues with diverse PAM sequences. We are using plant genome engineering to improve a number of high value crops by improving yield potential, changing grain composition, and enhancing disease resistance. Our most advanced genome edited product, CRISPR-Waxy corn, has successfully completed the development process in less than four years.

W286: **Cucurbit Genomics**

Whole-Genome Resequencing Study of the USDA Collection of *Citrullus amarus* Plant Introductions

Sandra E. Branham[1], Amnon Levi[1], Shan Wu[2], Zhangjun Fei[2], Alvaro G. Hernandez[3] and Patrick Wechter[1], (1)USDA-ARS, Charleston, SC, (2)Boyce Thompson Institute, Cornell University, Ithaca, NY, (3)University of Illinois at Urbana-Champaign, Urbana, IL

Sustained selection pressure for desirable horticultural characteristics during domestication has narrowed the genetic diversity of cultivated watermelon (*Citrullus lanatus* L.), resulting in a loss of genes/alleles conferring disease resistance. 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. Multiple independent evaluations of the USDA *Citrullus* plant introduction (PI) collection for resistance to various diseases and pests have identified resistant *C. amarus* germplasm, with a number of these screens failing to find resistance in any other *Citrullus* species. Despite the vital importance of *C. amarus* as a resource for improvement of cultivated watermelon, no genetic studies of the USDA collection of *C. amarus* have been reported. Here we present the analysis of a whole-genome resequencing study (20X coverage) of all available *C. amarus* PIs (N=133) from the USDA germplasm collection. Over 24 million genetic variants were identified and used to examine the genetic diversity, relatedness and population structure of citron melon.

W287: **Cucurbit Genomics**

Gene Expression Dynamics of Age-Related Resistance of Cucumber to *Phytophthora capsici*

Ben N. Mansfeld, Michigan State University, East Lansing, MI and Rebecca Grumet, Department of Horticulture, Michigan State University, East Lansing, MI

Cucumber (*Cucumis sativus*) fruit are largely susceptible to infection by *Phytophthora capsici.* However, some cucumber cultivars develop a fruit surface-associated age-related resistance (ARR) to *P. capsici.* Young, rapidly growing fruit are highly susceptible, but become resistant as they complete exponential growth [~16 days post-pollination (dpp); 2-3 weeks prior to ripening]. Analyses of peel of ARR expressing and non-expressing uninoculated fruit identified gene expression and metabolomic changes associated with resistance possibly functioning as preformed defenses. We further performed transcriptomic analyses of inoculated fruit at resistant (16 dpp) and susceptible (8 dpp) ages, providing a unique opportunity to examine compatible and incompatible interactions in the same genotype. Strong transcriptional changes were observed at 4 hours post inoculation (hpi), with approximately 2000 genes differentially expressed in either age, suggesting an early initial response to infection. At 24 and 48 hpi, susceptible 8 dpp fruit continued to mount defense along with strong downregulation of genes involved in photosynthesis, cell wall synthesis and modification, lipid and cuticle biosynthesis, cell division and growth. In contrast, resistant 16 dpp samples largely downregulated defense responses while upregulating photosynthesis and other biological processes. Together these results suggest that in ARR-expressing fruit, a successful defense is mounted within the first 24 hours. To understand the dynamics of infection during the first 24 hours, inoculated and control samples were collected at 0, 2, 4, 8, 12, 18, 24 hpi, from both resistant (16 dpp) and susceptible (8 dpp) fruit. 3'RNAseq was performed to assay gene expression and weighted co-expression networks will be analyzed to identify key regulatory hubs contributing to resistance during the first 24 hours.

W288: **Cucurbit Genomics**

An Old Dog Plays New Tricks: A Loss-of-Susceptibility Mutation of the *STAYGREEN* Gene Confers Durable Resistance against Multiple Pathogens in Cucumber

Yiqun Weng, USDA-ARS, Vegetable Crops Research Unit, Madison, WI

In the U.S., many modern cucumber varieties developed since 1960s' carry the *dm1*, *psl*, and *cla* loci for resistances against the oomyceteous downy mildew (DM, pre-2004 strains), the bacterial angular leaf spot (ALS) and the fungal anthracnose (AR) pathogens, respectively, all of which could be traced back to the plant introduction line PI 197087 from India. However, the underlying molecular mechanisms for the multiple disease resistance are unknown. Here, I will report results from our QTL mapping and cloning studies for DM, AR and ALS resistances in the Gy14 and WI 2757 cucumber inbred lines. We show that the triple-disease resistances in the two lines were controlled by the same locus, *STAYGREEN* (*CsSGR*) which encodes the magnesium dechelatase and plays a critical regulatory role in senescence-inducible chlorophyll degradation. Thus, host resistance conferred by *CsSGR* represents a novel function for this highly conserved gene in flowering plants. We found that the multiple-resistance was due to a SNP in the coding region of *CsSGR* that resulted in a nonsynonymous amino acid substitution in the CsSGR protein. Investigation of expression patterns of different genes in the chlorophyll degradation pathway in DM resistant and susceptible lines in response to pathogen inoculation suggested involvement of this pathway in host resistance. Association analysis in natural and breeding populations indicated that *CsSGR* has undergone selection in cucumber breeding. We conclude that *CsSGR* belongs to the loss-of-susceptibility type resistance gene; a model was proposed to explain the mechanisms of *CsSGR*-mediated multiple disease resistances.

W289: **Cucurbit Genomics**

Quantitative Trait Loci Associated with Resistance to Bacterial Fruit Blotch in *Citrullus amarus* USVL246-FR2.

Patrick Wechter, Sandra E. Branham and Amnon Levi, USDA-ARS, Charleston, SC

Acidovorax citrulli, the causal agent of bacterial fruit blotch (BFB) of cucurbits, has the potential to devastate production of watermelon and other cucurbits. Despite decades of research on host-plant resistance to *A. citrulli*, no germplasm has been found with immunity and only a few sources with moderate to high levels of BFB resistance have been identified, but the genetic basis of resistance in these watermelon sources are not known. The majority of these resistant sources are *Citrullus amarus* (citron melon) plant introductions, a species that crosses readily with cultivated watermelon (Citrullus lanatus L.). Working with a recombinant inbred line population derived from a cross between two *C. amarus* lines, BFB resistant (USVL246-FR2) and BFB susceptible (USVL114), we performed QTL analysis to identify QTL associated with resistance. QTL mapping of affected leaf area identified six QTL that each explained between 5 and 15% of the variation in BFB resistance in the population. This study represents the first identification of QTL associated with resistance to *A. citrulli* in any cucurbit.

W290: **Cucurbit Genomics**

Chromosome Evolution in Cucurbitaceae using a Reference-Quality Genome Assembly of Bitter Gourd (*Momordica charantia*)

Henri van de Geest[1], Peter van Dam[1], Almer van Belzen[1], Oskar Bruning[1], Mark de Heer[1], Renske A. Kuiper[1], Ferry Wijnhoven[1], Elio Schijlen[2], Linda V. Bakker[2], Venus Salutan[3], Eliot Cline[4], Darush Struss[4], Gabino Sanchez Perez[1], Jorn de Haan[1], Nikkie van Bers[1], Bart Nijland[1] and **Lidija Berke**[1], (1)Genetwister Technologies B.V., Wageningen, Netherlands, (2)Plant Research International, Wageningen University & Research, Wageningen, Netherlands, (3)East West Seed, Lipa City, Philippines, (4)East West Seed, Chiang Mai, Thailand

Bitter gourd (*Momordica charantia*) is a cucurbit species whose bitter-tasting fruits are of high economic significance in South-East Asia. Phylogenetically, it has a crucial position as an outgroup to the rest of the already sequenced cucurbit species (cucumber, melon, pumpkins, watermelon and bottle gourd) which enables evolutionary studies of traits and genome structure in cucurbits.

Genetwister Technologies in collaboration with East West Seed has generated a reference-quality assembly of bitter gourd. The PacBio-based assembly is 297Mb in size and comprises 132 contigs with the contig N50 statistic of 13Mb. It was next scaffolded by the BioNano Saphyr genome mapping technology with two nicking/labeling enzymes. Remarkably, the scaffolding step resulted in 11 chromosomes that contain more than 98% of the assembled sequence and were validated with a genetic map. Annotation incorporated both Iso-Seq and tissue-specific Illumina RNA-seq reads.

While some genomic resources for bitter gourd are available, our reference-quality genome assembly is for the first time able to answer questions about chromosomal rearrangements and duplications, reconstruction of the ancestral cucurbit genome, determine lineage-specific gene family expansions and large-scale genome evolution not only for bitter gourd but also of the entire cucurbit clade.

W291: **Cucurbit Genomics**

Genomewide Association Studies in Watermelon and Functional Validation

Umesh K. Reddy, Department of Biology, West Virginia State University, Institute, WV

Genotyping by sequencing was used to identify a total of 11,969 SNPs with minor allele frequency ≥0.05 and a call rate ≥70% for 190 watermelon accessions consisting of US Plant Introductions (PIs) and cultivars collected throughout the world. To identify genes that control trichome density and trichome length, 156 watermelon accessions were used for phenotyping during two different seasons followed by genome-wide association mapping. We identified significant SNP markers for showing different magnitude of associations with variation involving trichome length and trichome density. Ten Arabidopsis genes orthologous to watermelon genes were identified to be in association with the trichome traits. Similarly, GWAS was performed for citrulline variation among these collections that enabled us to identify two candidate genes that are involved in citrulline pathway. A third GWAS was performed for flavor related compounds including hexanal; 2-hexanal; octanal; 2-Decenal, Z-; 5-Heptenal, 2,6-dimethyl-; 1-Heptanol, 6-methyl- ;5-Hepten-2-one, 6-methyl-; Nonanal; 4-Nonenal, E-; 3-Nonen-1-ol, Z-; 2-Nonenal, E-; 3,6-Nonadien-1-ol, E,Z- and 2,6-Nonadienal, E,Z-. Functional validation and annotation of various genes will be presented.

W292: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

Get a Grip on Your Data Science Tools with Cyverse Vice (Visual Interactive Computing Environment)

Upendra Devisetty[1], Tyson Swetnam[2], Ian McEwen[2], John Wregglesworth[2] and Nirav Merchant[2], (1)University of Arizona, CyVerse, Tucson, AZ, (2)CyVerse, Tucson, AZ

How can you process 100GB of data using computational notebooks like Jupyter or Zeppelin, Rstudio and the Shiny apps or visualize a plot with millions of data points in your browser? How do you securely share these apps, data, and analyses with your collaborators and community? To effectively use these amazing tools with real-world data requires underlying computational resources (CPU, RAM, storage) beyond what is typically available on laptops/desktops or through hosted platforms like shinyapps.io and mybinder.org. Additionally, installation for these tools is often challenging and requires fairly specific versions of software libraries.

Visual Interactive Computing Environment (VICE) is a feature of CyVerse's data science workbench, called the Discovery Environment (DE). VICE provides easy to use access to RStudio, Shiny, Jupyter lab or any web-based application. Users can readily augment existing VICE apps by installing additional bioinformatics packages (over 4500 available) from Bioconda project (https://anaconda.org/bioconda/). Communities can customize their VICE app with requisite software tools for their analysis pipeline, thus providing a consistent analysis environment for performing reproducible analysis for anyone utilizing their apps. All of these resources are securely accessible via a web browser. In addition, all of the VICE tools will be available as normal DE apps for anyone wishing to do the non-interactive analysis. Some of the popular tools that are currently integrated as VICE apps include - Jupyter lab, Rstudio, Shiny, QIIME2, TBAS etc.,

Users can analyze data and host their results from large and complex analysis workflows without having to move large quantities of data. As a fully integrated piece of the CyVerse data management platform, VICE handles transferring data and returning results back to the Data Store. CyVerse users can take advantage of many existing VICE-based applications, create their own VICE application or request creation of a customized version. New users can sign up for free CyVerse account and can start running their interactive analysis right away.

W293: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

Reproducible Analysis of Microbiome Data on CyVerse's Atmosphere

Joslynn Lee, Howard Hughes Medical Institute, Chevy Chase, MD

CyVerse's Atmosphere is an open-source platform that allows researchers to launch their own virtual machine to access high-performance computing (HPC). Running command line based bioinformatics tools can have a large learning curve and even require substantial memory. How can we empower researchers to work with their own data using HPC resources? To address this, a guided reproducible computational workflow was developed on Atmosphere to run a microbiome analysis in Jupyter Notebooks. Using this for workflow for research is only the beginning as computational concepts (algorithms and file formats), statistics, accessing genomic data and running bioinformatics tools to analyze data are useful to introduce in the undergraduate level classroom.

W294: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

Investigating the Functional Landscape of Long Non-Coding RNAs in Spermatogenesis

Sateesh Peri, University of Nevada, Reno, Reno, NV

Spermatogenesis is a complex biological process that is strictly regulated by a large number of genes, including coding genes and non-coding RNAs. Environmental toxicants such as DDT have been shown to induce epigenetic trans-generational inheritance of disease (e.g., obesity) through the germline by affecting the process of spermatogenesis. However, the mechanism of transmission of these epigenetic marks remains a mystery with factors such as DNA methylation, long-non-coding RNAs (lncRNAs) potentially contributing to the phenomenon. RNA sequencing technology has made identification of lncRNAs relatively affordable and easy over the past decade. Our lab and several other groups have reported a repertoire of lncRNAs in sperm previously but, their functional role in spermatogenesis and epigenetic trans-generational inheritance remains yet to be understood. In this current study we have undertaken the identification and functional characterization of lncRNAs during several developmental stages of sperm using 'self-organizing maps'. We have used compute resources from CyVerse for this study specifically for: i) Data storage and sharing ii) Read mapping and transcript assembly iii) Identification of lncRNAs iv) Differential expression analysis.

W295: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

The Barcode Long Island Project as an Introduction to Scientific Research

John Halloran, Connetquot HS, Bohemia, NY

Authentic research experiences for high school students have been shown to improve critical thinking skills and increase student engagement in STEM fields. For teachers, implementing these experiences can be daunting, with numerous obstacles to successful student experiences. For the past five years, I have used Barcode Long Island (BLI) as a gateway program to teach students how to design, implement and complete a research project. Through this approach, students have experienced the process of scientific research while improving their critical thinking skills. Their research has led to the discovery of novel GenBank sequences, the detection of invasive species, and increased enrollment in advanced classes and internships.

W296: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

Using DNA Barcoding to Validate and Augment Observation-Based Biodiversity Studies

Alison Dell, St. Francis College, Brooklyn, NY

Crowdsourced and mobile apps are a growing part of citizen science efforts but when non-experts rely on apps for biodiversity studies, how reliable are the data? We used DNA barcoding to complement a plant biodiversity survey along New York City's waterway Newtown Creek, a Superfund site which is slowly recovering; in large part due to local environmental efforts. This survey, led by NGOs Newtown Creek Alliance and Hudsonia; relied on citizen scientists recording and cataloging their observations using the app *iNaturalist*. With each photographed observation, the app suggests possible identities for the plant. The user selects the best match, which can then be reviewed/verified by other members of the *iNaturalist* community. In addition to recording observations using *iNaturalist*, participants collected samples of each plant for an herbarium. This study employed DNA barcoding to both validate and extend this study using random quadrant sampling to select plants for recording, observation and collection. Sampling was performed in both spring and fall. A portion of each herbarium sample was used for DNA barcoding, and the resulting sequence was compared to suggestions by the *iNaturalist* app and its user community. Out of the fourteen samples initially collected; we found that *iNaturalist's* unaided prediction disagreed with experienced users about the identity of the samples in 64% of the cases (9/14). DNA barcoding supported the identification made by *iNaturalist* users in 6 of these 9 cases, with the remaining three samples identifying a different plant. Of our limited sample size, we never observed that the DNA evidence supported the app's eye over that of the *iNaturalist* user. However, when *iNaturalist* prediction's matched those of human users, DNA barcoding validated these findings. Before spring sampling, survey participants underwent a training session. Consensus in observation increased after this training. These findings underscore the critical role for human experts in crowdsourced biodiversity efforts. Resources like *iNaturalist* are an invaluable tool for community science; as long as local expertise is available to verify findings. DNA barcoding can be used in tandem with observation based biodiversity studies in training an app or when morphological identification is difficult for both humans and computers.

W297: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

FAIR (Findable, Accessible, Interoperable, Reusable) Data and Compute with iMicrobe

Ken Youens-Clark, University of Arizona, Tucson, AZ

The iMicrobe platform brings together analysis tools and microbiome data sets by leveraging National Science Foundation-supported cyberinfrastructure and computing resources from CyVerse, Agave, and XSEDE. The primary purpose of iMicrobe is to provide users with a freely available, web-based platform to: (1) maintain and share project data, metadata, and analysis products, (2) search for related public datasets, and (3) use and publish bioinformatics tools that run on highly-scalable computing resources. iMicrobe is not a data repository but rather a search engine that helps users find and analyze data from disparate repositories. Analysis tools are implemented in containers that encapsulate complex software dependencies and run on XSEDE resources via the Agave API. Taken together, iMicrobe promotes data integration, sharing, and community-driven tool development by making open source data and tools accessible to the research community in a simple web-based platform in accordance with FAIR principles.

W298: **CyVerse - Software, Tools, and Services for Data-Driven Discovery, Data Science, and Education**

Community Cyberinfrastructure Enabled Platform for Image-Based Phenotying and Analysis

Cory D. Hirsch[1], Nathan Miller[2], Roman Woodward[1], Tara A. Enders[3], Nathan M. Springer[1] and Edgar P. Spalding[2], (1)University of Minnesota, St. Paul, MN, (2)University of Wisconsin-Madison, Madison, WI, (3)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN

Image-based plant phenotyping requires data acquisition and analysis, which are inherently coupled together. Continued community efforts have focused on development of software, hardware, and languages to increase the throughput of imaging methods. Often, this has led to high-performance phenotyping focused on as much automation as possible, which results in specialized and centralized data gathering hardware. The

price and expertise associated with this limits widespread community adoption and supports a hub-and-spoke service model. Alternatively, distributed high-throughput phenotyping focuses on autonomy and demands a solution easily implemented, resulting in a decentralized network of users. To accomplish this, we designed a platform focused on stability, flexibility, and accessibility in metadata gathering, trait extraction, and analysis. In our system, plants are grown in a controlled environment setting and manually placed against a backdrop that includes a machine-readable user defined metadata sheet that is embedded in the image. Side-view RGB images are automatically acquired with consumer-grade cameras and can be pushed to NSF-funded high-throughput computing resources on CyVerse-managed data-storage infrastructure. Using a CyVerse enabled application, machine learning algorithms robustly segment each image into background, plant, and metadata components. The latter two components are analyzed with custom software that executes without user interaction and in parallel. The method was developed to measure maize growth and architecture and quantifies traits from images such as height, width, stem diameter, and center of mass. Prototypes of the entire platform have been used to quantify growth and architecture phenotypes of mapping populations of maize subjected to abiotic stress across multiple institutions. In addition, we have also developed an R Shiny application for initial data quality control and analysis. In all, the metadata sheet creation, trait extraction, and analysis functions are deployed on high-throughput computing resources accessed via the CyVerse Web interface. The simplicity of the image-acquisition hardware and the web-based trait and analysis functions make this image-based phenotyping method broadly accessible.

W299: Database Resources for Crop Genomics, Genetics and Breeding: NRSP10

Overview of NRSP10 - Achievements and Future Plans

Dorrie Main[1], Sook Jung[1], Cameron Peace[1], Jim McFerson[2] and Michael Kahn[1], (1)Washington State University, Pullman, WA, (2)Washington State Univesity, Wenatchee, WA

National Research Support Project 10(NRSP10,www.nrsp10.org) is a USDA NIFA, US Land Grant Universities, and industry funded project which provides standardized database and informatic resources for undeserved or specialty crops such as tree fruit, nuts, cotton, and berries. It builds on existing database resources developed for Rosaceae (Genome Database for Rosaceae), Citrus (Citrus Genome Database), Vaccinium (Genome Database for Vaccinium), Cool Season Food Legumes (Cool Season Food Legume Genome Database)and Cotton (CottonGen). Developed using Tripal, an open-source, resource-efficient, modular, well supported platform, these community databases provide centralized access to integrated genomic, genetic and breeding data and analysis tools for 25 crops representing a combined annual production value of over $26.6 Billion. We highlight the broader impacts of the current projects and driven by the research community it serves, we highlight plans for the next 5-year project, which focuses on providing tools and analysis capability to manage and utilize big data for both discovery and crop improvement, whilst leveraging funding from multiple sources as we develop solution for sustainability of these research-enabling resources.

W300: Database Resources for Crop Genomics, Genetics and Breeding: NRSP10

Using the Online GenSAS Platform for Community Genome Annotation

Taein Lee[1], Jodi L. Humann[2], Stephen P. Ficklin[3], Chun-Huai Cheng[1], Heidi Hough[2], Sook Jung[2], Jill L. Wegrzyn[4], David B. Neale[5] and Dorrie Main[2], (1)Washington State University, Pullman, Pullman, WA, (2)Washington State University, Pullman, WA, (3)Dept of Horticulture, Washington State University, Pullman, WA, (4)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (5)Department of Plant Sciences, University of California, Davis, Davis, CA

The Genome Sequence Annotation Server v6.0 (GenSAS, www.gensas.org) is a web-based annotation platform that allow for collaborative, community genome annotation of model and non-model organisms. GenSAS utilizes a series of user-friendly interfaces with embedded instructions and only requires a user account and internet access to annotate prokaryote and eukaryote genomes. The GenSAS annotation process includes structural annotation with a variety of gene model prediction tools, functional annotation, and manual curation via integrated JBrowse and Apollo. After annotation is complete, GenSAS generates the required files for publication and submission to a database. In addition to the current GenSAS platform, a Tripal compatible module called GenSAS Community, is in development. GenSAS Community will allow for Tripal crop databases to manage large, community powered, manual annotation curation projects using JBrowse/Apollo and several project coordination tools. The module will use different roles from Curator to Lead Curator to manage who is assigned to which section of the genome and which section is curated and ready for final approval. Approved annotations will then be loaded into the crop database for use by the community.

W301: Database Resources for Crop Genomics, Genetics and Breeding: NRSP10

Tripal v3, the Collaborative Online Database Platform for Genomic, Genetic and Breeding Databases

Stephen P. Ficklin, Dept of Horticulture, Washington State University, Pullman, WA, Bradford Condon, University of Tennessee, Knoxville, TN and Margaret Staton, University of Tennessee, Knoxville, Knoxville, TN

Tripal is an open-source software platform for building online community or project databases that house genomic, genetic and breeding data. A number of specialized community databases that use Tripal are tailored for horticultural and agricultural crops such as fruits, nuts, legumes, cotton, and forest trees. With active code contributors from 12 research groups in 3 countries, Tripal has emerged as a model of cooperative database development across specialty crops and as a mechanism for increased sustainability of community-level and community-built web resources. Based on the content management system Drupal, Tripal enables developers to easily write their own custom code and share with others. The community is building the primary infrastructure to support standardized biological data storage formats, intuitive data visualization, and commonly needed analysis tools. With increasing maturity of the software and a growing number of member databases, Tripal can now take advantage of the shared code base across groups by building cross-site interfaces that unify data across various specialty crop communities. This is largely enabled by the latest version of Tripal v3, a fully ontology-driven design with data structures and RESTful web services. With the new major expansion of the Tripal module that leverages ElasticSearch, sites are able to provide comprehensive full text search to users and also incorporate search results from other public Tripal databases. For example, the Hardwood Genomics Project can return relevant search results from its own data stores as well as results from other sites with tree data such as TreeGenes, the Genome Database for Rosaceae and the Citrus Genome Database. The growth of the Tripal community demonstrates how to drive advances in large scale cyberinfrastructure development and data integration through collaborations among smaller, specialized research communities.

W302: **Database Resources for Crop Genomics, Genetics and Breeding: NRSP10**
Using NRSP10 Data and Tools for Basic Research
Sook Jung[1], Chun-Huai Cheng[2], Ping Zheng[1], Heidi Hough[1], Jodi L. Humann[1], Jing Yu[1], Dorrie Main[1] and Katheryn Buble[3], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Washington State University, Pulmman, WA
NRSP10 provides many resources for basic research .....

W303: **Database Resources for Crop Genomics, Genetics and Breeding: NRSP10**
Using NRSP10 Data and Tools for Translational Research
Cameron Peace[1], Sook Jung[1], Katheryn Buble[2], Chun-Huai Cheng[3], Heidi Hough[1], Jodi L. Humann[1], Ping Zheng[1], Jing Yu[1] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pulmman, WA, (3)Washington State University, Pullman, Pullman, WA
DNA tests for traits of interest for genetic crop improvement are a key need for translating genomics research outcomes into breeding practice. DNA tests (trait-predictive markers) target particular Mendelian or quantitative trait loci (QTLs) to reveal genotypes that predict phenotype. NRSP10 databases streamline DNA test development and refinement for rosaceous, citrus, vaccinium, cotton and cool season food legume crops. Typically, a researcher wanting to develop a new DNA test starts with knowledge of the genomic location of a trait locus of interest and its genotypes for two or more cultivars (such as parents of the QTL discovery mapping population). The database is then consulted. For simple PCR-based DNA tests, first, the QTL position is zoomed into (MapViewer; JBrowse) and nearby DNA sequence obtained (Sequence Retrieval; Download Sequence File). Next, motifs of possible polymorphism are sought and primers are designed to flank them to produce amplicons of desired lengths (Primer3). Primers are checked to ensure they only amplify the target locus (BLAST). Sets of primers representing candidate assays can be commercially procured and then PCR conducted to examine segregation of amplicons and their match to known cultivar genotypes. DNA test primers reported from other research groups can be refined by adding a longer “GC clamp” to enhance annealing (Primer3) and their genomic position checked in the crop's latest genome assembly (BLAST; JBrowse). Comparative mapping can fast-track new DNA tests development, exploiting genomes synteny (Synteny Viewer). Further opportunities beckon to enhance genomics translation into breeding tools and knowledge.

W304: **Database Resources for Crop Genomics, Genetics and Breeding: NRSP10**
Using NRSP10 Data and Tools for Breeding
Ksenija Gasic[1], Taein Lee[2], B. Todd Campbell[3], Rebecca McGee[4], Sook Jung[5], Jodi L. Humann[5], Jing Yu[5], James Crabb[6], Heidi Hough[5] and Dorrie Main[5], (1)Clemson University, Clemson, SC, (2)Washington State University, Pullman, Pullman, WA, (3)USDA-ARS, Florence, SC, (4)USDA-ARS, Pullman, WA, (5)Washington State University, Pullman, WA, (6)Department of Horticulture, Washington State University, Pullman, WA
Advances in sequencing, sensor, drone and computational technology have led to increasing volumes of genotype and phenotype data being collected and tracked by modern breeding programs. To efficiently store, manage and integrate these large private and public research data sets, so breeders can use them efficiently in decision-making, we are developing the Tripal Breeding Information Management System (BIMS). BIMS is available in the rosaceae, citrus, cool season food legume, vaccinium and cotton NRSP10 Databases. It allows breeders to create and manage access to their breeding programs; upload phenotype data from the FieldBook App or Excel templates; upload genotype data; generate input files for the FieldBook App; archive their entire data to their own computers; search and filter by accessions/lines name, trial, location, cross, parent and trait values; and perform basic statistical analysis. BIMS is being developed in collaboration with public plant breeders. In this presentation we highlight current functionalities available in BIMS and demonstrate how it is being used to improve efficiency in a peach breeding program.

W305: **Database Resources for Crop Genomics, Genetics and Breeding: NRSP10**
Global Performance Prediction Tool in Tripal BIMS
Craig Hardner[1], Cameron Peace[2], Sook Jung[2], Taein Lee[3] and Dorrie Main[2], (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia, (2)Washington State University, Pullman, WA, (3)Washington State University, Pullman, Pullman, WA
Breeders Toolbox, initially created as part of the RosBREED program, provided a system to manage pedigree, phenotypic, and genotypic data from a breeding program. As breeding programs have begun to use information on large-effect trait loci and genome-wide predictions for selection of parents, seedlings, and advanced selections, the breeders toolbox is being extended into the Tripal platform as Breeding Information Management System (BIMS) to incorporate new types and amounts of genetic data, as well as to allow breeders to manage their own data. Horticultural tree crop breeding programs tend to be locally focused and there has generally been limited evaluation of the suitability of advanced selections across a broad range of target commercial environments. We propose to extend functionality of this publicly funded Tripal BIMS to support the evaluation of environmental stability of germplasm on a global scale, initially for horticultural crops. Our hypothesis is that a particular phenotype of an individual is a sample of its response to the environment to which it has been exposed, and SNP genotyping can track replicated genomic segments across otherwise unconnected germplasm trials. Our vision is that data from different sources can be compiled into an anonymous database that individual users can interact with to input genotype and phenotypic data and output performance predictions across the range of environments in the dataset.

W306: **Database Resources for Crop Genomics, Genetics and Breeding: NRSP10**
Open Discussion of NRSP10
Dorrie Main, Washington State University, Pullman, WA

In this open discussion session following the presentations of NRSP10 resources, we will have an open discussion about current functionality, new functionality needed and future plans for NRSP10.

W307: Data Resource Sustainability and Funding

Sustainability Lessons Learned at Phoenix

Eva Huala, Phoenix Bioinformatics, Fremont, CA

The TAIR database (www.arabidopsis.org) has served as a community database for Arabidopsis researchers since 1999. In 2014, NSF funding for the project came to an end and a new support model was established under the umbrella of Phoenix Bioinformatics, including subscription support from academic libraries and companies combined with a significant degree of open data access. Open data policies at TAIR include public release of data after one year, free access for students enrolled in courses, and metered access enabling occasional use of the resource without requiring a subscription. Recently we have extended this model to several additional databases. The impact of this innovative new sustainability model on funding, usage and data reuse will be presented.

W308: Data Resource Sustainability and Funding

Tripal Database Sustainability through Collaboration, Research, and New Tools

Margaret Staton[1], Abdullah Almsaeed[1], Bradford Condon[1], Ming Chen[1], Joseph Benjamin West[1], Amanda Devine[1], Casey Richards[1], Patrick Sisler[1], Raymond Senu[1], Christopher Childers[2], Monica Poelchau[3], Jill L. Wegrzyn[4], Dorrie Main[5] and Stephen P. Ficklin[6], (1)University of Tennessee, Knoxville, Knoxville, TN, (2)USDA/Agricultural Resarch Service/National Agricultural Library, Beltsville, MD, (3)National Agricultural Library, Fort Collins, CO, (4)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (5)Washington State University, Pullman, WA, (6)Dept of Horticulture, Washington State University, Pullman, WA

The Hardwood Genomics Database (hardwoodgenomics.org) is built on Tripal, a generic and flexible software system for building community databases for genetics, genomics and breeding data. Our development plan for the previous 10 years has relied exclusively on federal funds, but new and more diversified models of funding and sustainability are needed for future maintenance and expansion. The primary costs are sustaining a team who can contribute to core software development and maintenance, develop new code modules, and curate and load new data. We have employed professionals, postdoctoral associates, graduate students, and undergraduate students for these tasks. Our emerging sustainability strategy has three facets. First, we are reducing costs by sharing code development across different databases and involving more students in biocuration. Second, we are developing basic research ideas and pilot projects outside of, but dependent on, the database. These projects will form the foundation for new grant submissions. Third, we are moving beyond databases to develop software tools that have potential for independent streams of funding from nonprofit companies, for profit companies, or subscriptions from individuals. Our successes, failures, and course adjustments thus far will be discussed.

W309: Data Resource Sustainability and Funding

TreeGenes: Past, Present, and Future Development

Jill L. Wegrzyn[1], Emily Grau[2], Sean Buehler[1], Margaret Staton[3], Dorrie Main[4] and Stephen P. Ficklin[5], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Connecticut, Storrs, CT, (3)University of Tennessee, Knoxville, Knoxville, TN, (4)Washington State University, Pullman, WA, (5)Dept of Horticulture, Washington State University, Pullman, WA

The TreeGenes database (http://treegenesdb.org) was initiated as one of the original USDA-ARS funded databases under the name, Dendrome in the early 1990s. This web-based repository was continuously maintained via project-based funding with limited new development. At that time, the database supported a custom front-end and relational back-end that was populated with public data as well as direct user submissions. In 2015, TreeGenes transitioned into a Tripal database (tripal.info) and leveraged the module-focused framework to support a wide range of users and species. Currently, TreeGenes imports, curates, and integrates genetic, phenotypic, and environmental data for over 1700 forest tree species. These species are primarily non-model as TreeGenes hosts less than 30 reference genomes and only a handful of high quality drafts. Nearly 2000 registered users, including academic researchers, tree breeders, and government agency affiliates are associated with TreeGenes. This module-driven system enabled collaboration with established Tripal databases, including: Hardwood Genomics Web and Genome Database for Rosaceae. These collaborations implemented distributed development with specific teams taking the lead on modules that can be applied broadly. Recent development at TreeGenes has focused on those that serve the non-model plant community as well metadata collection and data integration. NSF-funded Tripal initiatives focused on bringing analysis to the data sources, has enabled TreeGenes to leverage existing tools, such as Galaxy. Sustainability is achieved through shared development, external funding, and inclusion on multi-institutional project proposals that will leverage the data.

W310: Data Resource Sustainability and Funding

Sustaining Life Sciences Data Resources in Europe

Chuck Cook, EMBL-European Bioinformatics Institute, Hinxton, United Kingdom

Research in the life sciences is becoming ever more data-driven, and researchers across the world are now dependent on the data integration and analysis enabled by open-access resources. Life sciences data resources are now, collectively, a major global infrastructure that life scientists worldwide depend on for their daily work, but support for this infrastructure is not well-coordinated and is fragile, with key data resources relying largely on short-term funding from a small set of funders.

I will give a brief overview of life science data resources in Europe, with emphasis on those at EMBL-EBI, of the importance of these resources for research, and of the ELIXIR infrastructure, which coordinates bioinformatics resources across its member states and encourages establishment of more sustained funding for these resources.

As part of its work to promote sustainable funding ELIXIR has identified a set of Core Data Resources that are of fundamental importance to the wider life-science community and the long-term preservation of biological data. I will describe the process by which Core Data Resources are chosen, which uses a suite of quantitative and qualitative indicators.
On a global scale, research funding bodies are aware that life sciences data resources are not well-coordinated and lack long-term funding. The funders are supporting a worldwide initiative, the Global Biodata Coalition (GBC), which I will describe. The GBC aims to increase efficiency via coordination of effort and identifying global Core Data Resources that should, in principle, receive sustained long-term support.

W311: **Data Resource Sustainability and Funding**
National Institutes of Health Perspective on Data Resource Sustainability Plans
Valentina Di Francesco, Computational Genomics and Data Science Program, NHGRI, NIH, Bethesda, MD

W312: **Degraded DNA and Paleogenomics**
The Goat Domestication Process Inferred by Ancient Genomic Data
Kevin Daly, Trintiy College Dublin, Dublin, Ireland
The wild goat *Capra aegagrus* was brought under human control c. 8,000 BC, leading to the domestic goat *Capra hircus*. Despite this long shared history, our knowledge of the goat domestication process is poor and reliant on modern populations. Using genomic data derived from over 80 ancient *Capra* remains from southwest Asia and Europe, the patterns of genetic diversity which characterize domestic goat through time will be described. Geographic structure of Neolithic and post-Neolithic goat populations will be illustrated, as will the role played by regional gene flow from ancient wild goat. A F_{ST} outlier scan will also be described, which identifies several regions strongly differentiated in Neolithic goat populations, including loci close or overlapping genes involved in pigmentation, detoxification, and other livestock traits. In addition, new data is presented addressing two of the unresolved questions relating to the goat domestication process: the makeup of the earliest managed goats at the beginning of the Neolithic period, and the potential role played by other *Capra* species in the gene pool of domestic goat.

W313: **Degraded DNA and Paleogenomics**
Revealing the Dynamics of Sunflower Domestication with Archaeological DNA
Benjamin K. Blackman, University of California, Berkeley, Berkeley, CA
Native American farmers living ~4000-5000 years ago transformed the common sunflower, *Helianthus annuus*, from a highly branched wild plant with small disks and small seeds into a staple oilseed crop that sports a single large head with large seeds on an unbranched stalk. We have assembled a time series of archaeological samples that spans the majority of this period, and we are using endogenous DNA sequences obtained from these samples to reveal how human cultivation altered genetic diversity through time. My talk will focus on how the genomic libraries obtained from these samples and from ethnographic collections from the historic period are proving fruitful for examining hypotheses about where in North America sunflower was domesticated and for highlighting reductions in sequence diversity at multiple time points in the history of sunflower cultivation. The intriguing patterns of haplotype turnover we observe through time also suggest that shifts in agricultural practices have occurred over this period. Finally, my talk will discuss our efforts to define candidate domestication genes through population genomics and transcriptomics approaches with extant germplasm, which will allow us to leverage our archaeological material further to learn how the sunflower domestication syndrome was assembled by Native American farmers through time.

W314: **Degraded DNA and Paleogenomics**
Ancient and Modern Genomes Reveal Population History of the Black Rat (*Rattus rattus*)
He Yu, Max Planck Institute for the Science of Human History, Jena, Germany
As a widely distributed commensal species, the black rat (*Rattus rattus*) is a witness of the urbanization process of human societies and reservoir of various diseases, i.e. bubonic plague, which is responsible for at least three catastrophic pandemics in human history. However, the population history of this species has remained largely unknown. Here, we present the first *de novo* genome assembly of the black rat, with a N50 scaffold reaching 110.5 Mb. We furthermore provide a genome annotation and comparative analysis with its closest relative the brown rat (*Rattus norvegicus)*. In addition we collected ancient black rat samples from various locations across Europe, including 6th century Byzantine remains from Serbia and combine them with the high-quality genome assembly in order to study the population history of the black rat. By screening and whole genome sequencing of these samples, we provide a better understanding of the genomic composition and migration patterns of the black rat across Europe. Our data might help to shed light upon the interaction of human populations and their commensal rodents through time and the potential origin and dissemination routes of plague.

W315: **Degraded DNA and Paleogenomics**
TBD
Joshua D. Kapp, University of California Santa Cruz, Santa Cruz, CA

W316: **Degraded DNA and Paleogenomics**
Synchronous Diversification of Sulawesi's Iconic Artiodactyls Driven by Recent Geological Events
Laurent Frantz, Queen Mary University of London, London, United Kingdom and Greger Larson, University of Oxford, Oxford, United Kingdom
The high degree of endemism on Sulawesi has previously been suggested to have vicariant origins, dating back 40 Myr ago. Recent studies, however, suggest that much of Sulawesi's fauna assembled over the last 15 Myr. Here, we test the hypothesis that more recent uplift of previously submerged portions of land on Sulawesi promoted diversification, and that much of its faunal assemblage is much younger than the island itself. To do so, we generated a novel geological reconstruction of the island over the last 4My and analysed genetic and/or morphometric data from a total of 1289 samples obtained from museums, zoos and wild populations of Sulawesi's three largest mammalian species: the Babirusa, Anoa, and Sulawesi warty pig. Our results indicate that although these species most likely colonized the area that is now Sulawesi at

different times (14 Myr ago to 2-3 Myr ago), they experienced an almost synchronous expansion from the central part of the island, which was driven by recent geological uplifts. Altogether, our results indicate that the establishment of the highly endemic faunal assemblage on Sulawesi was driven by geological events over the last few million years rather than ancient continental drift over the last 40My. In addition, our study demonstrates that while DNA from obtained from museum samples can be highly degraded, these collections represent an underused source of information that can be leveraged to 1) obtain samples from highly remote areas and/or from elusive species 2) to directly assess the impact that recent major anthropogenic disturbances have had on specific species by comparing genetic data from past and present day populations.

W317: Degraded DNA and Paleogenomics

Rethinking Maize Domestication in the Archaeogenomic Era

Logan Kistler, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, DC

Maize was domesticated from wild teosinte in Mexico beginning around 9,000 BP, and it traversed Central America to spread into South America by ~6,500 BP. However, recent genomes from archaeological maize dated to ~5,300 BP in the Tehuacan Valley reveal partial domestication—a mix of maize-like and teosinte-like alleles at loci involved in domestication. This creates a paradox: maize was still only partially domesticated near its site of domestication long after it became established as a crop species in South America, so it is unclear how the full complement of domestication syndrome genes came to fixation in South American lineages.

We sequenced forty indigenous landraces from traditional cultivation contexts in Peru, Chile, Argentina, the Brazilian Amazon, and the Brazilian Savanna, as well as nine complete archaeological maize genomes from the Andes and eastern Brazil. Using these and published datasets, we suggest that South American maize left the source region in Mexico and the primary domestication gene pool as a partial domesticate, and deeply structured, locally adapted lineages evolved *in situ* after arriving in South America. Thus while domestication began in a large single gene pool in Mexico, the linkage and fixation of domestication traits likely occurred in multiple regions and cultural contexts independently. The southwestern Amazon was likely an ecological and cultural incubator for the latter stages of maize domestication in South America.

W318: Degraded DNA and Paleogenomics

Recent History of Gombe Chimpanzees through Ancient DNA Analysis

Andrew T. Ozga[1], Timothy H. Webster[2], Ian C. Gilby[3,4], Rebecca Nockerts[5], Michael A. Wilson[5,6], Anne Pusey[7] and Anne C. Stone[3,4], (1)Center for Evolution and Medicine, Arizona State University, Tempe, AZ, (2)School of Life Sciences, Arizona State University, Tempe, AZ, (3)School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, (4)Institute of Human Origins, Arizona State University, Tempe, AZ, (5)Department of Anthropology, University of Minnesota, Minneapolis, MN, (6)Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, MN, (7)Evolutionary Anthropology Department, Duke University, Durham, NC

Ancient DNA methodologies can be used to answer questions about extant wild populations such as East African chimpanzees (*Pan troglodytes schweinfurthii*). In this study, we examine genetic diversity through analysis of both the host genome and the microbial ecosystems within (i.e. the microbiome) using ancient DNA lab methods, shotgun amplification, targeted capture, and Next Generation Sequencing (Illumina). Specifically, the goal of our research at Gombe National Park is to understand the extent to which mitochondrial and oral microbiome diversity has changed over time and how the oral ecosystem from wild chimpanzees compares to that seen in human populations. To date, we have recovered 33 full mitochondrial genomes from chimpanzees living at Gombe between the 1960's and the 2010's and found a slight decrease in genetic diversity within the park across decades (0.00257 to 0.00155). Additionally, we report that Gombe chimpanzees showed a significantly higher abundance of the common oral microbial phyla Bacteroidetes and Fusobacteria but a significantly lower abundance of Firmicutes and Proteobacteria when compared to historic human populations. Chimpanzees also showed a higher abundance of the genus *Porphyromonas*, a microbe associated with the Red Complex, which has been connected to periodontal disease in humans. These results allow us to evaluate the extent to which historic DNA samples (<200 years) can be used to answer questions regarding host health and diversity over time across a variety of at risk wild populations that may benefit from such analyses.

W319: Degraded DNA and Paleogenomics

Bulk-Bone Metabarcoding: Rapid Insights into Past Biodiversity and Ancient DNA Preservation

Michael Bunce, TrEnD laboratory, Curtin University, Perth, Western Australia, Australia

Fossil bones provide a unique window into the past but they are often difficult to interpret. Only a small proportion of animals are preserved as fossils – an even smaller fraction are then recovered and able to be identified morphologically.We have developed an ancient DNA method that rapidly profiles fossil assemblages. Our approach is called bulk-bone metabarcoding (BBM). BBM involves the conversion of largely non-diagnostic bone fragments into powder which is then genetically indexed, amplified and sequenced using metabarcoding.

This presentation will showcase some BBM data from a variety of sites across Australia, New Zealand, Hawaii, USA, Madagascar. The data generated using BBM provides some key insights into past biodiversity and faunal turnover. Moreover, the approach is an efficient way to assess DNA preservation both within and between fossil sites. Taken together, bulk-bone metabarcoding provides a powerful and cost-effective way to study past biodiversity with tangible benefits in conservation science, paleobiology and archaeology.

W320: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture

Engineered Genome-Editing Nucleases - Applications in Potato Improvement

Satya Swathi Nadakuduti, Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI

W321: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture

Plant Genome Engineering with Cas9 and Cas12a

Yiping Qi, University of Maryland, College Park, MD

W322: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture

Cloudy with a Chance of Mutations: Gene Editing and Functional Analyses in Soybean
Robert M. Stupar, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

W323: **Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**
Increased Tiller Production in Switchgrass by Targeted Mutagenesis of the Teosinte Branched 1 Genes with CRISPR/Cas9 Technology
Shui-zhang Fei and Yang Liu, Iowa State University, Ames, IA
Tillering is an important trait that affects biomass yield in switchgrass (*Panicum virgatum L.*), an important bioenergy crop. Understanding the tillering mechanism in switchgrass will facilitate breeding for high-yielding switchgrass cultivars. Using CRISPR/Cas9 system, we previously generated T0 switchgrass mutants for the *Teosinte Branched 1* (*tb1*) gene which has been shown to be a negative regulator on tillering in maize and other species, but its role in switchgrass has not been defined. In the present study, we fully characterized the nature of mutations in the primary mutants with the Next Generation Sequencing technology. Solid *Pvtb1a* and *Pvtb1b* mutants with various allelic compositions were isolated from chimeric T0 mutants using micropropagation. In addition, we demonstrated the transmissibility of CRISPR/Cas9-induced mutations in switchgrass and produced transgene-free mutants. By comparing the tiller numbers of heterozygous mutants for *Pvtb1a*, *Pvtb1b*, *Pvtb1a*-*tb1b* and wild type plants, we concluded that *Pvtb1* genes negatively regulate tillering in switchgrass, where *Pvtb1b* has a major effect. Transcriptome analysis showed that 831 genes were differentially expressed in the *Pvtb1*a-*tb1b* knockdown mutant compared to the wild type plant. Gene Ontology (GO) analysis revealed that downregulation of *Pvtb1* genes affects multiple biological processes.

W324: **Development and Application of Genome Engineering and Transgenic Technology to the Agriculture**
GAANTRY, a Novel *Agrobacterium*-based Transgene Stacking System for Improved Crop Biotechnology
Leyla Hathwaik[1], Ray Collier[1,2], James Thomson[1] and **Roger Thilmony**[1], (1)USDA-ARS, Albany, CA, (2)Wisconsin Crop Innovation Center, Middleton, WI
Genetic engineering provides a means 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. GA*A*NTRY (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 GA*A*NTRY's capabilities, we have assembled large T-DNAs carrying 10 or more cargo sequences and have used these constructs to generate transgenic rice and *Arabidopsis* plants. Approximately 90% of the *Arabidopsis* events that were identified using the dual antibiotic selection screen contained all 10 of the introduced sequences and exhibited all 8 of the expected traits. Similarly, GA*A*NTRY-generated transgenic rice plants that were identified using a single selection marker, carried the entire 37 kilobase T-DNA over 50% of the time and none of the events contained unintended sequence from outside the T-DNA. Our research results demonstrate that GA*A*NTRY is a powerful, yet simple to use, new tool for transgene stacking and crop biotechnology.

W325: **Domestication Genomics**
Phaseolus Domestication and Crop Evolution
Roberto Papa, Università Politecnica delle Marche, Ancona, Italy
Phaseolus can be considered as a unique model for the study of crop evolution, and in particular, for an understanding of the convergent phenotypic evolution that occurred under domestication. The almost unique situation that characterizes the *Phaseolus* genus is that five of its ~70 species have been domesticated (i.e., *Phaseolus vulgaris, P. coccineus, P. dumosus, P. acutifolius*, and *P. lunatus*), and in addition, for *P. vulgaris* and *P. lunatus*, the wild forms are distributed in both Mesoamerica and South America, where at least two independent and isolated episodes of domestication occurred. Thus, at least seven independent domestication events occurred, which provides the possibility to unravel the genetic basis of the domestication process not only among species of the same genus, but also between gene pools within the same species. Along with this, other interesting features makes *Phaseolus* crops very useful in the study of evolution, including: (i) their recent divergence, and the high level of collinearity and synteny among their genomes; (ii) their different breeding systems and life history traits, from annual and autogamous, to perennial and allogamous; and (iii) their adaptation to different environments, not only in their centers of origin, but also out of the Americas, following their introduction and wide spread through different countries. In particular for *P. vulgaris* this resulted in the breaking of the spatial isolation of the Mesoamerican and Andean gene pools, which allowed spontaneous hybridization, thus increasing of the possibility of novel genotypes and phenotypes.

W326: **Domestication Genomics**
The Evolutionary Road from Mallard to Pekin Duck
Zhengkui Zhou, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China
Comparative population genomics offers an opportunity to discover the signatures of artificial selection during animal domestication, however, their function cannot be directly revealed. We discovered the selection signatures using genome-wide comparisons among 40 mallards, 36 indigenous-breed ducks, and 30 Pekin ducks. Then, the phenotypes were fine mapped based on resequencing of 1,026 ducks from an F_2 segregating population generated by wild × domestic crosses. Interestingly, the two key economic traits of Pekin duck were associated with two selective sweeps with fixed mutations. A novel intronic insertion most possibly led to a splicing change in *MITF* accounted for white duck down feathers. And a putative long-distance regulatory mutation caused continuous expression of the *IGF2BP1* gene after birth which increased body size by 15% and feed efficiency by 6%. This study provides new insights into genotype-phenotype associations in animal research and constitutes a promising resource on economically important genes in fowl.

W327: **Domestication Genomics**
The Influence of Demographic History on Morphotype Diversification in Cole Crops

Sarah D. Turner, Division of Biological Sciences, University of Missouri, Columbia, MO
The vegetables comprising *Brassica oleracea*, commonly known as cole crops, can be categorized into six distinct and diverse morphotypes: kale, cabbage, Brussels sprouts, kohlrabi, broccoli, and cauliflower. In addition to being valued for their flavor, culinary, and nutritional attributes, these crops are an especially interesting model for domestication; wild mustard, a small, unpalatable plant, was selected to enrich different plant organs, producing an extreme range of phenotypic diversity in a single species. Despite widespread use of these crops to demonstrate the power of human selection on crop domestication, the influence of demographic history on the domestication of *B. oleracea* remains poorly understood. We present a demographic analysis for several major morphotypes in *B. oleracea* to examine how population size has changed over time and to estimate divergence of morphotypes. This work expands our understanding of domestication in cole crops, provides insight into Brassica evolution and, by observing patterns of recent changes in population size, facilitates future crop improvement efforts.

W328: **Domestication Genomics**
Selection Trajectories of Genetic Variants Underlying Domestic Animal Traits
Evan K. Irving-Pease[1], Laurent A. F. Frantz[1,2], Greger Larson[1] and Joshua G. Schraiber[3], (1)University of Oxford, Oxford, United Kingdom, (2)Queen Mary University of London, London, United Kingdom, (3)Temple University, Philadelphia, PA
The study of animal domestication is an important model system for understanding adaptive responses to changes in environmental conditions, demography and selective pressures over time. Despite decades of genetic research into traits associated with domestication, our understanding of the underlying genetic basis of adaptation to the domestic niche remains poor. Using genome-wide modern DNA, previous studies have contrasted populations of wild and domestic animals to scan for segregating signatures of selection in their respective genomes. Due to the intensive nature of modern breeding practices, it is unclear which candidate genes identified by these methods were under selection during the initial process of domestication, and which represent more recent improvement traits. Time series data, obtained from ancient DNA, can resolve these questions by directly observing changes in allele frequencies over time. Here, we reconstruct the allelic trajectory of >100,000 GWAS variants linked to quantitative traits, in four key domestic species (cattle, pigs, horses and goats). Using a novel dataset of >400 ancient nuclear genomes, spanning >12,000 years of evolutionary history, we are able to quantify the temporal origins and strength of selection for genetic variants associated with health, reproductive, performance, production, aesthetic and behavioural traits in domestic animal populations.

W329: **Domestication Genomics**
Domestication of Almond and Peach in the Evolutionary Context of *Prunus* subgenus *Amygdalus*
Dianne Velasco, University of California, Davis, CA
Prunus subgenus *Amygdalus* is comprised of tree and shrub species distributed across temperate latitudes of Asia, including domesticates *P. dulcis* (almond) and *P. persica* (peach). We set out to place their domestication within the broader context of evolution across the subgenus. Though interfertile, *dulcis* and *persica* were independently domesticated approximately 5000 BP in west and east Asia, respectively, and population structure corresponds to their centers of origin and diversity. Genetic diversity in *dulcis* is approximately seven-fold that of *persica*, which follows general expectations given their self-incompatible and self-compatible mating systems, respectively. The domesticates, when examined independently, were found to have little convergence in putative candidate selection loci. We then examined phylogeny, demography, and selection using an additional four Section *Amygdalus* (almond) and five Section *Persica* (peach) species with *P. cerasifera* (subgenus *Prunus*) as an outgroup. Phylogenetic analysis clearly separated species into two monophyletic groups corresponding to the two sections although Section *Persica* species were additionally split into two subgroups. Demographic modeling with SMC++ showed patterns of effective population size (*Ne*) in species across the subgenus had a nearly simultaneous decrease beginning approximately 2×10^6 generations ago in the early Miocene. However, *Ne* of self-incompatible species *dulcis* and *davidana* began stabilizing in the late Pleistocene with all species, except *ferganensis*, increasing in the early Holocene. Only the domesticates experienced a rapid increase of *Ne* to the present, suggesting population expansion post domestication. Selection in gene coding regions evaluated along the phylogenetic tree, with domestication considered as *a priori* knowledge of selection in the domesticates, dulcis and persica. Selection was detected in 4736 or 17.07% ($p < 0.05$) of the 27747 tested genic loci. Of these only 688 have some homology annotation, approximately 8% related to disease resistance.

W330: **Domestication Genomics**
The Genomic Consequences of Grape and Apple Improvement
Sean Myles[1], Zoë Migicovsky[1] and Gan-Yuan Zhong[2], (1)Dalhousie University, Truro, NS, Canada, (2)USDA-ARS, Geneva, NY
Apples and grapes are two of the earliest domesticated fruit crops and both have been shaped by human culture for thousands of years. To investigate the genomic consequences of domestication and breeding on these two valuable fruit crops, we present an analysis of genome-wide SNP data from the USDA's apple and grape germplasm collections. We find evidence of selection for hermaphroditism, white skin and muscat aroma in grapes and a strong signature for selection for red skin in apples. We also demonstrate that apples primarily used for making hard cider likely acquired their tannic and acidic qualities from the European crabapple, Malus sylvestris. In both apples and grapes, we find extensive first-degree relatedness in the germplasm collections, especially among today's most commercially successful cultivars. This suggests that today's consumers enjoy only a small fraction of the available genetic diversity in these species, and that there is tremendous future improvement that can be achieved through breeding.

W331: **Duckweed Research and Applications**
Comparative Analysis of Two Lemna Genomes
Rob Martienssen, HHMI-Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

W332: **Duckweed Research and Applications**
Cytogenomic Elucidation of the Spirodela Genomes
Ingo Schubert, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and Phuong Thi Nhu Hoang, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

The neotenous aquatic duckweeds of the monocot order Alismatales comprise 37 species of the genera Spirodela, Landoltia, Lemna, Wolffiella and Wolffia which vary as to their genome size (160 -2203 Mbp/1C) and chromosome number (2n=20-126). The ancestral genus Spirodela comprises only two species, *S. polyrhiza* (2n=40) and *S. intermedia* (2n=36), both with a genome size of 160 Mbp.
Using comparative multicolor FISH with pooled fingerprinted BACs as probes and integration of Oxford Nanopore assembly we i) resolved the discrepancies between previous maps for two *S. polyrhiza* clones, ii) revealed no chromosome rearrangements between seven studied *S. polyrhiza* clones and iii) established an updated reference genome map for *S. polyrhiza*.
By cross-hybridization with *S. polyrhiza* BACs to *S. intermedia* chromosomes, chromosome homeology and karyotype evolution between *S. polyrhiza* (n=20) and *S. intermedia* (n=18) were investigated and possible scenarios, depending on the direction of evolution, were suggested.
Based on PacBio read assemblies, reiterative rounds of manual curation to reduce the number of scaffolds and validation by FISH, 134.1 Mbp (~84% of the genome) could robustly be assigned to defined S. intermedia chromosomes so far. The number of genes and classes of repetitive DNA are established and compared to those of the sister species *S. polyrhiza.*

W33: **Duckweed Research and Applications**
Duckweed rDNA Organization and Evolution
Nikolai Borisjuk, Huaiyin Normal University, Huaian, China
Recent genome survey revealed that great duckweed, *Spirodela polyrhiza*, has an extremely low copy number of genes coding for ribosomal RNA (Michael et al., 2017; Hoang et al., 2018), compared to other plants. In all eukaryotes, ribosomal RNA genes (rRNA) are represented by tandemly repeated units encoding 45S rRNA precursor for 18S-5.8S-25S rRNAs and 5S rRNA, which account for about 65% of all cellular RNA. Because of rRNAs role in the assembly of ribosome, efficient functionality of rDNA plays pivotal role in development and growth of the organisms.
In this study we investigated molecular organization of rDNA in a number of geographic ecotypes of *S. polyrhiza*, as well as in the genomes of duckweed species representing other four genera of Lemnaceae family: *Landoltia, Lemna, Wolffia* and *Wolffiella*. The characterization of cloned 45S rDNA units, composed of coding region for the 18S-5.8S-25S rRNAs and the intergenic spacer (IGS) containing a Polymerase I promoter and transcription initiation site (TIS), has shown that the IGS length positively correlates to the species genome size. The shortest IGS of ~ 3.2 kb was revealed for *S. polyrhiza*, and the longest one, of ~10 kb, for *Wolffia globosa*. Compared to other plants, the sequence of *S. polyrhiza* IGS revealed a couple of unique features, such as: very low intra- and inter-genomic heterogeneity, high GC content and an unorthodox molecular structure of the TIS.
Characterization of the *S. polyrhiza* 5S rDNA repeats revealed an existence of two major IGS length variants. The European ecotypes contain a dominant type of about 1050 bp, whereas the Asian and American ecotypes have a single ~400 bp long 5S rDNA intergenic spacer. Estimation of rDNA copy number by quantitative PCR showed that all analyzed *S. polyrhiza* ecotypes have less than hundred copies of both 45S and 5S rDNA genes, the lowest number among the investigated flowering plants. Due to this low copy number and the combination of available rDNA manual sequencing data, ultra-long Oxford Nanopore reads and chromosome localization by FISH, we believe the rDNA loci of *S. polyrhiza* is the best characterized among the plant species up to date. The obtained data will be discussed in relation to molecular organization, evolution and transcription of rRNA genes in duckweed.

W34: **Duckweed Research and Applications**
Development of Efficient Protocols for Stable and Transient Gene Transformation for *Wolffia globosa* using Agrobacterium
Hongwei Hou, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China
Members of the *Wolffia* genus are fascinating plants for many biologists as they are the smallest flowering plants on Earth and exhibit a reduced body plan that is of great interest to developmental biologists. There has also been recent interest in the use of these species for bioenergy or biorefining. Molecular and developmental studies have been limited in *Wolffia* species due to the high genome complexity and uncertainties regarding the stable genetic transformation. We have developed new protocols for both stable and transient genetic transformation for *Wolffia globosa* using *Agrobacterium tumefaciens*. For the transient transformation, we used *Wolffia* fronds whereas we used clusters for the stable transformation. As proof of concept we transformed two synthetic promoter constructs driving expression of the GUS marker gene, that have previously been used to monitor auxin and cytokinin output in a variety of species. Using these approaches we obtained a Transformation Efficiency (TE) of 0.14% for the stable transformation and 21.8% for the transient transformation. The efficiency of these two methods of transformation are sufficient to allow future studies to investigate gene function. This is the first report for successful stable transformation of *W. globosa*.

W35: **Duckweed Research and Applications**
Single Cell Circadian Analysis in Duckweed
Tokitaka Oyama, Department of Biological Sciences, Graduate School of Science, Kyoto University, Kyoto, Japan

W36: **Duckweed Research and Applications**
Population Genomics of Spirodela
Shuqing Xu, AG. Plant adaptation-in-action, Institute for Evolution & Biodiversity, Münster, Germany

W37: **Duckweed Research and Applications**
Hyperpolymorphic NB-ARC Genes for High-Resolution Genotyping in *S. polyrhiza*
Eric Lam[1], Philomena Chu[2], Glen Matthew Wilson[3], Todd Michael[4] and Josh Honig[2], (1)Department of Plant Biology, Rutgers, the State University of New Jersey, New Brunswick, NJ, (2)Rutgers, State University of New Jersey, New Brunswick, NJ, (3)Norwegian University of Science and Technology, Gløshaugen, Norway, (4)J. Craig Venter Institute, La Jolla, CA

W38: **Ecological Genomics**

Species Range Limits, Recombination Rate Variation, and Chromosome Inversions Shape the Distribution of Deleterious Variants in Wild Barley

Peter L. Morrell, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

W339: **Ecological Genomics**

The GWAS Times: A Time-Series Aware GWAS to Detect Natural Climate Adaptations in *Arabidopsis* and *Populus*

Jared Streich[1], David Kainer[2], Jonathan Romero[3], Ashley Cliff[3] and Daniel Jacobson[4], (1)Oak Ridge National Labs, Oak Ridge, TN, (2)Oak Ridge National Laboratory, Canberra, TN, (3)University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, TN, (4)Oak Ridge National Laboratory, Oak Ridge, TN

We created a time-series aware association approach called GWATS (Genome Wide Association Time-series Studies) to detect climate adaptive alleles and demonstrated them on *Populus trichocarpa* and *Arabidopsis thaliana*. GWATS involves running a GWAS analysis using time-series phenotype data, in this case for each day of the year. BioClim has been an ideal data series for species distribution modeling and finding adaptive alleles but lacks a true seasonal component and leaves much of each seasonal period unexamined. Instead, from raw monthly climate data, we interpolated 365 daily values of 26 climate/environment layers using 265 locations of 970 *Populus trichocarpa* individuals from where our GWAS population were sourced using ~10M genome-wide SNPs. Additionally, we analyzed the 1001 (1135 indv.) genomes project lines of *A. thaliana* and their source locations using 12.9 M SNPs. Geographically isolated alleles tend to coincide with adaptive alleles making them difficult to distinguish from false positives. However, true positive p-values become distinct from false positives by using a Fourier Transform to analyze co-variate variation and output similarity matrix calculations across the time-series. Further, using a machine learning algorithm called iRF (iterative Random Forest) we can filter complex climate phenotypes into epistatic interactions across multiple suites of alleles ranging as high as five or more orders of epistasis. Using GWAtS we detected hundreds of candidate climate adaptive loci for Solar Radiation Stress, Temperature Stress, Aridity Stress, Light Quality and many more.

W340: **Ecological Genomics**

Population Genomics of American Bison

David Forgacs, Texas A&M University, College Station, TX

The American plains bison – the US National Mammal – provides an interesting case study for conservation genomics, as this iconic species has undergone a series of recent and drastic bottlenecks but some populations have rebounded beyond our wildest hopes and dreams. My research focuses largely on the bison herd at Yellowstone National Park, one of the only ecologically and genetically viable conservation herds in the US. I will demonstrate how population structure has shifted in the last two decades, necessitating an update to current management techniques. I will also show that by comparing the whole mitochondrial genomes of 65 bison, we have uncovered the signal for the two historic bison populations that comprise the modern herd. In order to facilitate easier, more randomizable and non-invasive sample collection methods, we have tested 50 paired blood and fecal bison samples at 35 microsatellite markers to directly compare the results. I will show evidence that while through careful selection methods and verification steps, some markers can be validated for use with fecal DNA, the lack of careful selection may result in a statistically significant reduction of heterozygosity estimates due to allelic dropout. This effect has caused many previously published fecal DNA studies of endangered species to over-report the level of inbreeding depression. Lastly, I will present the new bison SNP-based genotyping assay we are developing to provide a more robust analysis of identification, parentage testing, and cattle introgression than the microsatellite-based test used currently.

W341: **Ecological Genomics**

The Importance of Alternative Splicing in Adaptive Evolution

Nolan Kane, University of Colorado, Boulder, Boulder, CO

Although the importance of alternative splicing has been studied in many species, the evolution of alternative splicing is under-explored. Here we discuss several examples of evolutionary shifts in alternative splicing in sunflowers, both in the wild and during domestication. We also show that alternative splicing changes are associated with traits known to be under strong selection both in the wild and in agricultural systems.

W342: **Ecological Genomics**

Understanding the Role of Plant Genetics in Provisioning and Regulating Ecosystem Services: Two Examples in Sunflower

Stephan Reinert[1], Jason Corwin[1], Qingming Gao[2], Jarrad Prasifka[2], Nolan Kane[1] and Brent Hulke[2], (1)University of Colorado, Boulder, Boulder, CO, (2)USDA-ARS, Fargo, ND, Fargo, ND

Ecological genomics tries to disentangle questions about the evolution and adaptation of organisms in natural environments as well as how the genome affects its current distributions and interactions. These theories also help us to understand interactions of crops with their environment. Our goal is to develop breeding systems that enhance the crop's ability to attract beneficial organisms and deter pests. Specifically, we seek to understand the role of plant genetics on pollinator visitation and soil microbial symbiosis to improve plant health and enhance yield.

Pollinator visits can substantially increase sunflower yield, as much as 108%. Wild bees are the main pollinator of sunflower where it is cultivated in North America. In the pollinator project, we used 198 lines of the SAM association population with 152,606 SNP loci to identify floret and seed size QTL as well as their correlation. We have found 12 QTL associated with floret size traits and 25 QTL linked to seed sizes traits across the sunflower genome. The 27 QTL cover ~180 Mb of the sunflower genome including ~2,400 genes. Among those ~2,400 genes, we identified 16 potential candidate genes based of increased expression in flower or seed tissues. With corolla length inversely related to wild bee visits, and seed length positively correlated with corolla length, there is a perception that a bee-friendly, long seeded confectionery sunflower may be difficult to attain. However, these associations are not always seen at the gene level and provide opportunities for improving both traits simultaneously.

Sclerotinia is one of the most important disease organisms in sunflower and can cause immense yield losses. It is hard to manage due to the requirement of several genes to generate resistance.We used a subset of 100 lines of an OPV/landrace GWAS panel to conduct a field experiment where one treatment was inoculated artificially with Sclerotinia, and a mock-treated control. After analyzing rhizosphere samples of the

uninoculated group, we identified 44 prokaryotic taxa that appear to work as a community that were inversely related to *Sclerotinia* disease incidence in the treated plots and explained about 30% of the total phenotypic variation for basal *Sclerotinia* rot incidence. Since poor and good hosts for these 44 taxa are present in modern sunflower breeding pools, an understanding of the underlying genetics of host suitability provides a functional avenue to host-plant resistance to Sclerotinia and perhaps overall improved plant health.

W343: **Ecological Genomics**

Whole-Genome Pooled Sequencing for Differentiating Mānuka Provenances

David Chagné, The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Mānuka is a shrub that grows across many ecosystems in New Zealand. To date only one species of mānuka (*Leptospermum scoparium*) is recognised in New Zealand. Species of *Leptospermum* are also found in South East Australia (Victoria, New South Wales, Tasmania), including species currently considered as *L. scoparium*. However, the taxonomic characterisation of mānuka is still the subject of debate, and the only current knowledge of genetic diversity is through the study of terpene compounds in the foliage. A full genome assembly of the mānuka genome was recently developed, encompassing 297 Mb in total and scaffolded into the expected 11 pseudo-chromosomes of the Myrtaceae family. These 11 pseudo-chromosomes are syntenic to the related Myrtaceae model species *Eucalyptus grandis*. We undertook whole-genome pooled re-sequencing of mānuka and kānuka (*Kunzea ericoides*) using the Illumina NovaSeq6000 platform, to identify 685,813 genome-wide single nucleotide polymorphisms (SNPs) across 52 populations sourced from around New Zealand and Australia. These SNPs successfully differentiate local provenances within New Zealand, and between New Zealand and Australian related species. Our project is unique in its engagement with the indigenous people of New Zealand, as Māori stakeholders from the New Zealand mānuka honey industry were involved in planning, coordinating and undertaking the research.

W344: **Engineering NUE**

Using the Plant Microbiome to Reduce Reliance on Chemical Fertilizers

Sharon L. Doty[1], Andrew W. Sher[1], Andrea Firrincieli[1], Matthew R. Joseph[1], Pierre M. Joubert[1], Dehong Hu[2], William Chrisler[3] and Galya Orr[2], (1)University of Washington, Seattle, WA, (2)Pacific NW National Lab, Richland, WA, (3)Pacific Northwest National Laboratory, Richland, WA

Biological nitrogen fixation (BNF), the conversion of atmospheric dinitrogen gas into organic forms of N, can only be achieved by the microbial enzyme complex nitrogenase. Symbiosis with N-fixing (diazotrophic) bacteria therefore can provide plants with this essential macronutrient for growth. While the symbioses between Rhizobia and legumes as well as that between Frankia and actinorhizal plants are both well-known, BNF in root nodules is not the only way in which plants can obtain fixed N from bacteria. Diazotrophic endophytes, microorganisms living within the plant host, can also provide significant levels of organic N and have a much broader host range than nodulating bacteria. Native poplar (*Populus*) and willow (*Salix*) plants have a diverse microbiota including endophytes, some of which can fix dinitrogen gas and promote plant growth and health even under abiotic stresses. Our lab demonstrated using $^{15}N_2$ incorporation and acetylene reduction that N_2 is fixed at high levels by wild poplar by endophytes, though at variable levels depending on the endophytes present. Addition of a consortium of these wild poplar endophytes to hybrid poplar resulted in increased growth and nitrogen availability. Not only did the endophytes promote growth of poplar, they also increased growth and health of a wide range of plant species including grasses, rice, tomatoes, peppers, strawberries, and conifers. To successfully apply endophyte technologies for reducing reliance on chemical fertilizers, a greater understanding of endophytic N-fixation is necessary. Experiments utilizing random barcoded TnSeq, nif FISH, LCM, and RNAseq are underway.

W345: **Engineering NUE**

Nitrate Acts as a Signal, Not a Nutrient, to Promote Seed Germination

Eiji Nambara, University of Toronto, Toronto, ON, Canada

W346: **Engineering NUE**

How Hormones and Other Nutritional Signals Could Improve NUE?

Gabriel Krouk, CNRS, Montpellier, France

W347: **Engineering NUE**

Understanding the Linkage between Yield and Grain Protein Content in Wheat under a Hot and Dry Climate

Delphine Fleury, University of Adelaide, Adelaide, Australia

W348: **Engineering NUE**

Roles of the NIGT1-Centered Transcriptional Cascade in the Regulation of Nitrate Uptake and Responses

Mineko Konishi, The University of Tokyo, Tokyo, Japan

W349: **Equine 1**

NRSP-8 Bioinformatics Update

James M. Reecy, Iowa State University, Ames, IA

W350: **Equine 1**

Equine Y Chromosome Research Post Sequencing

Caitlin Castaneda, Texas A&M University, College Station, TX

Recent completion of the horse Y chromosome reference assembly and the emerging knowledge about Y chromosome haplotypes in equine populations, open the doors for novel research directions. First, we continue copy number (CN) and copy number variation (CNV) studies of Y-

linked ampliconic genes between individuals and breeds. Until now, we have generated reproducible data by ddPCR for 8 ampliconic genes and *SRY* across 12 selected breeds and observe high degree of CN conservation. We are currently expanding CN research to the same set of diverse, globally distributed horse breeds that were recently used for Y halplotype analyses and tracing sire lines. The findings are expected to reveal comparative contribution of CNVs and SNPs to Y chromosome variation in horses. Along with domestic breeds, we are studying Y CNVs in other Equus species and have generated comparative data for the domestic and Prezewalski's horses. Next, the observed conservation of gene CNs in normal populations suggests that CN stability of ampliconic genes may be functionally important. Therefore, another research direction is comparing CNs between reproductively normal males and those with various spermatogenic problems. Finally, since the horse Y reference sequence contains several gaps, work is ongoing to fill those by using Y sequence data from the cohort of males that contributed to SNP discovery and haplotype analysis. Improving the accuracy and contiguity of horse Y sequence assembly is the key for its proper functional annotation in course of the ongoing Equine FAANG initiative. Through a combination of results, we will gain a deeper understanding of horse Y function and evolution.

W351: Equine 1

Molecular and Cytogenetic Analysis of Centromere Repositioning in the Genus *Equus*

Elena Giulotto, University of Pavia, Pavia, Italy

The centromere is the site of kinetochore assembly required for chromosome segregation during cell division. The centromeric function is highly conserved and epigenetically specified by the protein CENP-A, while centromeric DNA sequences are variable, even among related species. Although centromeres are typically characterized by satellite DNA, we proved that in the genus *Equus* several centromeres are satellite-free. These Evolutionary New Centromeres (ENCs) arose recently during evolution through centromere repositioning (shifting of centromeric function without chromosome rearrangements).

In the genus *Equus*, speciation events occurred rapidly, in the evolutionary time scale. Using a cytogenetic approach, we previously demonstrated that the rapid karyotypic evolution of equids is marked by exceptionally frequent events of centromere repositioning (Piras et al, PLoS Genetics 2010).

Following the identification of one satellite-less centromere in the horse (Wade et al, Science 2009) and 16 satellite-less centromeres in the donkey (Nergadze et al, Genome Research 2018), we investigated, by ChIP-seq with an anti-CENP-A antibody, other equid species, uncovering an extraordinarily high number of fully functional satellite-less centromeres: 15 in kiang, 16 in onager, 11 in Grevy's zebra, 14 in Burchell's zebra and 9 in Hartmann's zebra. Comparative sequence analysis revealed the presence of centromeres at different maturation stages, ranging from classical satellite-based to "immature" satellite-less centromeres. Comparison of the genomic position of all these neocentromeres in the different species suggests that hotspots for neocentromere formation were re-used during evolution. Furthermore, using a combination of molecular and cytogenetics approaches, we demonstrated that the phylogenetic history of specific chromosomes is determined by two main interconnected mechanisms: neocentromere formation and rearrangement.

W352: Equine 1

Genomic Comparisons of the Persian Kurdish Horse to Persian Arabian and American Thoroughbred Populations

Navid YousefiMashouf[1], Ernest Bailey[1], Jessica L. Petersen[2], Hassan Mehrabani Yeganeh[3] and Ardeshir N. Javaremi[4], (1)University of Kentucky, Lexington, KY, (2)University of Nebraska, Lincoln, NE, (3)University of Tehran, Karaj, Iran (Islamic Republic of), (4)University of Tehran, Tehran, Iran

The Persian Kurdish horse is an ancient indigenous breed originating from western Iran. DNA samples from 58 Kurdish horses, 38 Persian Arabians (a geographically-neighboring population) and 84 American Thoroughbred horses were genotyped across 670,796 SNP markers. The dataset was pruned using call rate, number of alleles, minor allele frequency, Hardy-Weinberg equilibrium, identity-by-descent and linkage disequilibrium, resulting in 9,170 SNPs in 50 Kurdish, 24 Persian Arab and 59 Thoroughbred horses. The data were analyzed using principal component analysis, F statistics, cluster analyses, analysis of molecular variance and genetic distance. The first eigenvalue explained 6.84% of the variance, discriminating Thoroughbred from Persian breeds, and the second explained 2.06% distinguished between Kurdish and Persian Arabian populations. Pairwise F_{st} between the two Persian breeds was calculated as 0.014, several fold less than that between the Thoroughbred and either of Kurdish or Persian Arab populations (0.054 and 0.063 respectively), reflecting a close genetic relationship between the Persian Arabian and Kurdish and distance from the Thoroughbred. Cluster analysis assigned Kurdish and Thoroughbred horses to distinct clusters (0.96 and 0.95 respectively), but the Persian Arab's genome formed a partially (0.53) distinct cluster with several individuals overlapping those in the Kurdish population. The expected heterozygosity was 0.328, 0.324 and 0.341 for Thoroughbred, Persian Arab and Kurdish populations, respectively. These results have implications for developing conservation strategies to achieve sound breeding goals while maintaining genetic diversity.

W353: Equine 1

Genetic Investigation of Distichiasis in Friesian Horses

Erin A. Hisey, University of California, Davis, Davis, CA, Hanneke Hermans, Utrecht University, Utrecht, Netherlands, Mary Lassaline, Department of Clinical Surgical & Radiological Sciences, University of California, Davis, Davis, CA, Willem Back, Ghent University, Merelbeke, Belgium and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Distichiasis, an ocular condition reported in Friesian horses, occurs when aberrant lashes grow from openings of the Meibomian glands along the inner eyelid. These lashes can cause irritation and corneal ulcers, which can lead to corneal scarring and vision loss. Because of its bilateral nature and prevalence in a single breed with known inherited monogenic disorders, this condition is hypothesized to be inherited as a Mendelian trait. To test this hypothesis, a genome wide association study (GWAS) was performed utilizing the Equine Affymetrix 670K array (MNEc670k) on fourteen cases and thirty-eight controls that were phenotyped for distichiasis. A chi-squared test for a basic allelic association identified a locus on ECA13 as associated with the disorder ($p_{corrected}=2.58x10^{-4}$), however genomic inflation was high (l=1.52). To correct for this, a single locus mixed linear model (EMMAX) approach was employed. Under a recessive model, the locus on ECA13 was further supported

($p_{corrected}=1.41x10^{-2}$). A haplotype analysis identified a 360 Kb run of homozygosity on ECA13 in thirteen of the fourteen cases, providing support for a recessive mode of inheritance. This locus contains three annotated genes, none of which are obvious functional candidate genes. Whole genome high-throughput sequencing data from four horses (2 cases and 2 controls) is being analyzed to identify variants in the associated region on ECA13 for further interrogation.

W354: Equine 1

Patterns of Copy Number Variants in the European Horse Genome: Preliminary Results

Marina Solé[1], Michela Ablondi[2], Amrei Binzer[3], Brandon D Velie[4], Nina Hollfelder[3], Nadine Buys[5], Bart Ducro[6], Liesbeth François[5], Steven Janssens[5], Anouk Schurink[6], Åsa Viklund[1], Susanne Eriksson[1], Anders Isaksson[3], Sofia Mikko[7] and Gabriella Lindgren[1,5], (1)Department of Animal Breeding & Genetics, Swedish University of Agricultural Science, Uppsala, Sweden, (2)Department of Veterinary Science, Università di Parma, Parma, Italy, Parma, Italy, (3)Department of Medical Sciences, Array and analysis facility, Uppsala University, Uppsala, Sweden, (4)Faculty of Life and Environmental Science, University of Sydney, Sydney, Australia, (5)Livestock Genetics, Department of Biosystems, KU Leuven, Leuven, Belgium, (6)Animal Breeding and Genomics, Wageningen University & Research, Wageningen, Netherlands, (7)Dept of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

Copy Number Variants (CNVs −gain or loss of genomic material) are a common form of genetic variation underlying phenotypic diversity across a wide range of species. It has been previously hypothesized that high frequency CNV differentiation between breeds may be linked to population-specific selection; and highly inbred populations may also accumulate abundance of CNVs. Using 670K Affymetrix Axiom™ Equine genotyping array data from a large cohort of individuals (N=1755) belonging to nine European horse breeds (cold blooded draughts to several warmblood populations), we aimed to explore the pattern of CNVs and to identify CNV-associated genes involving biological processes in the European horse genome. In total, we identified 13502 genome segments that displayed CNV gains (average length of 410.63kb) and 5622 genome segments that displayed CNV losses (average length of 199.53kb). *E. caballus* (ECA) 12 was the most enriched in CNV gains and losses, but most of the CNVs were detected on ECA1 and ECA20. The Friesians showed high frequency of unique CNV gains in ECA1 and Exmoors displayed high frequency of unique CNV losses in ECA24. We identified 81.83% of genes harboring CNV regions (±500 kb), which showed significantly overrepresented GO terms involved in sensory perception of smell, G-protein coupled receptor signaling pathway, response to stimulus and regulation of biological process (Bonferroni P-value < 0.05). This work provides evidence to support the hypothesis that high frequency breed-specific CNVs may potentially be responsible for diverse phenotypes and biological processes. Further research is needed to investigate how these unique CNVs affect phenotypic changes.

W355: Equine 1

A Frameshift Variant in MFSD12 Explains the Mushroom Coat Color Dilution in Shetland Ponies

Jocelyn L. Tanaka, University of California - Davis, Davis, CA, Tosso Leeb, Institute of Genetics, University of Bern, Bern, Switzerland, Maura Mack, Colorado State University Veterinary School, Fort Collins, CO, Christine Flury, Bern Univeristy of Applied Sciences, Department of Agriculture Forestry, Food Science & Management, Bern, Switzerland, Iris Bachmann, University of Bern, BERN, Switzerland, M. Cecilia Penedo, Veterinary Genetics Laboratory, University of California, Davis, Davis, CA and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Mushroom is a unique coat color phenotype in the Shetland Pony. It is characterized by the dilution of the chestnut coat color to a light sepia tone and is hypothesized to be a recessive trait. A genome wide association study (GWAS), utilizing the Affymetrix 670K array (MNEc670k) and a single locus mixed linear model analysis (EMMAX), identified a locus on ECA7 for further investigation ($P_{corrected}=7.64X10^{-5}$). This locus contained a 3.5 Mb run of homozygosity in the 12 mushroom ponies analyzed. Replication testing of nine markers from this locus in 42 additional Shetlands (28 mushroom and 14 chestnut) provided further support of a recessive mode of inheritance and confirmed the association ($P_{combined}=4.51X10^{-16}$). Analysis of high throughput Illumina sequencing data from one mushroom Shetland pony compared to 80 genomes from horses of various breeds uncovered a frameshift variant in the *major facilitator superfamily domain containing 12* gene (*MFSD12* p.Asp201fs). This variant was perfectly concordant with phenotype in a total of 90 Shetlands ($P=2.38X10^{-21}$) and was absent in over 300 individuals screened from eight breeds not reported to have the mushroom phenotype. MFSD12 is highly expressed in melanocytes, localized to the lysosomal membrane, and variants in humans and mice are associated with altered pigmentation. Given the reported role of this gene in melanogensis, the nature of the identified variant, and the perfect concordance in our sample set, we propose that p.Asp201fs is the causal variant for the reduction in pheomelanin observed in mushroom ponies.

W356: Equine 1

Genetic Investigation of Idiopathic Hypocalcemia in Thoroughbred Foals

Victor Rivas[1], Gary Magdesian[1] and Carrie J. Finno[2], (1)University of California Davis, Davis, CA, (2)UC Davis, Davis, CA

Thoroughbred foals affected with equine idiopathic hypocalcemia are presented with hypocalcemic tetany, a stiff gait, and hyperhidrosis. The condition appears to be inherited and is invariably fatal. A candidate gene approach was performed as a genetic variant in one of the genes homologous to human *CASR*, *GNA11*, or *TRPM6* was hypothesized to be causative of idiopathic hypocalcemia in Thoroughbreds. Clinicopathologic data from affected and non-affected foals were collected to document the exclusion of other causes of hypocalcemia. Samples of two affected foals and their dams from the UC Davis Veterinary Teaching Hospital and from Hagyard's Equine Medical Institute (Lexington, Kentucky) were collected. DNA underwent next-generation sequencing on an Illumina HiSeq2500. Reads were mapped to the Equcab2.0 equine-reference sequence using Burrows-Wheeler Aligner (BWA) for Illumina mapping program and variants (single nucleotide polymorphisms [SNPs] and insertions/deletions) were identified using freebayes and DELLY2 programs. SNPeffect was used to characterize variants by phenotype. No segregating putative genetic variants associated with idiopathic hypocalcemia were identified in *CASR*, *GNA11*, or *TRPM6*. To further investigate other potential variants, a whole-genome association study was performed. A segregating putative functional variant, causing a nonsense mutation, in a gene encoding for a guanine nucleotide exchange factor was found associated with idiopathic hypocalcemia. To

validate such mutation, immunohistochemistry of this protein is currently being performed to assess protein expression in paraffin-embedded tissue of affected Thoroughbred foals.

W357: Equine 1

Differential Gene Expression in Skin from Foals with Macrolide-Induced Anhidrosis

Laura Patterson Rosa, L. Chris Sanchez, Martha F. Mallicote, Robert J. Mackay and Samantha A. Brooks, University of Florida, Gainesville, FL

Hyperthermia is a common and occasionally fatal side-effect of treatment with macrolides, the antimicrobial class most commonly used for the treatment of foals with *Rhodococcus equi* pneumonia. In a previous study, macrolide-induced hyperthermia was shown to be caused by profound impairment of sweat responses (Stieler at al. 2015). To better understand these findings, we performed a genomics study to evaluate differential gene expression in skin biopsies acquired from six foals before, during, and after treatment with the macrolide erythromycin (ERY). Quantitative sweat tests performed at the time of biopsy were used to show that sweating was suppressed during treatment but had recovered by the time of the last biopsy 10 days after treatment. RNA extraction was performed according to gold standard protocols, and sequencing of RNA was done by Illumina RNA-seq at the University of Florida ICBR Sequencing Core. After checking for read quality using FastQC and MultiQC, differential expression analysis was accomplished using the kallisto and sleuth tools (Bray et al. 2016, Pimentel et al. 2017) in our HiPerGator (University of Florida supercomputer) computer resources. A Wald Test model including day as variable and a false discovery rate of 0.1 was applied to the abundance of pseudo-aligned reads.

There were highly significant effects of day on terbutaline sweat responses ($P<0.0005$; 2-factor day x terbutaline concentration repeated measures interactions). Compared with baseline (pretreatment days 1, 2, 3) values, sweat weights in foals given ERY were significantly lower ($P<0.05$) on all treatment days (4, 5, 8), and on post-treatment days 12 and 27, but were not significantly different on day 42. Due to formalin degradation of the skin biopsies, sequenced reads were on average just 40bp/strand. Yet, between 47,253 and 995,557 reads were successfully pseudoaligned per sample. We tested 2197 loci for differential expression and detected 214 potential candidate genes (qval<0.05) up and down-regulated during ERY-induced temporary hypohidrosis. These pathways involve and can be influenced by the broad class of macrolide drugs including erythromycin, clarithromycin and telithromycin. Analysis of additional pathways and genes is currently ongoing. These results will provide insight into gene function and pathways correlated to the thermoregulatory dysfunction seen with the use of macrolides. Results of this research will help identifying components to the pathophysiology of anhidrosis, and might contribute to the future development of safer drugs for neonatal and pediatric care in horses.

W358: Equine 1

Site-Specific Differential Gene Expression in Cerebellum and Cervical Spinal Cord of Standardbred Pacers and Trotters

Annette M McCoy and **Ann M. Kemper**, University of Illinois, Urbana, IL

North American Standardbred horses race at either the trot, a diagonal 2-beat gait, or the pace, an ipsilateral two-beat gait. Breeding practices over the past century have distinctly separated pacing and trotting lines, but despite this, approximately 20% of the offspring of trotting stallions race as pacers, and there are rare individuals who excel at both gaits. All North American Standardbreds carry the mutation in DMRT3 that is considered to be permissive for "gaitedness", which suggests that additional genetic factors must play a role in the ability to pace. These genes may interact with *DMRT3* or act independently. In mice, *DMRT3* was localized to the spinal cord both pre- and post-natally, but to our knowledge this has not been confirmed in horses. The aim of this study was to catalogue gene expression in the cerebellum and cervical spinal cord, assess differential expression between pacers and trotters in each location, and assess co-expression of differentially expressed genes with DMRT3 to identify putative functional candidate genes for pacing in the Standardbred.

Tissue was collected from two trotters and two pacers euthanized for reasons unrelated to the study. Tissue was collected from two sites in the cerebellum (vermis, hemisphere) and four sites in the cranial cervical spinal cord (dorsal, lateral, deep ventral, superficial ventral). RNA was sequenced using an Illumina HiSeq4000, yielding ~27-35 million paired-end reads per sample. Reads underwent quality control prior to mapping to EquCab3.0 with STAR. Normalized gene counts were used for differential expression analysis in both limma and EdgeR. A total of 18,615 genes were expressed across all tissues. A MDS plot revealed tight clustering of the cerebellar samples, which were distinct from the cervical spinal cord samples. There was minimal evidence of differential expression between pacers and trotters in the cerebellum. In contrast, the lateral and superficial ventral cervical spinal cord samples showed marked expression differences between pacers and trotters (787 and 131 significantly differentially expressed genes, respectively). *DMRT3* was expressed at low levels only in the deep ventral samples of all four horses. Preliminary evaluation of genes from previously identified GWAS regions of interest (ROIs) for gait in Standardbreds has identified at least 10 genes that are co-expressed with *DMRT3* in the deep ventral cervical spinal cord, but these are not differentially expressed between pacers and trotters. In contrast, several genes from GWAS ROIs found in the superficial ventral and lateral spinal cord do appear to be differentially expressed; these analyses are ongoing.

W359: Equine 2

Functional Investigation of Putative Variant for Atypical Equine Thrombasthenia in Thoroughbreds

Anna R. Dahlgren, Fern Tablin and Carrie J. Finno, University of California - Davis, Davis, CA

Atypical Equine Thrombasthenia (AET) is a frequent cause of bleeding in Thoroughbreds, affecting one in every 150 horses. Aberrant cell signaling after thrombin stimulation prevents platelets from efficiently binding to fibrinogen, leading to abnormal bleeding. Affected Thoroughbreds commonly experience epistaxis during racing and prolonged bleeding after a vascular injury. Despite the negative effect on horse health and performance, the underlying etiology of AET is unknown, though pedigrees of affected horses indicate that AET is heritable. A whole genome association study using six affected and twelve control Thoroughbreds identified an associated 1.6 kilobase deletion within a long non-coding RNA (lncRNA) upstream of *suppressor/enhancer of lin-12-like (SEL1L*). When *SEL1L* is knocked down in zebrafish, vascular leakage occurs leading to blood pooling. To investigate the putative role of this deletion in AET, the expression of the lncRNA and *SEL1L* was determined. The lncRNA expression was confirmed by isolating RNA from equine testes, the tissue with the highest expression in humans. After reverse transcription and PCR amplification, Sanger sequencing the product confirmed the lncRNA is expressed in horses. *SEL1L* expression was also confirmed in the equine platelets. In a preliminary study, the protein expression of SEL1L in platelets was investigated using Western

Blot in two affected horses and one unaffected horse. Both affected horses had lower levels of SEL1L protein than the control horse, suggesting this protein plays a role in the etiology of AET. Elucidation of the mechanism of AET in Thoroughbreds can lead to improved health and performance of the breed.

W360: Equine 2

Transcriptome Diversity and Differential Expression Supporting Limb Laminitis

Heather M. Holl[1], Caitlin Armstrong[2], Hannah L. Galantino-Homer[3] and **Samantha A. Brooks**[1], (1)University of Florida, Gainesville, FL, (2)University of Pennsylvania, Kennett Square, PA, (3)University of Pennsylvania, Philadelphia, PA

Laminitis is a painful and debilitating disease of hoofstock caused by inflammation and damage to the lamellae, a deeply folded pattern of dermal and epidermal tissues responsible for suspending the digit within the hoof capsule. Next-generation sequencing enables improvements to annotation by incorporating equine- and hoof-specific expressed sequences. Pacific Biosystems (PacBio) long read technology specifically provides the ability to sequence full length transcripts using the Iso-Seq method. As a result, neither alignment to a reference genome nor *de novo* assembly is required to generate a transcriptome. Instead, individual sequencing reads are clustered into distinct isoforms and polished using consensus calling. Our objective is the identification of genes and signaling pathways that are differentially expressed in lamellar tissue from naturally-occurring supporting limb laminitis and control cases from the Laminitis Discovery Database in association with severity and duration of gross and histopathological evidence of digital destabilization. We first generated a novel hoof transcriptome resource using the Iso-Seq method, then utilized Illumina short read sequences from 12 acute cases, 12 developmental stage cases and 12 controls quantify gene expression changes in SLL. Among the 113 differentially expressed loci we identified genes contributing to immunomodulation, cytoskeleton/extracellular matrix and keratinocyte function. Additionally, 17% of DE loci were not protein coding genes, but novel lncRNA transcripts. Knowledge of these disease pathways will inform future studies to devise improved prevention and treatment strategies, alleviating a significant cause of suffering in horses and other hooved species.

W361: Equine 2

Application of Single-Cell RNAseq to Identify Genes that escape X-Inactivation in the Horse

Brian W. Davis[1], Caitlin Castaneda[1], Andrew Hillhouse[2] and Terje Raudsepp[1], (1)Texas A&M University, College Station, TX, (2)Department of Veterinary Pathobiology, Texas A&M University College of Veterinary Medicine, College Station, TX

To balance X chromosome gene dosage between mammalian males and females, one of the X chromosomes in females undergoes epigenetic silencing, known as X Chromosome Inactivation (XCI). The process is random and the female body is a mosaic of cells with maternal or paternal XCI. However, about 15% human and 3% mouse X genes escape XCI. These genes are important dosage sensitive regulators of normal development, but can cause sexual dimorphism in X-linked diseases, and contribute to disorders associated with X aneuploidies. Despite the biological importance, research on XCI escape genes is currently limited to human and mouse, likely because random XCI confounds the discovery of escape genes by conventional gene expression analysis in tissues. Here we initiated systematic search for XCI escape genes in the horse using single cell (SC) transcriptome analysis. We obtained blood mononuclear cells and skin fibroblasts from one normal female horse, used Chromium 10X platform to isolate millions of SCs and prepare sequencing libraries, followed by RNAseq on Illumina HiSeq4000 platform. Bioinformatics analysis of the data is ongoing and includes *de novo* assembly of the transcriptome, as well as cell-by-cell analysis of allele specific expression based on the EquCab3 genome. Putative XCI escape genes will be identified based on the presence of variation consistent across the majority cells. These findings will advance functional annotation of horse sex chromosomes and provide novel information about X chromosome regulation in mammals.

W362: Equine 2

The Lung Transcriptome of Horses with Pasture-Associated Severe Equine Asthma Identifies a TH17-High TH2-Low Phenotype

Christa M. Frodella[1], Kensey A Thomas[1], Jacquelyn Bowser[2], Cathleen Mochal[1], Alison Eddy[1], Andrew Claude[3] and Cyprianna E. Swiderski[1], (1)College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, (2)Johnson & Wales University, Providence, RI, (3)College of Veterinary Medicine Michigan State University, East Lansing, MI

Severe equine asthma (SEA) is characterized by reversible airway obstruction, non-specific airway hyper-responsiveness and chronic neutrophilic airway inflammation. Two forms of SEA have been described. One is elicited by barn dust in association with indoor housing in continental climates. The second form is associated with grazing pastures during conditions of high heat and humidity. Airway inflammation is predominantly neutrophilic in SEA, presenting a conundrum because TH2 cytokine responses identified in both conditions are predicted to precipitate eosinophilic inflammation. Increased IL-17 has been identified in horses with barn dust SEA, leading us to hypothesize that TH17 and TH2 phenotypes co-exist in pasture-associated equine asthma. To test this hypothesis and identify relevant upstream regulators, we contrasted the lung transcriptomes of horses with pasture-associated SEA (N=6) in serial lung biopsies collected during disease exacerbation and remission. Reads were aligned to EquCab3.0 with differential expression analysis and modeling using CLC Genomics Workbench and Ingenuity Pathways Analysis, respectively (Qiagen). IL-17 signaling (p=7.6x10E-11) was the top canonical pathway, supporting predominance of TH17 responses in pasture-associated SEA. TH2 signaling was also significantly increased (p=3.2x10E-4). HMGB-1 signaling, documented to support both TH17 and TH2 differentiation, was significantly increased. HMGB-1 is increased via TLR4 signaling and identified in serum and sputum of severe asthmatics where TH17 high and TH2 low phenotypes are described. We conclude that horses with pasture-associated SEA are similarly characterized by a TH17 high/TH2 low phenotype and that environmentally induced HMGB-1 signaling mediates a shift toward TH17 signaling during seasonally induced exacerbations of pasture-associated SEA.

This project was supported by the Agriculture and Food Research Initiative (AFRI) Animal Health Program competitive grant no. 2015

W363: Equine 2

Importance of Precise Phenotyping in Equine Genome-Wide Association Studies

Erin N. Burns[1], Stephanie J. Valberg[2], Monica Aleman[1] and Carrie J. Finno[3], (1)University of California - Davis, Davis, CA, (2)Michigan State University, East Lansing, MI, (3)UC Davis, Davis, CA
Despite the availability of equine single nucleotide polymorphism (SNP) arrays, there are limited published studies demonstrating genome-wide significance in the mapping of equine traits and diseases. Of the 40 traits and diseases for which a genetic test is currently available, only 13 were identified using a SNP-based genome-wide association study. We hypothesize that the difficulty in obtaining sufficient and accurate phenotype data may partially account for the limited success. Using very stringent phenotyping data, we have successfully mapped Immune Mediated Myositis (IMM) and Juvenile Idiopathic Epilepsy (JIE) in a relatively small population (IMM: 90 horses [36 affected and 54 unaffected], JIE; 38 horses [9 affected and 29 control]). With IMM, it was imperative to use a control group maintained under the same environmental risk factors as affected foals. Phenocopies were also noted in the follow up cohort of the IMM mutation. In mapping JIE, genome-wide significance was only achievable when using the most stringently phenotyped horses that had been diagnosed via electroencephalogram. Additionally, covariates such as sex were included using a linear mixed model. Utilizing these methods, we were recently able to identify a genome-wide significant region of interest for equine neuroaxonal dystrophy (eNAD). A strong candidate gene is currently under investigation for eNAD using whole-genome sequencing and a mouse model.

W364: **Evolution of Genome Size**
Gene Families and Rates of Sequence Evolution in Gymnosperms
Amanda de la Torre, Nothern Arizona University, Flagstaff, AZ

W365: **Evolution of Genome Size**
Genome Evolutionary Patterns from Size and Composition Perspectives
Jinyu Wang[1], Xianran Li[1], Kyung Do Kim[2], Michael Scanlon[3], Scott A. Jackson[2], Nathan M. Springer[4] and Jianming Yu[1], (1)Iowa State University, Ames, IA, (2)University of Georgia, Athens, GA, (3)Cornell University, Ithaca, NY, (4)University of Minnesota, St. Paul, MN
Plant domestication provides a unique model to study genome evolution. Many studies have been conducted to examine genes, genetic diversity, genome structure, and epigenome changes associated with domestication. Interestingly, domesticated accessions have significantly higher [A] and [T] values across genome-wide polymorphic sites than accessions sampled from the corresponding progenitor species, and this pattern was found in multiple comparisons of accessions separated by domestication. However, the relative contributions of different genomic regions to this genome divergence pattern and underlying mechanisms have not been well characterized. Here we show that non-genic part of the genome has a greater contribution than genic SNPs to the [AT]-increase observed between wild and domesticated accessions in maize and soybean. The separation between wild and domesticated accessions in [AT] values is significantly enlarged in pericentromeric regions. Moreover, motif frequency and sequence context analyses showed the motifs (PyCG) related to solar-UV signature are enriched in pericentromeric and nongenic regions, particularly when they are methylated. Additional analysis using population-private SNPs also implicated the role of motifs related to solar-UV signature in relatively recent mutations. With base-composition across polymorphic sites as a genome phenotype, genome scans identified a set of putative candidate genes involved in UV damage repair pathways. Our findings establish the important links among UV radiation, mutation, DNA repair, methylation, and genome evolution.

W366: **Evolution of Genome Size**
T.B.D.
Pankaj Jaiswal, Oregon State University, Corvallis, OR

W367: **Evolution of Genome Size**
Biased Subgenome Fractionation occurs *via* Fertility-Based Selection in *Brassica* Hybrids
Annaliese S Mason, Justus Liebig University, Giessen, Germany
Subgenome fractionation is the process by which redundant and/or duplicated genes and sequences are lost over time after two or more genomes come together in a polyploidization event. When these subgenomes differ in their rate of sequence loss, such that one genome preferentially loses sequences and the other genome preferentially retains sequences during diploidization (loss of redundant gene copies), this process is called "biased" subgenome fractionation. Biased subgenome fractionation has important implications for genome evolution, but the exact mechanisms which might be responsible for biased subgenome fractionation have yet to be experimentally validated. Synthetic *Brassica* hybrids undergo frequent non-homologous chromosome exchanges, where sections of one subgenome are replaced by copies from the other subgenome (duplication-deletion events). These hybrids hence have a "sped up" rate of genomic change, such that it is possible to observe the effects of allopolyploidization in only a few generations rather than over evolutionary time. In our segregating populations, self-pollinated seed fertility was decreased in individuals with C -> A genome translocation events relative to individuals with A -> C genome translocation events for the same chromosome segments (duplications in the A genome and deletions in the C genome were favoured). Subsequently, after generations of selection for fertility, later-generation hybrids show significant evidence of biased subgenome fractionation processes, with more deletions of the *Brassica* C genome relative to the *Brassica* A genome. Our results highlight fertility-based selection as playing an important role in genome evolution in allopolyploids, and suggest interesting avenues for further research into genome evolution using synthetic polyploid hybrids.

W368: **Evolution of Genome Size**
How to Build a Plant Immunity- and Development-Altering "Toolbox": Large Scale Expansion and Neofunctionalization of Housekeeping Genes in the Plant-Parasitic Cyst Nematodes
Sebastian Eves-van den Akker, University of Cambridge, Cambridge, United Kingdom
Plant pathogens and parasites are a major threat to global food security. Plant parasitism has arisen four times independently within the phylum Nematoda, resulting in at least one parasite of every major food crop in the world.

While the basis for the evolution of nematode parasitism is largely unresolved and widely debated, it is likely that a series of evolutionary transitions gave rise to the biologically complex sedentary plant endoparasites. Surprisingly little is known about the genetic changes that occurred with these transitions. In general, parasites lose functions as they further rely on their host. For the sedentary endoparasitic cyst nematodes, this is evidenced by a reduction in genes involved in detoxification of xenobiotic compounds, and the absence of whole classes of antibacterial and antifungal genes. However, concurrent with this process, cyst nematodes have evolved a large repertoire of 'effector proteins' that facilitate their remarkable abilities to suppress plant immunity and induce plant cells to re-differentiate into a novel tissue. The evolution of sedentary endoparasitism must therefore be additionally characterized by acquisition of novel functions.
We analysed the genomes and/or transcriptomes of nine plant-parasitic and non-plant-parasitic nematode species to identify gene family expansions that have co-occurred with the evolution of the most complex forms of sedentary endo-parasitism to understand how these nematodes evolve new 'effectors'. We found that a massive, two step, expansion of endogenous housekeeping Glutathione Synthetase (GS)-like genes has occurred. New GS paralogues acquired multiple dorsal gland promoter elements, altered spatial expression to the secretory dorsal gland, altered temporal expression to primarily parasitic stages, and a signal peptide for secretion. The gene products are delivered into the host plant cell during infection, giving rise to 'GS-like effectors'. Remarkably, by solving the structure of GS-like effectors we show that during this process they have also diversified in biochemical activity, and likely represent the founding members of a novel class of GS-like enzyme.
Our results demonstrate the re-purposing, and re-deployment of an endogenous housekeeping gene to form a family of effectors with modified functions. We anticipate that our discovery will be a blueprint to understand the evolution of other plant-parasitic nematode effectors, and the foundation to uncover a novel enzymatic function.

W369: **Evolution of Marine Mammal Genome**
A Genomic Framework for Understanding Life History Evolution in Sea Urchins
Lingyu Wang, Duke university, Durham, NC

W370: **Evolution of Marine Mammal Genome**
To be Added Please
Ira R Cooke, James Cook University, Townsville, Australia

W371: **Evolution of Marine Mammal Genome**
Population Genomics Evidences Support Independent Species Status of the Yangtze Finless Porpoise
Guang Yang, College of Life Sciences, Nanjing Normal University,, Nanjing, China

W375: **Evolution of Marine Mammal Genome**
Chromosome Level Genome of Indo-Pacific Humpback Dolphin (*Sousa chinensis*) reveal its Chromosome Evolution and Population Feature by Re-Sequencing
Guangyi Fan, BGI, Shenzhen, China

W376: **Exploring Phytobiomes**
vConTACT 2: A Tool to Automate Genome-Based Prokaryotic Viral Taxonomy
Olivier Zablocki, Ho Bin Jang, Benjamin Bolduc and Matthew B Sullivan, The Ohio State University, Columbus, OH
Viruses of bacteria and archaea impact natural, engineered and human ecosystems, but their study is hampered by the lack of a universal or scalable taxonomic framework. Here we introduce vConTACT 2.0, a network-based application to establish prokaryotic virus taxonomy that scales to thousands of uncultivated virus genomes, and integrates distance-based hierarchical clustering and confidence scores for all taxonomic predictions. Performance tests using vConTACT 2.0 demonstrate near-identical (96%) correspondence to current International Committee on Taxonomy of Viruses (ICTV) viral taxonomy where genus-level assignments are available. Beyond "known viruses", vConTACT 2.0 suggested automatic genus assignments for 1,364 previously unclassified reference viruses, with perfectly scoring assignments submitted as new taxonomic proposals to ICTV. Scaling experiments with 15,280 large viral genome fragments from the global oceans demonstrated that the reference network was rapidly scalable and robust to adding large-scale viral metagenomic datasets. Together these efforts provide a systematic reference network and a critically-needed accurate, scalable, automatable taxonomic analysis tool.

W377: **Exploring Phytobiomes**
Rhizosphere Microbial Communities of *Sorghum bicolor* Cultivars with Contrasting Flavonoid Production Patterns
Mara Cloutier[1], Dinakaran Elango[2], Jin Cui[3], Debamalya Chatterjee[3], Mary Ann Bruns[1] and Surinder Chopra[3], (1)Department of Ecosystem Science and Management, Pennsylvania State University, State College, PA, (2)International Potato Center, La Molina, Peru, (3)Department of Plant Science, Pennsylvania State University, University Park, PA
Flavonoid production and exudation by plant roots can impact the rhizosphere environment by stimulating *nod* gene expression, attracting rhizobia, or deterring root pathogen colonization. It therefore follows that plants producing flavonoids may be altering microbial communities in the rhizosphere. In this study, the rhizospheres of five near isogenic lines of *Sorghum bicolor*, two of which produce root flavonoids (F) and three that do not (NF), were collected and analyzed using 16S rRNA amplicon sequencing. Sequences were grouped into exact sequence variants (ESVs) and comparisons were made across cultivars and flavonoid production patterns (FPPs). Evaluation of alpha- and beta-diversity did not indicate differences in rhizosphere community richness, evenness, or structure across the cultivars or FPPs. Rhizosphere communities were dominated by Proteobacteria (average relative abundance of 74.1% across the rhizospheres), which primarily consisted of Gamma- (31%), Beta- (28%), and Alpha-Proteobacteria (13%). At the genus level, *Pseudomonas* was the most abundant (16.4%), followed by *Burkholderia* (7.2%). Of the 27 ESVs assigned to the *Pseudomonas* genus, 4 were significantly lower in the F sorghum rhizospheres when comparing FPPs, while 5 were differentially abundant across cultivars. Similarly, the abundances of ESVs assigned to the *Burkholderia* genus also revealed differences across

Sorghum cultivars. These results lead us to hypothesize that individual variants may be up-regulated or selected for by different sorghum cultivars and FPPs. Targeting ESVs for culture-based hypothesis driven research that were identified as being differentially abundant across the sorghum cultivars and FPPs will help us to better characterize plant-microbe interactions in sorghum systems.

W378: Exploring Phytobiomes
AgriVectors: A Data and Systems Resource for Arthropod Vectors of Plant Diseases
Surya Saha[1], Wayne Hunter[2], Prashant S Hosmani[1], Mirella Flores-Gonzalez[1], Lukas A. Mueller[1] and The AgriVectors Consortium, (1)Boyce Thompson Institute, Ithaca, NY, (2)USDA Agricultural Research Service, Fort Pierce, FL
Arthropod vectors of pathogens cause enormous economic losses and are a fundamental challenge for sustainable increases in food production, yet agricultural pathosystems remain an underserved area of research. To more effectively fight plant diseases, data pertaining to a disease system needs to be consolidated, made searchable and amenable to data mining. The AgriVectors platform is an open access and comprehensive resource for growers, researchers and industry working on plant pathogens and pathosystems spread by arthropod vectors. The portal connects established public repositories with pathosystem-specific data repositories. The AgriVectors system will provide tools to enable technologies such as RNAi, CRISPR, screening bioassays, etc. to leverage current and emerging knowledge across disciplines. It will also include private and unpublished data, using passwords and secure protocols for restricted access. The portal will be based on the Citrusgreening.org (https://citrusgreening.org/) community resource that was developed as a model for systems biology of tritrophic disease complexes. Citrusgreening.org provides omics and biology resources for the Huanglongbing pathosystem. In addition, it includes a biochemical pathway database for each organism in this disease complex, and an expression atlas with proteomics and RNAseq data from psyllids (http://pen.citrusgreening.org) and citrus (http://cen.citrusgreening.org) across multiple infection states. The AgriVectors portal will extend this model beyond gene-centric omics data to the broader Pathosystem-wide information, with integrated pest management, behavioral, plant health, soil health and climate data to incorporate rapid phenotyping information from research trials, building a foundation for more effectively identifying solutions to combat plant diseases.

W379: Exploring Phytobiomes
Insect-Bacteria-Plant Interactions: Microbiomes of Russian Wheat Aphid (*Diuraphis noxia*) contain Bacteria that Increase Virulence to Wheat
Emily Luna and Jessica Metcalf, Colorado State University, Fort Collins, CO

W380: Exploring Phytobiomes
G.E.M.S: Data Sharing and Analysis enabling Data-Driven Agricultural Innovation while Respecting IP
Kevin A. T. Silverstein, Supercomputing Institute, University of Minnesota, Minneapolis, MN
G.E.M.S is not another database. The College of Food Agriculture and Natural Resource Sciences (CFANS) and the Minnesota Supercomputing Institute (MSI) at the University of Minnesota have merged domain expertise in the food and agricultural sciences with HPC and bioinformatics expertise to drive the development of a next- generation agroinformatics data discovery and analysis platform called G.E.M.S. G.E.M.S brings together these individually powerful components to unlock novel insights from Genetic, Environmental, Management, and Socioeconomic data that will inform and enable innovative agro-economic decisions at different temporal and spatial scales. But innovation in agriculture requires reaching well beyond academia/the ivory tower. G.E.M.S's data parsing and data sharing capabilities are uniquely designed to enable new, agile and mutually advantageous public-private collaborations in the pre-commercialization space. G.E.M.S also enables users to turn data into actionable information along the entire innovation value chain, from technology development, to deployment and stewardship.

W381: Exploring Phytobiomes
Viral Ecology in Agricultural Soils
Laura A. Zinke and **Joanne Emerson**, University of California, Davis, Davis, CA
Marine viruses impact ocean carbon and nutrient cycling, microbial ecology, and climate, and recent work suggests that viruses play similarly important roles in terrestrial ecosystems. Our group uses metagenomic approaches to characterize viral communities in soil and their contributions to soil and plant health in natural and agricultural ecosystems. Here we report the recovery of 3,488 viral populations from 16 near-surface (top 15 cm) soil samples from eight agricultural tomato plots at two time points (pre-planting in April 2018 and at harvest in August 2018). Samples were chosen with the intended purpose of measuring differences in viral community composition in response to four biochar and two nitrogen amendments. However, the most significant differences in viral community composition were between pre-planting and harvest samples and between viral size-fraction metagenomes ("viromes") and standard ("bulk soil") metagenomes. Nearly all (186 of 194) viral populations recovered in bulk soil metagenomes were also recovered in viromes, and the vast majority of viral diversity was recovered in viromes alone (3,294 of 3,488 populations), highlighting the utility of purifying viral particles prior to metagenomics. Differences in soil viral community composition between pre-planting and harvest samples suggest the potential for recruitment of a distinct virome to rhizosphere soils.

W382: Farm Animal Genome Editing
Genome Edited Livestock are Here
Bruce Whitelaw, The Roslin Institute, Midlothian, United Kingdom
Since the first pigs were produced using genome editing technology in 2011 we have come a long way. We have seen animal produced addressing welfare issues, disease resistance and production traits. We have seen use of ZFN, TALEN and CRISPR. We can deliver the genome editor reagents by a variety of platforms; although zygote injection and SCNT are the two mainstays, other cell-based platforms are emerging focussing on SSC and ESC. In poultry it is via PGCs, while the technology is just beginning in aqua. All genomic targets tested to date have come from known biology – in the future we will see both targets identified through genetics and through in vitro screening – with the opportunity for multiplex to deliver QTL benefits. The 'genie-is-out-of-the -bottle' with much more to come from this transformative technology.

W383: **Farm Animal Genome Editing**
Surrogate Sires: Germline Ablation and Transplantation in Livestock
Jon Oatley, Washington State University, Pullman, WA
Spermatogonial stem cells (SSCs) possess the remarkable capacity to regenerate spermatogenesis following isolation from testes of a donor male and transplantation into testes of a recipient male. Over two decades ago, methodology to exploit this feature was developed for mice to produce recipient males that transmit donor haplotype to offspring via natural mating. A paradigm in which SSCs can be isolated from testes of a genetically desirable male and transplanted into the testes of battery of recipient males could provide an advanced reproductive technology for enhancing genetic gain in commercial beef cattle and pig production. Realizing this potential requires recipient males that serve as surrogates for donor-derived spermatogenesis. Ideally, testes of a surrogate sire would completely lack endogenous germline but normal somatic support cell function would be intact. To achieve this, we have targeted the *NANOS2* gene. As a key proof-of-concept, we generated *Nanos2* knockout mice via CRISPR/Cas9 embryo manipulation and found that males are germline ablated. Importantly, *Nanos2* knockout mice attained natural fertility following transplant with wild-type SSCs. To translate this paradigm to livestock, we targeted *NANOS2* in porcine and bovine embryos using CRISPR/Cas9 methodology to create inactivating mutations. For pigs, *NANOS2* knockout boars were found to be sterile due to germline ablation. Importantly, we discovered that testes of *NANOS2* knockout boars are capable of harboring regeneration of donor-derived spermatogenesis following transplantation of wild-type SSCs. These findings represent key steps in the development of surrogate sires as a breeding tool in livestock production.

W384: **Farm Animal Genome Editing**
Modelling Surrogate Sire Technology
John Hickey, The Roslin Institute, Edinburgh, United Kingdom

W385: **Farm Animal Genome Editing**
Genome Edited Small Ruminants
Irina Polejaeva, Utah State University, ADVS, Logan, UT

W386: **Farm Animal Genome Editing**
Genome Editing Pigs
Bhanu Telugu, Department of Animal and Avian Sciences, University of Maryland, College Park, MD

W387: **Farm Animal Genome Editing**
Pig Breeding Industry's Persepctive on Genome Editing Technology
Jonathan E Lightner, Genus, DeForest, WI

W388: **Farmed Insects to Feed Future Populations**
Genetic Engineering of Insects for Food: Where We Are, and Where We Need to be
Brenda Oppert, USDA ARS Center for Grain & Animal Health Research, Manhattan, KS

W389: **Farmed Insects to Feed Future Populations**
The Genetics and Microbiome of *Tenebrio molitor*, an Insect Bred for Animal Feed
Virginia Emery and Hans Kelstrlup, Beta Hatch, SeaTac, WA

W390: **Farmed Insects to Feed Future Populations**
Development of Insects for Food, Feed, Pharma and other Valuable Applications
Aaron T. Dossey, All Things Bugs LLC, Oklahoma City, OK

W391: **Farmed Insects to Feed Future Populations**
CRISPR-Cas9 in the Mealworm *Tenebrio molitor* for Improvement as a Crop
Fu-Chyun Chu, All Things Bugs LLC, Oklahoma City, OK

W392: **Farmed Insects to Feed Future Populations**
Lessons from Triboilum: Transferring Genetic Technologies to other Coleoptera
Marce Lorenzen, North Carolina State University, Raleigh, NC

W393: **Feline and Canine Genomics Workshop**
The Dog Aging Project: A Multi-Omic National Study of Healthy Aging in Companion Dogs
Daniel Promislow, University of Washington, Seattle, WA, Kate Creevy, Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, TX, Kelly Jin, Department of Pathology, University of Washington, Seattle, WA, Jessica M. Hoffman, Department of Biology, Birmingham, AL, Josh Akey, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ and Elinor K. Karlsson, U Mass Med School and Broad Institute, Cambridge, MA

The domestic dog is one of the most phenotypically diverse species in the world. This diversity is seen not only in widely recognized features such as size, coat color, or behavior, but also in lifespan and age-related disease risk. Moreover, pet dogs live in the human environment, and have a highly sophisticated health care system. Thus, the pet dog provides a powerful opportunity to understand the genetic and environmental determinants of variation in aging and age-related disease, to uncover the underlying mechanisms by which genes and environment influence aging, and to do so more quickly and cheaply than would be possible in a human cohort. With this goal in mind, we have created the Dog Aging Project (DAP), an NIH-funded long-term study of aging in pet dogs. The DAP will follow a Foundation Cohort of at least 10,000 pet dogs, including companion dogs that represent the full diversity of breed, age, sex, geography, and socioeconomic status. We will collect a large amount of information about each dog, including frailty measures, genotype, electronic health records, activity levels, diet, home environmental information, and more. For a subset of approximately 1200 dogs, we will also collect annual measures of the metabolome, epigenome, and microbiome. In addition, in a small sub-cohort of 600 dogs, we will test whether low-dose rapamycin increases healthspan. We are committed to making the Dog Aging Project an Open Data Project, ensuring that the data we collect will be available to researchers around the world. We will also encourage interested researchers to submit ancillary proposals to various funding agencies to extend the breadth and depth of research carried out with this unique DAP cohort. The Dog Aging Project has the potential to greatly increase understanding of the causes and consequences of aging outside of the lab, and to help veterinarians predict, diagnose and treat age-related diseases.

W394: Feline and Canine Genomics Workshop
The Genomic Landscape of Selection in Domestic Cat Breeds
Hasan Alhaddad, Kuwait University, Safat, Kuwait
The study identifies signatures of selection in over 35 cat breeds using a comprehensive dataset of over 2000 cats and over 58K autosomal SNP markers. The findings include: (1) breed specific candidate selective sweep regions, (2) an overall regions under selection in the domestic cat genome, and (3) regions/chromosomes that represent selection “deserts”. A general comparison of the signatures of selection of cats to that of other domesticated animals illustrates that the selection phenotypes largely controlled by of single genes resulted in overall fewer candidate selective sweep regions.

W395: Feline and Canine Genomics Workshop
Genetic Discovery in Canines using a Direct-to-Consumer Genotype/Phenotype Dataset
Aaron J. Sams, Embark Veterinary, Inc., Ithaca, NY
Direct-to-consumer genomics has the potential to rapidly accelerate genetic discovery in household pets such as dogs and cats. Embark Veterinary has been marketing direct-to-consumer canine genetic tests using a research-grade 200,000+ SNP array for the past two years and has already begun to generate novel insights into the genetic basis of phenotypic traits relevant to dog morphology and health. Here, we discuss the past and ongoing discoveries made possible by the large sample sizes generated by Embark’s platform, as well as Embark’s ethics policies aimed at protecting consumers and their data.

W396: Feline and Canine Genomics Workshop
African Lion Genome Assembly and Comparative Genomic Analysis
Ellie Armstrong, Department of Biology, Stanford University, Stanford, CA
The African lion (*Panthera leo*) has seen population declines of over 40% over the last two decades. This once widespread species is under threat due to habitat loss, human-wildlife conflict, disease, and now the illegal wildlife trade. Here, we present a chromosome-level assembly of the lion, using the combined technologies of 10x Genomics Chromium, Dovetail Hi-C, and Nanopore. We compare this genome with those of other genomes within the genus *Panthera* in order to gain insight into the population history, genomic architecture, and genome evolution of one of the world's most iconic cats. Further, we demonstrate the impressive and low-cost combination of these technologies for high-quality genome assemblies.

W397: Feline and Canine Genomics Workshop
Darwin's Ark: Building a Open Data Platform for Citizen Science Pet Genomics
Elinor K. Karlsson, U Mass Med School and Broad Institute, Cambridge, MA
Our pets – with huge populations, and human owners invested in their wellbeing – are potentially unparalleled natural models for investigating genome function and advancing complex disease research. To harness this power, we need genomic data resources with detailed phenotypic data from thousands of individuals. We’re engaging directly with pet owners as citizen scientists to collect data at this scale. Since 2014, we have enrolled over 21,000 dogs in Darwin’s Ark, an open data, collaborative platform for pet research. Owners sign up through our web-based portal, complete a series of questionnaires, and provide a dog DNA sample using our saliva collection kit. To encourage owner engagement, we return basic genomic analyses, including breed identification, through our website. To date, owners have cumulatively answered more than 2.3 million behavioral survey questions and we have collected saliva samples for over 5000 dogs. We have generated dense genetic data for over 1000 dogs using a sequencing based approach that types more than ten million markers per dog. Our first genomewide association studies demonstrate we have remarkable power to map, and refine, trait associated loci. We welcome new collaborations to expand the scope of phenotypes collected through Darwin’s Ark and add new companion animal species, with the goal of building a massive, shared genomic resource that will accelerate progress from variant discovery to new therapeutic approaches for pets and their people.

W398: Feline and Canine Genomics Workshop
Ancestry Dynamics and Trait Selection in Domestic Cat-Wildcat Hybrids
Gregory S. Barsh, HudsonAlpha Institute for Biotechnology, Huntsville, AL
The Bengal cat breed was established 40 years ago based on an intercross between the Asian leopard cat (ALC, *Prionailurus bengalensis*) and the domestic cat (*Felis silvestris catus*), with a last common ancestor 6 million years ago. The Bengal breed is characterized by pelage traits and ornate color patterns similar to those of ALCs, ocelots, and jaguars, and thus serves a tractable genetic model for understanding the basis and

evolution of pattern diversity in the felid phylogeny. We explore ancestry distribution and selection signatures in the Bengal breed using reduced representation and whole genome sequencing from 500 cats. ALC ancestry in the Bengal breed is low (mean 3.3%), nonrandomly distributed across the genome, and responsible for a breed specific partial melanism known as *Charcoal*. Selective sweeps in the Bengal breed, however, are associated with domestic cat, rather than ALC, haplotypes. One such selection interval overlaps a genome-wide association peak for *Glitter*, a desirable trait affecting the texture and light reflectivity of the pelage. We use a combination of genetic, transcriptomic, and histological approaches to characterize *Charcoal*, *Glitter*, and color pattern traits in the Bengal breed.

W399: **Feline and Canine Genomics Workshop**
An Ancient Genomic Perspective on the Origins of Dogs
Greger Larson, University of Oxford, Oxford, United Kingdom
Despite numerous investigations leveraging both genetic and archaeological evidence, the geographic origins of dogs remain unknown. On the basis of an ancient Irish dog genome and an assessment of the spatiotemporal appearance of dogs in the archaeological record, a recent paper suggested that dogs may have been domesticated independently in Eastern and Western Eurasia from distinct wolf populations. Following those independent origins, a mitochondrial assessment suggested that the Mesolithic dog population in Western Europe may have been replaced by a population from the East. To test this hypothesis, we are generating nuclear genomes of ~30 ancient dogs sampled from sites across Eurasia, and mitochondrial genomes from ~400 dogs spanning the last 15,000 years across Eurasia. The results of this analysis will reveal the phylogenetic affinities of dogs that were present across the Old World prior to the introduction of dogs associated with farming communities. This study will also allow us to infer the timing of the European mitochondrial turnover and to assess whether there was a commensurate turnover at the nuclear level, thus directly addressing whether dogs were domesticated from more than one population.

W400: **Feline and Canine Genomics Workshop**
99 Lives: Leveraging Genome Analysis in Cats for Disease Model Discovery
Reuben M. Buckley, University of Missouri-Columbia, Columbia, MO
The 99 Lives project, a research community based effort, aims to whole genome sequence (WGS) a wide variety of cats with unique disease relevant phenotypes. WGS data from over 200 individual cats was collated and genotyped using the genome analysis toolkit (GATK) best practices. The project has explored the role of structural variation in cat diseases and identified novel candidate variants for two separate diseases. Chediak-Higashi syndrome, a lysosome trafficking disorder, is caused by a large 20 kb tandem duplication within the *LYST* gene, spaning exons 30 to 38. Feline disproportionate dwarfism, is associated with a complex rearrangement within a metabolic gene, providing a new potential cause for dwarfism in humans.

W401: **Feline and Canine Genomics Workshop**
A Whole Genome Discovery of Domestic Cat Variants for Practicing Genomic Medicine
Wesley Warren, University of Missouri, Columbia, MO
Whole genome sequence variation, single nucleotide (SNV) and structural variants (SV), across domestic cat breeds are poorly characterized to date. Since knowledge of SNV and SV frequencies will be crucial in the discovery of causal disease variants as well as their association with cancer occurrence we characterize SNVs and SVs for 74 various cat breeds each sequenced to at least 28x average coverage. After stringent SNV filtration for depth, quality, and missingness, we identified 36M SNVs, with a median per individual of 13M and 38,920 private per cat. Using independent SV detection algorithms and reporting only those convergent within 500bp of the estimated sequence breakpoint we find a total of 21,064 less than one megabase and 10,008 to be unique to individual cats. We have collectively used this preliminary set of SNV and SVs to discover a possible causal variant of cat dwarfism, compared normal to a tumor genome for an oral carcinoma and to start classifying variants with predicted deleterious impact. We also will discuss a new cat exome resource and our observed capture efficacy in detecting variants possibly associated with either common or rare diseases as well as tumors. Overall, our new variant resources in the domestic cat represent a significant step forward in our efforts to discover segregating variation associated with disease and the characterization of genes promoting cancer progression.

W402: **Feline and Canine Genomics Workshop**
PEA15 Defect causes Cerebral Dysgenesis in Domestic Cats
Emily Graff, Auburn University, Auburn, AL
Cerebral cortical size and organization are critical features of neurodevelopment and evolution in animals. In domestic cats, a homozygous mutation in PEA15 (phosphoprotein expressed in astrocytes-15) results in profound malformation of the cerebrum with ~50% decrease in overall brain weight, distinct disorganization of cerebral cortical layers and grey matter astrocytosis. When compared to unaffected cats, fibroblasts from affected cats had increased caspase-8 mediated apoptosis and increased proliferation. This is the first report of PEA15 as a critical factor for cerebral development in the domestic cat and these findings suggest that PEA15 plays an important role in neurodevelopment of gyrencephalic animals.

W403: **Feline and Canine Genomics Workshop**
Dog10K: An International Sequencing Effort to Advance Studies of Canine Domestication Phenotypes, and Biomedical Research
Brian W. Davis, Texas A&M University, College Station, TX
While genomic advances have led to whole genome sequencing of hundreds of canines, until now no concerned effort has been organized to determine the near and long-term goals of the field. Such organization is critical, as it will position the community best take advantage of the canine system for mapping genes associated with morphology, disease susceptibility and behavior. It will further allow investigators to understand the evolution and development of modern domestic dogs. In the next five years Dog10K aims to do whole genome sequence on 10,000 canids including wild canids, village, feral and non-breed dogs as well as breed dogs worldwide. Centralized, community analysis will

result in databases of single nucleotide, copy number, and structural variants, along with mobile element detection, phased haplotypes, and *de novo* assemblies of not only the boxer dog reference genome but other dog breeds and wild canids. This unprecedented dataset will provide reference data sets for researchers worldwide interested in evolution, behavioral genetics, mammalian biology, and veterinary and human biomedicine.

W404: Feline and Canine Genomics Workshop

Using QTL Analysis of Serum Metabolite Levels to Identify Genetic Loci Involved in Dog Metabolism

Jeff Brockman, Hills Pet Nutrition, Topeka, KS

Metabolomics using LC/GC – mass spectrometry for metabolite profiling has emerged as a valuable tool for understanding metabolism in biological systems. We have used genome wide association analysis to identify genetic loci associated with serum metabolite levels. Carnosine, a dipeptide with potent antioxidant activity, is highly concentrated in brain and muscle tissue and is thought to slow the onset and progression of age related disease. We have used a linear model with 225 mixed breed dogs to identify a locus associated with serum carnosine levels that exceeds genome wide significance (3.67e-08). The locus contains the genes encoding carnosine dipeptidase 1 and 2 (CNDP1, and CNDP2). We have reproduced this result (8.12e-11) in a cohort of 857 purebred dogs with a mixture of over 80 breeds. The minor allele frequency in each cohort is 0.08 and 0.09 respectively. The minor allele effect appears to be additive with the homozygous individuals having an approximate 50% reduction in serum carnosine levels suggesting that the minor allele is associated with an increase in carnosine dipeptidase activity. These findings demonstrate the value of combining genetics with metabolite profiling as a tool for understanding metabolism in companion animals.

W405: Flax Genomics

Transgenics for High Throughput Functional Genomics in Flax: Exploitation of an Apical Meristem-Targeted *in planta* Transformation Strategy

Prasanta K Dash, ICAR-National Research Centre on Plant Biotechnology, New Delhi, India and Rohini Sreevathsa, Indian Centre of Agricultural Research, National Research Centre on Plant Biotechnology, New Delhi, India

Genetic engineering through transgenesis offers a directed method of plant breeding that selectively introduces one or a few traits into crop plants, with *A. tumefaciens* T-DNA-directed gene transfer as the method of choice. Flax (*Linum usitatissimum*) depicts poor regeneration capacity of explants as in various other economically important crops. Hence, it is imperative to develop alternate methods that minimize or eliminate the steps of regeneration. The present study depicts amenability of flax to an apical meristem-targeted *in planta* transformation. The strategy essentially involves *in planta* inoculation of embryo axes (plumule) of germinating seeds, allowing them to grow into seedlings *ex vitro* and identification of the putative transgenics in the T_1 generation. We have demonstrated the genotype-independent nature of this methodology in flax along with transgene integration and inheritance. The approach can be effectively used in both functional genomics as well as in translational research in flax improvement programmes.

W406: Flax Genomics

Using Transcriptomics (RNA-seq) to Dissect Molecular Responses of Flax to *Fusarium oxysporum*

Leonardo Galindo, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada and Michael K. Deyholos, University of British Columbia, Kelowna, BC, Canada

Fusarium wilt caused by the fungal pathogen *Fusarium oxysporum* f.sp. *lini* (*Foln*), is a major soil-borne disease of flax that can cause up to 100% crop yield loses. Resistant cultivated varieties have been used to limit the spread of the disease, but a wide range of susceptibility still exists among these cultivars.

Study of the disease has focused on plant symptoms, physiology and targeted gene expression and transgenics, but whole-genome underlaying mechanisms are not fully understood. With the advent of next generation sequencing (NGS) technologies it is now possible to uncover how transcriptomes in plants are responding to pathogen attack, highlighting key candidate genes that can be used as targets for gene editing.

We used an *in*-vitro assay to evaluate a time-course response (2, 4, 8 and 18 days post-inoculation) of a moderately resistant flax cultivar to infection with a fusarium pathogenic isolate. Plants presented wilting, leaf chlorosis, leaf spots and root browning and decay, and the pathogen was detected colonizing the vascular system of the host. RNA-seq transcriptome analysis indicated a full deployment of plant defense molecular mechanisms at 18 days post-inoculation. The defense response included membrane and cytoplasmic receptors that potentially interact with the pathogen-associated molecular patterns (PAMPs) and effectors. The interaction causes a signal transduction through activation of MAP kinases that translates into a large transcriptional reprogramming (76 transcription factors were significantly regulated). Additionally, most of the key enzymes in ethylene and jasmonate synthesis were triggered, demonstrating that these hormones play a key role in regulation and signaling. Genes responsive to hormone modulation like ethylene response factors – important markers of plant defense against fungal pathogens- were detected as part of this transcriptional response. Along with these, pathogenesis-related proteins including chitinases, thaumatins, lipid transfer proteins and protease inhibitors constitute a first line of defense for flax against *Foln.* However, flax also uses secondary metabolites to increase lignin formation (phenylpropanoid, laccases and peroxidases were modulated), and create an arsenal of antioxidants to control the reactive-oxygen species (ROS) burst - peroxidases found are likely involved in a localized hypersensitive response through ROS generation -. The flavonoid and isoprenoid-related genes found on this study, probably aid on ROS control and can also directly impair fungal function and growth when translocated by transporters – ATP-binding cassette transporters were significantly regulated as well -. Finally, we found some genes that may favor pathogen colonization: certain upregulated auxin genes and cell wall remodeling enzymes may favor growth and cell wall weakening over defense. Additionally, the increased transcription of certain aquaporins may facilitate hyphal colonization, while amino acid transporters promote nutrient deposition that could be used by the fungi.

Besides depicting the transcriptome response, we show the importance of utilizing different sets of bioinformatics tools to dissect these whole-genome responses. We also point out how candidate genes coming out of this type of studies can be used as a starting point for gene-editing towards gene functional validation and potential increases in cultivar resistance.

W407: Flax Genomics

Using Genomics Approaches to Examine the Early Flowering Trait in a Mutant Selected from a Population of 'Royal', Treated with 5-Azacytidine

Akshaya Vasudevan[1], Raja Ragupathy[2], Megan House[1], Lester W. Young[1], Helen Booker[1] and **Stephen J. Robinson**[3], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)AAFC SCRDC, Swift Current, SK, Canada, (3)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

Canada is a world leader in flax production, and the expansion of this crop into the northern region of the prairies requires the development of early flowering cultivars to avoid the risks from frost. New sources of variation for flowering time thus hold great interest. Flax genomics resources are now sufficiently developed to examine traits with complex inheritance. Here we describe the characterization of the early flowering mutant 'RE2' originally selected from the cultivar 'Royal' after treatment with 5-azacytidine (5-azaC). A large recombinant inbred population derived from the cross 'Royal x RE2' was used in this analysis to dissect this quantitative trait. Next generation sequencing data was generated from 24 lines and two genomic regions having significant association with the early flowering trait were identified on chromosomes 9 and 12 using bulked segregant analysis. As 'RE2' was derived using the demethylating agent 5-azaC, Whole Genome Bisulfite Sequencing (WGBS) data was generated to identify variation in methylation patterns. Interestingly, a cluster of significant differentially methylated regions (DMR) was also identified on the chromosome 12. The challenges of performing these analyses in the non-reference 'Royal' background and the integration of these data with RNAseq data will be discussed.

W4**08: Flax Genomics**

Genetic Diversity of the Flax Rust Resistance Gene *L*

Sridhar Ravichandran[1], Frank You[1], Gabriel Sullivan[1], Khalid Rashid[2] and Sylvie Cloutier[1], (1)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)Agriculture and Agri-Food Canada, Morden, MB, Canada

Breeding crops for durable disease resistance relies on the existence of a diversity of resistance (*R*) genes/alleles to recognize and mount defense against the rapidly coevolving pathogens. Comprehensive studies of flax and flax rust, caused by the fungal pathogen *Melampsora lini*, led to the postulation of the gene-for-gene resistance theory (Flor 1956), now called effector-triggered immunity (ETI). To date, five flax rust resistance loci named *K*, *L*, *M*, *N* and *P* have been identified to contain race-specific rust resistance genes. The *L* locus harbors a single functional gene with 13 known *L* alleles. A core collection of more than 400 worldwide accessions was sequenced and a bioinformatics pipeline utilizing mapping and *de novo* assembly was developed to accurately obtain their *L* sequences. Diagnostic STARP markers were designed to distinguish and validate *L* alleles. Comparative analysis of SNPs and other structural variations uncovered novel haplotypes representing potential new resistance sources. This in-depth characterization of the *L* alleles aims to assist breeders in maintaining high level of rust resistance; this should be particularly useful when crossing with more exotic germplasm to select against alleles conferring susceptibility to the predominant Canadian rust races or in the event of shifting in rust races.

W4**09: Flax Genomics**

The Flax Breeding Database (FlaxDB)

Gaofeng Jia, Crop Development Centre, Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada, Frank M. You, Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, Sylvie Cloutier, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, Michael K. Deyholos, University of British Columbia, Kelowna, BC, Canada and Helen Booker, University of Saskatchewan, Saskatoon, SK, Canada

FlaxDB is a comprehensive database that includes ~1.7 million SNPs from 407 flax accessions and 43,000 predicted genes and annotation from a reference genome CDC Bethune. Agronomic, fibre, and quality traits were evaluated under multi-environments for the flax core collection and a pedigree database of Canadian cultivars registered from 1910 to 2015 was built. Three main function models, phenotype, pedigree, SNPs, and genomic information were integrated to apply to various queries by genotype, gene, scaffold, sequence, and chromosome, generating results that could be associated with specific trait performance. FlaxDB bridging phenotype and genotype can facilitate MAS and GS applications in flax breeding.

W4**10: Flax Genomics**

Flax Genetic and Eco-Geographic Elasticity: Hint to Breeders and Conservationists

Demissew Sertse Desta[1,2], Frank You[1], Sridhar Ravichandran[1] and Sylvie Cloutier[1,2], (1)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)University of Ottawa, Ottawa, ON, Canada

Flax is a versatile plant that has been grown as a fiber and oil crops during the entire length of agrarian history. The crop has a broad eco-geographic range spanning the warm Indian subcontinent and North East Africa to the cool European and America's temperate zones. With the aim of understanding the global genetic diversity of flax, a genome wide variation assessment of 383 accessions from 37 countries representing the eco-geographic regions of the crop was performed using ~51K SNPs. The accessions clustered into four major groups: Temperate (TEMP), South Asian (SA), Abyssinian (ABYS) and Mediterranean (MEDT). The TEMP and SA groups were subdivided into eight and two populations, respectively, while ABYS and MEDT had single populations. The TEMP group embraced closely related fiber and oil morphotypes widely distributed in Europe and the new world regions, likely as a consequence of gene flow via human migrations. Genetic variations among groups significantly correlated with variations in latitude and day length. Significantly higher GC content than predicted was tallied for two TEMP and the MEDT populations, hinting at the adaptation mechanism of the crop to cold and dry climatic conditions. The SA, MEDT and ABYS groups, representing the postulated flax centers of diversity, harbored high frequency of endemic haplotypes, suggesting regions of focus for breeders and conservationists.

W4**11: Forage, Feedstocks & Turf**

Taking Phenomics to the Field: Non-Destructive Measurement of Herbage Yield, Persistence and Nutritive Value Traits in Forage Grass Breeding and Evaluation

Kevin F. Smith[1], Pieter E. Badenhorst[2], Alem Gebre[1], Chaya Smith[3], Chinthaka Jayasinghe[1], Noel O.I. Cogan[4] and German C. Spangenberg[4], (1)University of Melbourne, Hamilton, Australia, (2)Agriculture Victoria, Hamilton, Australia, (3)Lat Trobe University, Hamilton, Austria, (4)Agriculture Victoria, Bundoora, VIC, Australia
The contribution of pasture to farm profitability is based on the amount and quality of the herbage grown over the lifetime of the sward. Easy to say but hard to measure, especially when thousands of genotypes are measured as spaced plants, families and plots over multiple sites and years. For forage plant breeders to capitalise on the recent advances in genomic sequencing the need exists to develop new methods to measure trait variation that are rapid, accurate and repeatable. In this paper we discuss our progress in the development of an integrated platform of sensor-based assessment of pasture yield, persistence and forage quality in a genomic selection program for perennial ryegrass. The combination of multispectral, vegetative indices with plant height has enabled the rapid and accurate assessment of biomass from both ground-based and aerial measurement platforms. To measure forage quality we are developing hyperspectral calibrations based on canopy based reflection for a range of traits and are undertaking further research to optimise the sample collection and calibration processes. Ultimately this research will provide an integrated suite of analyses to support the introduction of genomic selection in ryegrass breeding, the description of cultivar performance in evaluation programs and the on-farm assessment of pastures to support management decisions.

W412: **Forage, Feedstocks & Turf**

Boosting Genetic Gain in Tall Fescue *via* Speed Breeding and Genomic Selection

Abdulqader Jighly[1,2], Luke Pembleton[3], Noel O.I. Cogan[3], German C. Spangenberg[2,3], Ben Hayes[4] and **Hans D. Daetwyler**[1,2], (1)Agriculture Victoria, Bundoora, Australia, (2)La Trobe University, Bundoora, Australia, (3)Agriculture Victoria, Bundoora, VIC, Australia, (4)University of Queensland, St Lucia, Australia
Breeding schemes that utilize modern breeding methods like genomic selection (GS) and speed breeding (SB) have the potential to accelerate genetic gain for different crops. We investigated through stochastic computer simulation the advantages and disadvantages of adopting both GS and SB into commercial breeding programs for allogamous crops. In addition, we studied the effect of omitting one or two selection stages from the conventional phenotypic scheme on GS accuracy, genetic gain, and inbreeding. As an example, we simulated GS and SB for five traits with different genetic architectures and heritabilities for a tall fescue breeding program. The phenotypic selection scheme required eleven years, while the proposed GS/SB schemes required four to nine years per cycle. Our results showed that running more SB rounds resulted in higher genetic gain per cycle when compared to phenotypic or GS only schemes and this increase was more pronounced per year when cycle time was shortened by omitting cycle stages. While GS accuracy declined with additional SB rounds, the decline was less in round three than in round two, indicating that it may be stabilizing. However, more SB rounds resulted in higher inbreeding rate, which could limit long-term genetic gain. The inbreeding rate was reduced by approximately 30% when generating the initial population for each cycle through random crosses instead of generating half-sib families. Our study demonstrated a large potential for additional genetic from combining GS and SB. Nevertheless, methods to mitigate inbreeding should be considered for optimal utilization of these highly accelerated breeding programs.

W413: **Forage, Feedstocks & Turf**

Implementation of Genomic Selection in Commercial Breeding of Perennial Ryegrass

Luke Pembleton[1], Courtney Inch[2], Zibei Lin[1], Preeti Thakur[1], Michelle Drayton[1], Maiko Shinozuka[1], Yugo Watanabe[1], Yvonne Ogaji[1], German C. Spangenberg[1], Hans D. Daetwyler[3] and **Noel O.I. Cogan**[1], (1)Agriculture Victoria, Bundoora, VIC, Australia, (2)New Zealand Agriseeds, Christchurch, New Zealand, (3)La Trobe University, Bundoora, Australia
Genomic selection, which uses genome-wide sequence polymorphism data and quantitative genetics techniques to predict plant performance, has large potential for the improvement in pasture plants. Global diversity of the important forage species perennial ryegrass is high and so would require a large reference population in order to achieve moderate accuracies of genomic selection. However, the diversity of germplasm within individual breeding programs is likely to be lower. Consequently, historical phenotypic records for seasonal biomass yield and heading date over an 18-year period within a commercial perennial ryegrass breeding program have been analysis and breeding lines have been characterised with a high-density transcriptome-based genotyping-by-sequencing assay. To represent the true implementation of genomic selection a forward prediction approach was assessed whereby the performance of breeding lines in subsequent years were predicted using only those breeding lines in previous years. Moderate to high prediction accuracies were achieved for biomass yield and heading date, respectively, over a period of nearly 2 decades. These results are supported by simulation studies, demonstrating the ability to predict sward based phenotypic performance early in the process of individual plant selection, so shortening the breeding cycle and dramatically increasing the rate of genetic gain. Supported by computation crossing design scripts that mitigate inbreeding depression, genomic selection has now been implemented directly into the breeding program for multiple years and the first genomically selected synthetic varieties are currently being evaluated.

W414: **Forage, Feedstocks & Turf**

Polymorphism of Genes Involved with Regulation of Flowering Time in *Miscanthus* as C_4 Bioenergy Crop

Zhihui Guo[1], Fiorella D. B. Nunez[1], Hironori Nagano[2], Lindsay V. Clark[3], Erik J. Sacks[3] and **Toshihiko Yamada**[2], (1)Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan, (2)Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Japan, (3)Dept of Crop Sciences, University of Illinois, Urbana, IL
Global climate change and energy security issues have promoted interest in the production and increased availability of alternative energy sources. Lignocellulosic biomass is a promising feedstock source for biorefineries producing biofuel, which can mitigate greenhouse gas emissions and reduce dependency on fossil oil. Perennial C_4 bioenergy crops such as Miscanthus provide good targets as non-edible plant species. Flowering time is a key target trait for extending the vegetative phase to increase biomass in bioenergy crops. Molecular genetic studies allow the identification of genes involved in the control of flowering in different species. The use of model species such as *Arabidopsis* and rice has played a major role in understanding the molecular mechanisms involved in flowering time to help in the genetic improvement of crop development. Nevertheless, little is known about the mechanisms promoting flowering in perennial C_4 bioenergy crops. We identified sequences for some flowering genes such as *Heading date1 (Hd1*), *Heading date7* (*Ghd7*), and *Early heading date1*(*Ehd1*) from *Miscanthus* genome. Genetic polymorphism of these genes was analysed among a wide range of genetic accessions of *M. sinensis*. These findings enable us to

understand allelic dispersal of *Miscanthus* in genes that existed between Asian mainland and Japanese archipelago, but also to be useful for implementation of *Miscanthus* breeding program to develop varieties with high adaptability and productivity in different regions.

W415: Forage, Feedstocks & Turf

Rhizosphere Microbial Communities associated with Switchgrass Grown in Marginal Soils

Yuan Wang[1], Erin Nuccio[2], Javier Ceja Navarro[3], Renmao Tian[4], Travis Simmons[1], Abelardo Arellano[3], Jialiang Kuang[4], Colin Bates[4], Katerina Estera-Molina[5], Lauren Hale[4], Na Ding[1], Josh Barbour[1], Kateryna Zhalnina[3], Nameer Baker[5], Eoin Brodie[3], Trent Northen[3], Wolf R. Scheible[1], Michael K. Udvardi[1], Jennifer Pett-Ridge[2], Malay C. Saha[1], Liyou Wu[4], Jizhong Zhou[4], Mary Firestone[5] and Kelly Craven[1], (1)Noble Research Institute, Ardmore, OK, (2)Lawrence Livermore National Laboratory, Livermore, CA, (3)Lawrence Berkeley National Laboratory, Berkeley, CA, (4)University of Oklahoma, Norman, OK, (5)University of California, Berkeley, Berkeley, CA

Switchgrass (SG; *Panicum virgatum* L.) is a perennial C4 grass native to the tallgrass prairie and a promising feedstock for U.S. bioenergy production. A major barrier to large-scale cultivation of SG under low-input management is seedling establishment. We hypothesize that successful establishment and sustainable cultivation of SG in marginal soils is in part enabled by beneficial plant-microbial interactions in the rhizosphere, potentially through enhanced nutrient acquisition, water uptake and/or pathogen suppression. In this study, we tracked the succession of rhizosphere microbial communities associated with SG plants grown in marginal soils at two Oklahoma field sites. The outcome of this research will provide a better genomic basis for SG cultivation in marginal soils, expand our knowledge of the interactions between soil microbiomes, plants and ecosystems, and ultimately guide efforts for translation into agronomic row crops.

W416: Forest Tree

Complex Regulation of Condensed Tannin Biosynthesis in Poplar by R2R3 MYB Activators and Repressors

C. Peter Constabel and Dawei Ma, University of Victoria, Victoria, BC, Canada

Condensed tannins are the most widespread plant secondary metabolites and highly abundant in woody plants. In poplar, condensed tannins are highly responsive to environmental stresses and are induced by wounding, pathogen infection, UV-B, high light stress, and nitrogen deficiency. A suite of poplar R2R3 MYB transcription factors that control CT accumulation has been identified, and includes both positive and negative regulators acting on CT pathway enzymes. For example, MYB134 and MYB115 activators, when overexpressed in transgenic poplars, lead to a fifty-fold over-accumulation CTs. The activator MYBs also induce expression of R2R3 MYB repressors; these, in turn, downregulate the CT pathway. How these MYBs cooperate during stress responses is not yet known, however. Promoter activation assays in transiently transformed poplar cells demonstrate that flavonoid genes are the targets of activator MYBs, but that this activation can be inhibited by the repressors. Furthermore, the MYB115 and MYB134 activators also regulate each other, a bHLH co-factor gene, and the MYB repressor genes. Transcriptomic analysis of both MYB activator- and MYB repressor-overexpressing transgenics have provided new potential target genes and pathways. Our work suggests that the induction of CTs in poplar is controlled by a complex network of positive and negative regulators, which act to fine-tune CT biosynthesis. How these genes are controlled by environmental stress, and how they interact with signaling pathways, such as the jasmonate pathway, is currently being investigated.

W417: Forest Tree

Expression Quantitative Trait Locus Mapping in Populus

Jay Chen, Oak Ridge National Laboratory, Oak Ridge, TN

Populus hybrids are widely used for biomass production for a suite of industrial applications including biofuels conversion. Our understanding of biomass productivity and quality is limited by the fact that this complex trait requires the regulation and coordinated interactions of many genes. Identification of genetic networks regulating biomass productivity and quality remains largely unaccomplished and is urgently needed to inform genetic improvement of *Populus* feedstocks for biomass production and conversion. We report here the Expression Quantitative Trait Locus (eQTL) Mapping to identify cis-genetic and trans-acting genetic elements as well as genetic networks underlying genome-wide transcript variations by leveraging the large-scale RNAseq dataset generated by a DOE Joint Genome Institute Community Science Program project. We performed the analysis of RNAseq data on a total of 438 biological samples of developing xylems collected from the progeny of a *P. trichocarpa* × *P. deltoides* pseudo-backcross pedigree. Among these samples, 124 samples have biological replicates. In total, the samples represent 312 unique genotypes of the QTL mapping pedigree and two parents. We found that the transcripts of over 15,000 genes were detected (FPKM value ≥1) in the developing xylem of both parents and progeny. We identified a total of 9,161 trans-eQTLs, 2,561 cis-eQTLs, and 15 eQTL hotspots (>150 regulation targets). By performing targeted eQTL mapping of lignin biosynthesis pathway genes, a number of unconventional transcriptional regulators have been identified, which offers opportunities to discover new regulators to fine tune the lignin biosynthesis pathway for cost-effective conversion of *Populus* biomass to biofuels and bioproducts.

W418: Forest Tree

Computational Model for Predicting Monolignol Transcript and Protein Abundances in *Populus trichocarpa* under Single and Combinatorial Monolignol Gene Knockdowns

Megan Matthews, Jack P. Wang, David Muddiman, Ronald Sederoff, Vincent L. Chiang and Cranos Williams, North Carolina State University, Raleigh, NC

Accurate manipulation of metabolites in the monolignol biosynthesis pathway is a key step for controlling lignin content, structure, and other wood properties that are important to the bioenergy and biomaterial industries. A crucial component of this is predicting how single and combinatorial knockdowns of monolignol specific genes at the transcript level influence monolignol proteins, which are the driving mechanisms of the monolignol biosynthesis pathway. Computational models have been developed to estimate protein abundances from transcript perturbations of monolignol specific genes. The accuracy of these models, however, are hindered by their inability to capture indirect influences that arise when one or more genes are perturbed. We created a computational model based on a sparse maximum likelihood approach to estimate the resulting monolignol transcript and protein abundances in *Populus trichocarpa* based on desired single or combinatorial knockdowns of

specific monolignol genes. Using *in-silico* simulations of this model and root mean square error, we show that our model more accurately estimates transcript and protein abundances from xylem tissue when individual and families of monolignol genes were perturbed. Our model captures relationships such as those between the *Ptr4CL* and *PtrCAld5H* gene families, potentially explaining the observed decrease in S/G ratio reported in the literature when *Ptr4CL* is knocked down. This approach provides a useful computational tool for guiding and further exploring the cascaded impact of single and combinatorial modifications of lignin specific genes on lignin and other wood properties. Additionally, it can be used to guide future experiments for elucidating the mechanisms responsible for the estimated indirect influences.

W419: **Forest Tree**

Synthetic-Genetic-Array Based Yeast One-Hybrid System for the Construction of Gene Regulatory Network in Wood Formation

Ying-Chung Lin, National Taiwan University, Taipei, Taiwan

W420: **Forest Tree**

Identification of WRKYs in *Eucalyptus globulus* and their Role in Cold Tolerance

Sofia Valenzuela, Universidad de Concepcion, Concepción, Chile
Catalina Lagos[1]; Paula Aguayo[1]; Marta Fernández[1], Claudio Balocchi[2], Sofia Valenzuela[1*]
[1]Facultad de Ciencias Forestales, Universidad de Concepción, Concepción, Chile
[2]Bioforest SA, Concepción, Chile

The WRKY transcription factor family plays important roles in plants. These genes participate in the transcriptional regulation associated with plant stress responses. However, this important gene family has not been characterized in *Eucalyptus globulus*. This specie is sensitive to freezing temperatures; however it is capable of developing a certain degree of frost tolerance when exposed for a few days to low, but non-freezing temperature, phenomenon known as cold acclimation (CA). A total of 51 *WRKY* genes were identified in *E. globulus* transcriptome in response to CA profile. The amount and group distribution of *EglWRKY* genes were similar to those previously reported in other woody plants, such as *E. grandis* (79), *Vitis vinifera* (59), *Camellia sinesis* (50) and *Prunus persicae* (61). Based on the sequence similarity and distribution of conserved domains, *EglWRKY* genes were distributed into three groups namely, I, II and III. The expression patterns of the 51 *EglWRKY* genes were determined by DEG analysis *in silico* in leaf and root. Transcriptome profiling in response to CA profile indicates that the expression of multiple *EglWRKY*s genes is activated under cold (4 ºC) and chilling temperatures (-6 ºC). A total of 15 cold-induced *EglWRKY* genes with significant changes in their expression levels were identified in leaf. These results suggest that *EglWRKY* genes are involved in cold stress response. This study provides a basis for further studies and analysis of *WRKY* genes to identify their function and molecular mechanisms involves in plant abiotic stress responses in *E. globulus*.

Acknowledgments: This work was financed by Fondecyt 1161063 to S.V and plant material provided by Bioforest SA

W421: **Forest Tree**

Unraveling the Role of Cytosine-5 Methylation on Plant Development in *Salix purpurea* L.

Jorge Paiva[1], Carolina Gomes[2], Dariusz Kruszka[2] and Piotr Kachlicki[2], (1)Institute of Plant Genetics, Poznań, Poland, (2)Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

The study of epigenetics in woody species is important to understand its phenotypic plasticity to a changing environment. Cytosine-5 methylation (Cy5Met) is a major and dynamic epigenetic DNA modification, affecting gene expression. However, Cy5Met machinery and its role on the modulation of Cy5Met levels during plant development is still largely unknown in *Salix purpurea*.

A genome-wide scan of *S. purpurea* genome allowed the identification of seven DNA methyltransferases (*SpMET1a, SpMET1b, SpMET1c, SpCMT2, SpCMT3, SpDRM2 and SpDNMT2*), divided in four clades, and three DNA demethylases (*SpROS1, SpDML1 and SpDML2*), divided in two clades. These genes were characterized in terms of phylogeny, domain conservation, motif structure and chromosomic location. Furthermore, plants with artificially altered Cy5Met patterns were generated using the DNA demethylating agent zebularine, to understand the effect of Cy5Met disruption on plant development. Plantlets of an *in vitro* micropropagated genotype of *S. purpurea* were grown in 0 µM, 25 µM and 50 µM of zebularine for 20 days. For concentrations ³25 µM, growth inhibition both at root and aerial components was observed. After 20 days, plantlets were placed in zebularine-free medium to monitor the recovery from the zebularine treatment. In zebularine-exposed plantlets, root growth was reestablished but shoot elongation remained impaired. Plant material was collected at two time points and characterized in terms of global DNA methylation patterns, secondary metabolites and Cy5Met-related gene expression. Correlation analysis between these variables provided new clues on the role of Cy5Met during developmental processes.

Acknowledgments: The authors thank the NCN (Poland) SONATABIS 5 bis grant UMO-2015/18/E/NZ2/00694 (project PurpleWalls). J. A.P. Paiva also acknowledges his research contract financed by the EU FP7 BIOTALENT project [GA621321] and financial sources for education in the years 2015-2019 allocated to an international co-financed project no W26/7.PR/2015 [GA 3413/7.PR/2015/2]"

W422: **Forest Tree**

Molecular Response to Stem Inclination in Radiata Pine Seedlings: A Functional Characterization of the Differentially Expressed PrMADS10

Raul Herrera, Maria A. Moya-León, Tamara Mendez and Yazmina Stappung, Institute of Biological Sciences - University of Talca, Talca, Chile

Plants have the ability to reorient their vertical growth when they are exposed to inclination. The molecular response to stem inclination is an intricate mechanism, and to explore this response, young radiata pine seedlings were tilted, and stem samples were collected after 2.5, 10 and 24 hours of inclination, and dissected into upper and lower stem sides. Analysis of transcripts allowed the identification of several functional groups which showed a particular signature both spatial and temporal. Among others, several transcription factors (TF), proteins involved in cell wall remodeling, and membrane transporters were identified. Particularly intriguing is the role of *PrMADS10* TF, which modulates the expression of key genes, activating particular metabolic pathways. To further explain its role, *Arabidopsis thaliana* plants over-expressing *PrMADS10* were

prepared. Affymetrix AraGene chip enabled us to identify up and down regulated genes in over-expressed *PrMADS10* plants. Genes from the phenylpropanoids pathway leading to lignin synthesis were modulated in their expression in over-expressed plants, as well as, other genes related to cell wall remodelling. As expected a higher content of lignin was observed in the transgenic lines. Also, several MYBs, NACs, and other TFs showed an interaction with genes of the phenylpropanoid pathway. These studies increased our knowledge of the mechanism underlying the molecular response to stem inclination, particularly the role of MADS-box from pine.
Fondecyt Grant # 1150964.

W423: **Forest Tree**
Genome-Wide Association Mapping for Wood Formation and Tracheid Traits in Norway Spruce to Identify Novel Candidate Loci
John Baison[1], Linghua Zhou[1], Zhiqiang Chen[1], Carolina Bernhardsson[2], Douglas Scofield[3], Amaryllis Vidalis[4], Tommy Morling[5], Thomas Grahn[6], Lars Olsson[6], Bo Karlsson[7], Pär Ingvarsson[2], Ewa J. Mellerowicz[5], Sven-Olof Lundqvist[6], Harry Wu[1] and María Rosario García-Gil[1], (1)Umea Plant Science Center, Umea, Sweden, (2)Swedish Univeristy of Agricultural Sciences, Uppsala, Sweden, (3)Uppsala Multidisciplinary Center for Advanced Computational Science, Uppsala, Sweden, (4)Center of Life and Food Sciences Weihenstephan, München, Germany, (5)Swedish Univeristy of Agricultural Sciences, Umea, Sweden, (6)RISE Bioeconomy, Stockholm, Sweden, (7)Skogforsk, Svalov, Sweden
Through the use of genome-wide association (GWAS) mapping it is possible to establish the genetic basis of phenotypic trait variation. Our genome-wide association mapping study presents the first such an effort in Norway spruce (*Picea abies* (L). Karst.) for the traits related to wood formation and tracheid characteristics. The study employed an exome capture genotyping approach that generated 178 101 high quality Single Nucleotide Polymorphisms (SNPs) from 40 018 probes within a population of 517 Norway spruce maternal trees. We applied a LASSO based association mapping method using a functional multi-locus mapping approach that utilizes latent traits, with a stability selection probability method (Frequency) as the hypothesis testing approach to determine significant Quantitative Trait Loci (QTLs). Latent traits were derived from estimated breeding values (EBVs) by the application of a quadratic spline model and the central peak regression for MFA. The findings presented from the GWAS analysis have provided loci that are potentially controlling wood formation, tracheid dimensions, their cell wall thickness and microfibril angle. The presence of many traits with several significant QTLs supports the notion that the majority of these traits are polygenic in nature.

W424: **Forest Tree**
High Throughput RGB and Hyperspectral Imaging in Support of Machine Vision and GWAS to Identify Genes Controlling Regeneration and Transformation in *Populus trichocarpa*
Michael Nagle[1], Anna C. Magnuson[1], Chris Draves[2], Howland Jones[2], Gabor Kemeny[2], Cathleen Ma[1], Ritesh Mewalal[1], Jialin Yuan[1], Nihar A. Doshi[1], Fuxin Li[1], Yuan Jiang[1], Brett Pierce[1] and Steven H. Strauss[1], (1)Oregon State University, Corvallis, OR, (2)Middleton Spectral Vision, Middleton, OR
Phenomics methods lag behind advancements in genomics and computing in genetic studies of trees and other plants. In vitro phenotypes are particularly difficult due to the need to maintain sterility during growth, image through partially transparent vessels such as Petri dishes, and the wide range of developmental phenotypes observed as callus, shoots, roots, and embryos develop. Transformation rates are also particularly difficult to quantify due to factors such as chimaerism and proximity of transgenic shoots to those that escape selection. In support of a GWAS (genome-wide association study) to identify genes that control rates of regeneration and transformation in Populus trichocarpa, we have been developing a novel phenotyping system based on visible (RGB, red-green-blue wavelengths) and hyperspectral imaging systems. Working with Middleton Spectral Vision (MSV, Wisconsin), we designed an imaging system that can record visible and hyperspectral images of plants in Petri dishes at high speed. This data can then be directly analyzed for growth, health, and transformation (via green fluorescent protein (GFP) fluorescence) using MSV multivariate statistical methods, and by using machine vision analysis to characterize spatially explicit developmental patterns of growth, differentiation, and transformation. We will describe our work to date with the MSV image acquisition and analysis system, then describe our progress and reclassification accuracy in annotation and analysis of images using machine vision (neural network) methods, including preliminary GWAS results based on machine vision estimates. We thank the National Science Foundation Plant Genome Research Program for support (IOS # 1546900.)

W425: **Forest Tree**
The Redwood Genome Project: Conservation and Restoration of California Icons
Alison Dawn Scott, Department of Plant Sciences, University of California, Davis, Davis, CA

W426: **Forest Tree**
Refining Annotation Methodology for Comparative Genomics in Conifers
Sumaira Zaman, Department of Computer Science and Engineering, University of Connecticut, Storrs, CT

W427: **Forest Tree**
Predicting Spanish Cedar (*Cedrela odorata*; Meliaceae) Origin with 144 SNPs
Kristen Finch, Oregon State University, Corvallis, OR

W428: **Forest Tree**
Combining Exome Capture and Pool-Seq: An Efficient Method to Genotype Species with Large and Complex Genomes
Brandon Lind[1], Mengmeng Lu[2], Dragana Obreht Vidakovic[1], Sam Yeaman[2] and Sally N. Aitken[1], (1)University of British Columbia, Vancouver, BC, Canada, (2)University of Calgary, Calgary, AB, Canada

Conifers exhibit some of the largest genome sizes across tree species, thus developing efficient, reproducible, and cost-effective genotyping methods is a top priority for both conservation and breeding efforts as well as for empirical tests of evolutionary theory. We explored the utility of combining exome capture and pool-seq in Douglas fir (*Pseudotsuga menziesii [Mirb.] Franco*) by quantifying capture efficiency, discovered SNPs, and allele frequency estimates from both individual- and pool-sequencing (*N*= 20 individuals) of exome targets across 37,091 genes via probes designed from transcriptome analysis. Individual and pooled samples showed similar fractions of reads on target (35-38%), near target (<500bp from a probe region, 52-59%) and off-target regions (3-13%) while pooled samples had higher sequencing depth than individual samples for these categories. Individual samples and pooled samples also showed similar fraction of reads on different genic regions — between 16-18% on coding regions, 22-22% on untranslated regions, 5-6% on intron regions, and 44-49% on intergenic regions. After implementing pre-processing best practices, calling genotypes, and hard filtering, we acquired 8 and 4 million SNPs from individual and pooled samples, respectively, where about 2.5 million SNPs overlapped. We will also present data relating to differences in allele frequency estimates between individual and pooled sequencing approaches. In short, combing exome capture and pool-seq methods is an efficient method to genotype species with large and complex genome such as conifers.

W4**29: Forest Tree**
Multiple Approaches to Mapping Resistance to Melampsora Leaf Rust in Shrub Willow (*Salix* spp.)
Dustin G. Wilkerson, Cornell University, Geneva, NY

W4**30: Forest Tree**
Predicting Disease Resistance in Radiata Pine using Genomics
Natalie Graham[1], Mari Suontama[2], Jaroslav Klápště[1], Nari Williams[1], Emily J. Telfer[1] and Heidi Dungey[1], (1)Scion, Rotorua, New Zealand, (2)Skogforsk, Sävar, Sweden
Long-term health and survival are critical for long rotation forestry species to attain their genetic potential in terms of growth and wood quality. Resistance to diseases, such as Dothistroma needle blight and Cyclaneusma needle cast, have been extensively assessed and incorporated as selection criteria for several decades in New Zealand's radiata pine breeding programs. For newer diseases, like red needle cast, development of laboratory-based artificial screening methods have offered an alternative for assessing resistance and estimating breeding values (EBVs). Using this approach, we have generated EBVs for RNC resistance and identified phenotypic extremes in a clonal population of radiata pine, for which no resistance data was previously available. However, screening remains labour-intensive and expensive and relies on the availability of clonal material to ensure statistical robustness.
Genomic selection (GS) is proving particularly useful for traits that are difficult to measure, require individuals to reach a certain age, or require exposure to a certain set of conditions or pathogens. We have already shown the ability of GS to predict breeding values with high accuracies for certain form and wood quality traits in other radiata pine populations. Recently, we have genotyped the RNC-screened clonal population using our 44K exome capture panel. We report on the development of prediction models for RNC resistance, and generation of the first RNC resistance genomic estimated breeding values.

W4**31: Fruit/Nuts**
Improvement of Genome Resources for Hazelnut
Kelly Vining, Shawn A. Mehlenbacher and Jacob Snelling, Oregon State University, Corvallis, OR
We present a reference genome for hazelnut (*Corylus avellana*), cultivar 'Jefferson.' A total of 29 Gb (~50x genome coverage) was obtained from 42 SMRT cells on a Pacific Biosciences RSII instrument. Contig assembly was performed with Falcon v. 0.3, and error-correction was done using Pilon. Chromosome conformation capture (Hi-C) enabled chromosome scaffolding. The genome includes 557.84 Mb in 11 chromosomes. The Maker v. 2.31 pipeline was used to annotate 34,982 genes. BLAST2GO v. 4.1. was used to assign gene ontology categories. This resource is now available to the research community.

W4**32: Fruit/Nuts**
Advances in Prunus Genetics: Horticulturally Valuable QTL Hotspots
Ksenija Gasic, Clemson University, Clemson, SC

W4**33: Fruit/Nuts**
New Insights into the Genome Evolution and Domestication of Apricots
Shuo Liu, INRA, BORDEAUX, France
Shuo Liu.[1,2], Amandine Cornille.[3], Tatiana Giraud.[4], Stéphane Decroocq.[1], David Tricon.[1], Aurelie Chague.[1], Alexis Groppi.[5], Macha Nikolski.[5], Wei-sheng Liu.[2], Veronique Decroocq.[1]
([1]UMR BFP, INRA-Univ. Bordeaux, Villenave d'Ornon, France. [2]Liaoning Institute of Pomology, Xiongyue, Yingkou City, China. [3]Génétique Quantitative et Évolution - Le Moulon, Gif-sur-Yvette, France. [4]Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France. [5]Bordeaux Bioinformatics Center, Université de Bordeaux, Bordeaux, France)
Long-lived perennials present evolutionary and domestication processes distinct from annual counterparts. Tree crop species tend to have been domesticated more recently, they are generally outcrossing, have extended juvenile periods, and are propagated clonally. Unravelling their histories of divergence should provide insights into the processes of adaptation and diversification and may help to identify the genomic bases underlying important agronomic traits that have been under selection during domestication.
Apricot, *Prunus armeniaca* L., is an excellent fruit tree for studying species evolution and domestication: while it is cultivated worldwide, it still exists in its wild form in natural populations of the Central Asian mountainous forests. Based on microsatellite data and approximate Bayesian computation, we inferred that the origin center of European/Irano Caucasian cultivars was Central Asia and substantial genetic differentiation between Chinese landraces and both European cultivars and wild Central Asian apricot trees (Decroocq et al, Molecular Ecology 2016).

Approximate Bayesian computation further revealed that the wild species *P. armeniaca* and *P. sibirica* diverged ca. 8 to 16 Mya ago, followed by divergence of the two cultivated apricot clusters, Chinese and European/Irano-Caucasian from wild *P. armeniaca* in north Central Asia. Furthermore, we assembled *de novo* a high-quality apricot reference genome sequence based on long reads and Illumina sequenced multiple sequences of wild and cultivated Armeniaca species. We performed a comparative genomic approach to assess the contribution of each wild related species into the apricot cultivated germplasm and identify loci under selection during apricot domestication, both in China and in Central Asia.

W434: **Fruit/Nuts**

Population Genomics Reveals the Postglacial Colonization History of European Woodland Strawberry and Adaptation to Extreme Climates

Tuomas Toivainen, University of Helsinki, Helsinki, Finland

W435: **Fruit/Nuts**

Emerging Genomic Resources for Wild and Elite Genotypes Shed Light on the Domestication of Strawberry

Michael A Hardigan, Department of Plant Sciences, University of California, Davis, Davis, CA

W436: **Fruit/Nuts**

Patricio Munoz: Exploring the Complexity of Autotetraploid Genomics for Plant Breeding

Patricio R. Munoz, University of Florida, Gainesville, FL

W437: **Fruit/Nuts**

Genetic Control of Primocane Fruiting and Dwarfism in Rubus

David Chagné, The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Like many perennial fruit crops, most *Rubus* species have a biennial cycle of flowering and fruiting and require a period of chilling to break bud dormancy in the spring of year two. Annual fruiting (AF) cultivars of raspberry and blackberry can flower and fruit in one season, without an intervening dormant period. . The ability to bypass dormancy enables AF cultivars to be grown in areas that do not accumulate significant winter chill. AF is a complex trait controlled by an interplay between genetic factors,the environment and crop management. We used a red raspberry (*Rubus idaeus*) population segregating for AF obtained from a cross between NC493 (AF) and 'Chilliwack' (biennial fruiting; BF). Genotyping by sequencing was performed to generate saturated linkage maps of both parents. Trait mapping indicated that AF is controlled by two loci (*RiAF3* and *RiAF4*) located on *Rubus* chromosomes 3 and 4. Both loci are syntenic with loci linked to recurrent flowering in strawberry. We verified the identified loci by using high resolution melting-based markers on independent AF x BF raspberry and blackberry populations segregating for the AF trait.

W438: **Fruit/Nuts**

Development and Application of Genomics Tools for Kiwifruit Breeding

John A. McCallum[1], Luis Gea[2], Elena Hilario[3], Paul Datson[4], Susan Thomson[5], Andrew Catanach[6], Ross N Crowhurst[3], Amardeep Nath[7], Sarah Pilkington[8], Jibran Tahir[9], David Chagné[10], Susan E Gardiner[10], Mareike Knäbel[10], Meeghan Pither-Joyce[11], Ed Morgan[10] and Alan Seal[12], (1)New Zealand Institute for Plant and Food Research, Christchurch, New Zealand, (2)The New Zealand Institute for Plant & Food Research Limited, Te Puke, New Zealand, (3)The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, (4)Plant & Food Research, Auckland, New Zealand, (5)The New Zealand Institute for Plant & Food Research LTD, Christchurch, New Zealand, (6)Plant and Food Research, Christchurch, New Zealand, (7)Plant and Food Research, Te Puke, New Zealand, (8)Plant and Food Research, Auckland, New Zealand, (9)New Zealand Institute for Plant and Food Research, Palmerston North, New Zealand, (10)The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand, (11)New Zealand Institute for Plant and Food Research, Christchurch, New Zealand, (12)New Zealand Institute for Plant and Food Research, Te Puke, New Zealand

Horticultural varieties of Kiwifruit (*Actinidia* spp.) have been developed by recent selection and hybridization of accessions originating mainly in southern China. The genus includes multiple species and varieties with differing ploidy, adaptation, architecture and fruit quality, which readily hybridize, offering many possibilities for innovative breeding. Development of commercial varieties involves pyramiding multiple production, post-harvest and consumer traits. Over the past decade both breeding and genetic studies have been greatly challenged by spread of the bacterial canker caused by *Pseudomonas syringae* pv. *actinidiae* (PSA). We describe development of resources to enable genomic analysis to understand domestication and historical selection targets in these pedigrees, dissect trait architecture and enable forward selection strategies. Development of genome assemblies for the female red kiwifruit 'HongYang' and more recently 'Red5' have been key advances enabling targetted and genome-wide variant analysis and QTL mapping. We will describe development and application of two new assemblies from a male and a monohaploid genotype which provide advantages for studies of sex-linked and quantitative traits. Whole-genome sequencing reveals a high level of genetic diversity in *Actinidia* and abundant haplotype-informative reads. We used the annotated 'Red5' genome and resequencing data to design a 10k bait targetted genotyping panel. Some preliminary results from this panel for GWAS and development of a high density genetic map in a tetraploid full-sib family will be discussed.

W439: **Fruit/Nuts**

Transcriptional Regulatory Network Controlling Strawberry Fruit Ripening and Quality

Carmen Martín Pizarro[1], Eva Lucas Reina[1], Almudena Trapero Mozos[1], Victoriano Meco[1], Maria A. De Luis Balaguer[2], José L Rambla[3], Tuomas Toivainen[4], Rosangela Sozzani[2], Antonio Granell[3], Miguel A. Botella[1], Victoriano Valpuesta[1], Timo Hytönen[4]

and **David Posé**[1], (1)IHSM - University of Málaga - CSIC, Málaga, Spain, (2)North Carolina State University, Raleigh, NC, (3)IBMCP, Valencia, Spain, (4)University of Helsinki, Helsinki, Finland
Ripening is a critical step for the development of flavor quality in fruits. This character has significantly declined in many fleshy fruits over recent decades. This is particularly significant in strawberry (*Fragaria* × *ananassa*), where current cultivars are derived from a narrow germplasm collection. Improving fruit quality requires two important breakthroughs: 1) a precise understanding of the fruit ripening process that will allow the targeting of relevant genes, and 2) the identification of novel alleles responsible for fruit quality traits.
In our project, we aim at the identification and characterization of key transcription factors involved in fruit ripening regulation and their target genes, in order to infer the Gene Regulatory Network controlling this process. On the other hand, we are carrying out a Genome-Wide Association Study using a germplasm collection of the woodland strawberry (*Fragaria vesca*) in order to identify loci involved in important traits such as aroma, fruit size or resistance to pathogens. Finally, we have implemented the use of the genome-editing tool CRISPR/Cas9 in the cultivated strawberry, which we expect it might open opportunities for engineering this species to improve traits of economic importance.

W4**40: Fruit/Nuts**

Glutathione-S-Transferase: A Candidate Gene for Berry Pigmentation in *Vitis rotundifolia*

Aruna Varanasi, Lacy Nelson and Margaret Worthington, University of Arkansas, Fayetteville, AR
Muscadines (*Vitis rotundifolia* Michx.), belonging to the Vitaceae family, are a specialty crop cultivated in the southern United States. Muscadines (2n=40) belong to the subgenus *Muscadinia* and are related to 'bunch' grapes of the subgenus *Euvitis* that includes *Vitis vinifera* (2n=38). Muscadines are often processed into juices and wines; however, they have poor color stability during storage which affects the quality of the processed products. Fruit color in muscadines is determined by anthocyanin accumulation in berry skins. In *V. vinifera*, berry color variation is primarily controlled by *VvMYBA1* located on chromosome 2. However, the color locus in muscadine was mapped to a region syntenic with 11.1 to 11.9 Mbp on chromosome 4 of *V. vinifera* in two mapping populations segregating for berry color. The objective of this research was to identify and characterize the candidate gene for color variation in muscadine berries. Of the 21 predicted genes spanning the 0.8 Mbp locus, glutathione S-transferase (*GST4*) was identified as a likely candidate gene involved in anthocyanin biosynthesis pathway. Other members of the GST family have essential roles in anthocyanin transport and accumulation including *Bz2* in maize, *An9* in petunia, *Fl3* in carnation, and *TT19* in *Arabidopsis*. PCR and sequencing from black and bronze muscadine DNA revealed a nonsynonymous SNP (C/T) in the *VrunGST4* sequence corresponding to a proline to leucine$_{171}$ mutation within the VrunGST4 protein (213 aa). An intragenic KASP genotyping assay confirmed the association of both CC and CT genotypes with black (dominant) and TT genotypes with bronze (recessive) phenotypes in 64 breeding selections, 32 cultivars, and 320 progeny from the two mapping populations segregating for berry color. Furthermore, *VrunGST4* expression was significantly higher in the ripe berries compared to unripe berries of black muscadines unlike in the bronze muscadine. These results validate the role of *VrunGST4* in muscadine berry color variation and suggest that berry pigmentation in muscadines may be regulated by a mechanism distinct from *V. vinifera*. The *VrunGST4* sequence from both black and bronze muscadines was cloned into pET19b vector for expression and purification of the VrunGST4 protein. Further research is in progress to determine the catalytic and/or the ligandin activity of the black and bronze VrunGST4 proteins and to study anthocyanin profiles in homozygous and heterozygous black muscadines to determine gene action (additive or dominant). These findings will have important implications for muscadine breeding programs and the muscadine processing industry.

W4**41: Fruit/Nuts**

Numerous Apple Quality Traits are Controlled by Large Effect Loci

Sean Myles[1], Zoë Migicovsky[1], Gan-Yuan Zhong[2], Bjarne Larsen[3], Carsten Pedersen[3] and Jun Song[4], (1)Dalhousie University, Truro, NS, Canada, (2)USDA-ARS, Geneva, NY, (3)University of Copenhagen, Frederiksberg, Denmark, (4)Agriculture and Agri-Food Canada, Kentville, NS, Canada
The degree to which marker-assisted selection (MAS) is useful depends crucially on the genetic architecture of the traits targeted by breeders: a marker must predict a large effect on an important phenotype to be worth adopting. Here we present the results of genome-wide association studies in various apple populations and argue that single markers may be useful for accelerating apple breeding for several apple quality traits. Several of our results confirm previous work. For example, the widely used marker to predict acidity at the MA locus was found to be the best predictor of acidity genome-wide and is thus deemed a useful marker for MAS. Similarly, markers at the well-characterized MYB1 locus are confirmed to be useful for selecting for skin colour. For firmness, we identify a marker in a NAC transcription factor that is a strong predictor of firmness, and show that markers currently used to predict firmness have no predictive power in our GWAS populations. In addition, we show that single markers may be useful for selecting for volatile compounds like butyl acetate, hexyl acetate; for the sucrose/fructose composition; for firmness retention during storage; and for several phenolics linked to health benefits like epicatechin, catechin and procyanidin. These initial results suggest that measuring chemical phenotypes linked to apple quality using high-throughput metabolomics approaches may significantly accelerate breeding for apple flavour and nutritional qualities using MAS. These insights also provide the foundation for future work aimed at apple quality improvement through genome editing.

W4**42: Fruit/Nuts**

GWAS and Genomic Prediction for Evaluating Breeding Populations and Germplasm Collection: Case Studies in Japanese Pear, Apple and Pecan

Hiroyoshi Iwata, The University of Tokyo, Tokyo, Japan

W4**43: Fruit/Nuts**

Genetic Diversity and Population Structure Analysis of the California Avocado Germplasm using SNP Markers

Nilwala S Abeysekara[1], Barbara Freeman[2], Brandon McKee[1], Aidan Shands[1], Mary Lu Arpaia[3], David N. Kuhn[2] and Patricia Manosalva[1], (1)Department of Microbiology and Plant Pathology, University of California Riverside, Riverside, CA, (2)USDA-ARS-SHRS, Miami, FL, (3)Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA

Avocado (*Persea americana*) is an economically important major fruit crop grown across the world with California being the largest producer in the United States. Avocado is subdivided into three botanical races: *Persea americana* var. *americana* (West Indian); var. *guatemalensis* (Guatemalan); and var. *drymifolia* (Mexican). Clear identification of the genetic diversity and population structure of avocado accessions is crucial for conservation and breeding purposes. Over 230 avocado accessions of the University of California-Riverside (UCR) germplasm collection were genotyped with 384 avocado single nucleotide polymorphism (SNP) markers. Markers were evenly distributed among all chromosomes and selected from a set of avocado SNPs generated from a previous study. Accessions were grouped by affinity propagation. A phylogenetic tree was generated with the neighbor joining method and data were further analyzed with Principal Components Analysis (PCA). All three analyses demonstrate that the UCR germplasm collection consists of accessions from all three landraces and hybrids with the majority being of Mexican origin. No West Indian x Mexican hybrids were detected in our collection. Mislabeling of accessions within the germplasm collection were also identified. Based on affinity group analysis, self-pollinated and outcrossed progeny of known maternal parentage were identified. Inference of potential paternal parents of progeny of known maternal parentage and assessment of the population structure of the germplasm collection are in progress. This study provides a new set of tools for an efficient avocado germplasm characterization and curation as well as for improving avocado rootstock breeding.

W44**4: Functional Annotations of Animal Genomes (FAANG)**
Skeletal Muscle eQTL and Allele-Specific Expression Associated with Phenotypic Traits in Pigs
Catherine W. Ernst, Department of Animal Science, Michigan State University, East Lansing, MI

W44**5: Functional Annotations of Animal Genomes (FAANG)**
Identifying Active Deubiquitinases and Kinases in Chicken
Bindu Nanduri, Mississippi State University, Mississippi State, MS

W44**6: Functional Annotations of Animal Genomes (FAANG)**
Livestock Rampage for High Definition Transcription Start Site Annotation
Pablo J. Ross, Department of Animal Science, University of California, Davis, CA

W44**7: Functional Annotations of Animal Genomes (FAANG)**
Epigenomic Landscapes from Various Cells and Tissues of *Gallus gallu*
Yvonne Drechsler, Western University of Health Sciences, Pomona, CA

W44**8: Functional Annotations of Animal Genomes (FAANG)**
Functional Annotation of the Porcine Genome and ISAFG Update
Christopher K. Tuggle[1], Huaijun Zhou[2], Catherine W. Ernst[3], Crystal Loving[4], James E. Koltes[5], James M. Reecy[5], Pablo J. Ross[6], Dan Nonneman[7], Timothy P.L. Smith[7], Juan P. Steibel[3] and Wen Huang[8], (1)Animal Science, Iowa State University, Ames, IA, (2)Animal Science, University of California, Davis, CA, (3)Department of Animal Science, Michigan State University, East Lansing, MI, (4)USDA-ARS-NADC, Ames, IA, (5)Iowa State University, Ames, IA, (6)Department of Animal Science, University of California, Davis, CA, (7)USDA, ARS, USMARC, Clay Center, NE, (8)Michigan State University, East Lansing, MI
While the porcine genome assembly has been significantly improved, the lack of functional annotation prevents full exploitation of single nucleotide variant analysis for genetic improvement through traditional breeding, and limits efficiency of gene editing of specific genome components to understand function and accelerate genetic improvement. With new funding from the USDA-NIFA-AFRI Foundational program, we will dramatically increase the depth and breadth of current porcine functional annotation and generate a community resource for scientists in both the public and private sectors. Across all objectives, we will use 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 annotate a tremendous breadth of biological states. In Objective 1, we will triple the number of adult tissues currently being annotated. In Objective 2, we will annotate four tissues at important stages of fetal development, and link allele-specific expression with allele-specific chromatin modifications through analyzing divergent-breed reciprocal crosses. In Objective 3, we will annotate circulating white blood cell populations in healthy pigs, as well as macrophage responses to bacterial and viral mimics. We will also profile gene expression of single cells in complex cell populations to maximize annotation of critical immune tissues. In Objective 4, we will use all these data as well as public FAANG data to generate a first-generation chromatin state segmentation map. At Livestock Genomics 2018, we will describe the current proposed research in detail as well as gather community input on biological priorities. Supported by USDA-NIFA-AFRI- 2018-67015-27501.
In November 2018, the Seventh International Symposium on Animal Functional Genomics (ISAFG) was held in Adelaide, Australia. This talk will include a short summary of this meeting, which included an accompanying FAANG session. This session consisted of updates from several groups working in FAANG research areas as well as small group discussions on wet lab methods and quality control and bioinformatics approaches. Summaries of these small group discussions were prepared and will also be recapitulated at the 2019 FAANG Workshop.

W44**9: Functional Annotations of Animal Genomes (FAANG)**
Genome Wide Identification and Annotation of Functional Regulatory Regions in Livestock Species
Huaijun Zhou, Animal Science, University of California, Davis, CA

W44**50: Functional Annotations of Animal Genomes (FAANG)**
FR-Agencode Update
Elisabetta Giuffra, INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France

W451: **Functional Annotations of Animal Genomes (FAANG)**
The FAANG Data Coordination Centre: Submitting and Retrieving Rich Datasets
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 develops the core infrastructure to support the community to create this rich genome to phenome resource. A key advantage of the project is its focus on ensuring high quality and rich supporting metadata to describe the project's samples and experimental assays. The DCC supports this endeavour through the provision of metadata validation tools (http://www.ebi.ac.uk/vg/faang/), a dedicated helpdesk (dcc-faang@ebi.ac.uk), and file conversion software, that ensures rich data submissions to the public archives and the rapid pre-publication of data in line with FAANGs data release policy (http://www.faang.org/data-share-principle). For retrieval of these rich FAANG datasets the DCC has created and hosts the FAANG data portal (http://data.faang.org/home). This 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. This site acts as a focal point for the FAANG community and the DCC aims to continually improve the filtering mechanisms and search interfaces to aid researchers in identifying appropriate data for their research. The portal provides direct links for downloading data files direct from the public archives and supports programmatic API access. 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.

W452: **Functional Annotations of Animal Genomes (FAANG)**
Global Run-on Sequencing and Computational Pipelines for FAANG
Douglas Antczak, Cornell University, Ithaca, NY and CJ Finno, RR Bellone, EN Burns and NB Kingsley (UC Davis), JL Petersen (U Nebraska), and DC Miller, EJ Rice, and CG Danko (Cornell).
The genome sequences of domestic livestock have dramatically advanced research in many areas. However, although the protein coding genes of most livestock species are now quite well defined, the non-coding elements that control gene transcription andother aspects of gene expression remain largely uncharacterized and are poorly understood. Such information is crucial to exploiting the full potential of genomics to advance animal production, health, and well-being. Members of the Equine Genome Community are engaged in the international Functional Annotation of Animal Genomes (FAANG) project. However, the approaches and assays of the FAANG project are expensive, time-consuming, and difficult to execute. Furthermore, resources for equine genomic research are limited. New methods for identifying regulatory elements have been developed since the human ENCODE Project was completed, and we are applying one of those to annotate gene expression in the horse. Chromatin run-on and sequencing (ChRO-seq) is a variant of global run-on and sequencing (GRO-seq), a technique that can precisely capture active transcription events in a cell population by focusing on engaged RNA polymerases at sites throughout the genome. When combined with a suite of computational programs recently developed at Cornell, a single ChRO-seq assay can reveal the location of thousands of non-coding functional elements, including enhancers, promoters, and long intergenic non-coding RNAs (lincRNAs). In addition, ChRO-seq generates information on nascent transcripts equivalent to conventional RNA-seq. We have tested ChRO-seq with three equine tissues: CD4+ T-lymphocytes, liver, and placenta (trophoblast), and compared ChRO-seq data from liver with results obtained using ChIP-seq and ATAC-seq. Our experience with ChRO-seq has shown it to be robust, reproducible, and sensitive. We found high correlation between marks identified by ChRO-seq and regions of the genome identified by ChIP-seq and ATAC-seq. ChRO-seq appears to produce data equivalent to that acquired using the traditional methods of the ENCODE Project, but with greater precision. ChRO-seq also has the potential to identify new regulatory elements not detected by earlier assays. Finally, ChRO-seq is efficient and cost effective. Thus, application of ChRO-seq should accelerate progress in the functional annotation of animal genomes.
Chu et al. 2018 Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. Nat Genet. 50:1553-1564.
Danko et al. 2015 Identification of active transcriptional regulatory elements from GRO-seq data. 2015 Nat Methods. 12:433-438.
Danko et al. 2018 Dynamic evolution of regulatory element ensembles in primate CD4+ T cells. Nat Ecol Evol. 2:537-548.
Supported by the Harry M. Zweig Memorial Fund for Equine Research in New York State and the Grayson Jockey Club Research Foundation. DFA is an investigator of the Dorothy Russell Havemeyer Foundation, Inc.

W453: **Functional Annotations of Animal Genomes (FAANG)**
FAANG-Related Project Updates - Canadian Efforts
Angela Cánovas, Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada

W454: **Functional Annotations of Animal Genomes (FAANG)**
The Ovine FAANG Project: What's Newe
Brenda Murdoch[1], Stephen N. White[2], Michelle R. Mousel[2], Kim C. Worley[3], Alan L. Archibald[4], Emily Clark[4], Brian Dalrymple[5], James Kijas[6], Shannon M. Clarke[7], Rudiger Brauning[7], Timothy P.L. Smith[8], Tracy Hadfield[9], Noelle Cockett[9] and Ovine FAANG, (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (3)Baylor College of Medicine, Houston, TX, (4)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (5)Institute of Agriculture, The University of Western Australia, CRAWLEY, Australia, (6)CSIRO Agriculture, Brisbane, Australia, (7)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (8)USDA, ARS, USMARC, Clay Center, NE, (9)Utah State University, Logan, UT

W456: **Functional Genomics**

Navigating NCBI Resources for Plant Genomics

Anjana Raina Vatsan[1], Terence D. Murphy[2] and Francoise Thibaud-Nissen[2], (1)National Center for Biotechnology Information, National Library of Medicine,, Bethesda, MD, (2)National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

The Eukaryotic Genome Annotation Pipeline at NCBI generates annotation of coding and non-coding genes, transcripts, and proteins from finished and unfinished genome assemblies publicly available in the International Nucleotide Sequence Database Collaboration (INSDC). These annotated genomes are in scope for manual curation by the Reference Sequence (RefSeq) group at NCBI (https://www.ncbi.nlm.nih.gov/refseq/) to generate high-quality, non-redundant transcripts with accessions starting with NM_/NP_, NG_, or NR_. Expert curation ensures accurate and full-length representation of nucleotide and protein sequences and to resolve data conflicts and ambiguities. Gene and protein names are assigned, and publications added, when available. Gene-specific data is available in NCBI's Gene resource (https://www.ncbi.nlm.nih.gov/gene/) which integrates information from multiple data sources, both internal and external. Gene annotation and related data can be explored in NCBI's Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) . RefSeq data can be accessed from the RefSeq homepage https://www.ncbi.nlm.nih.gov/refseq/ or can be downloaded from the FTP directory at ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Glycine_max/.

W4**57: Functional Genomics**

Comparative Omics between *Arabidopsis* and the Extremophyte, *Schrenkiella parvula* Identify Candidate Contributors to Boron Toxicity Tolerance

Guannan Wang, Dong-Ha Oh and Maheshi Dassanayake, Louisiana State University, Baton Rouge, LA

Boron toxicity is a worldwide problem for crop production. The genetic mechanism for adapting to high borate soils is largely unknown. *Schrenkiella parvula* is an extremophyte, naturally adapted to the high borate soils in the shores of Lake Tuz, Turkey. It can survive high borate at levels toxic to most plants. The *S. parvula* genome shares high macro-synteny with the genome of closely related *Arabidopsis thaliana* and both species have similar lifecycles. Therefore, *S. parvula* provides an ideal system to study the genomic blueprint for adapting to high soil borate. To identify the potential key contributors for boron tolerance, we have compared RNAseq profiles of borate-treated and control samples of shoot and root tissues from *S. parvula* and *A. thaliana* to explore the transcriptomic adaptations underlying boron tolerance with integrated insight from genome and ionome comparisons. We see a significantly reduced level of global expression changes in *S. parvula* to borate treatments compared to *A. thaliana*. The differently expressed genes (DEGs) between the two species were clustered to reveal genes that are differentially regulated in *A. thaliana*, compared to those clusters that exemplified "stress readiness" in *S. parvula*. One of the most highly expressed genes in *S. parvula* was annotated as *SpBOR5* due to its sequence similarity to the Arabidopsis *AtBOR5*. However, *AtBOR5* is an uncharacterized gene in the family of putative borate transporters. *SpBOR5* was able to complement the yeast *bor1* mutant much more efficiently than other boron transporters from both *S. parvula* and *A. thaliana*. We have also developed loss of function mutants of *SpBOR5* via the CRISPR-Cas9 system to test whether *SpBOR5*, compared to its ortholog in Arabidopsis, have been a significant contributing factor in the improved borate tolerance exemplified by *S. parvula*. This is likely one of the first extremophyte plants to have transgenic lines with targeted genome editing. We plan to present potential key contributing mechanisms for boron tolerance by comparing *A. thaliana* and *S. parvula* at the genome and transcriptome levels as well as preliminary data on the functional verification of selected candidates.

W4**58: Functional Genomics**

Transcriptome Analysis of Wheat Cultivar Molly Under Heat Stress and Hessian Fly Infestation

Jiazheng (John) Yuan[1], Lieceng Zhu[1], Ming Chen[2], Jordan O'Neal[1] and Daria Brown[1], (1)Fayetteville State University, Fayetteville, NC, (2)Kansas State University, Manhattan, KS

The Hessian fly [*Mayetiola destructor* (Say) HF] is one of the most devastating pests of wheat plants (*Triticum aestivium* L.) in North America and North Africa. The interaction of wheat and HF appeared to be gene for gene fashion where the expression of R gene of wheat plant can successfully defeat the pest attack resulting in the death of insect. However, the resistant effects of R genes of on wheat plants to HF infestation can be compromised by heat stress. Wheat variety Molly possesses HF resistance genes H13. In this study, we analyzed the RNA-Seq data generated from the seedlings of Molly under 6 hours of 35ºC heat stress with and without HF infestation. A significant gene expression change has been identified in almost all categories of cellular functions caused by heat stress compared to that of various corresponding treatments. The heat stress following HF infestation resulted in both up- and down-regulated major pathways associated with biotic and abiotic stresses such as hormonal signaling, cell wall, heat shock, and PR proteins. A total of 8839 differentially expressed genes (DEGs $\leq$-1 or $\geq$1 on $\log_2$ scale) were identified compared with that of HF infestation. In contrast, Hessian fly infestation caused 12 DEGs compared with uninfected controls and the up-regulated DEGs in this comparison were related with the functions of cell wall, abiotic stress, and signaling etc. Three down-regulated DEGs were found belonging to ethylene responsive element family associated with the biotic stress pathway. Our results also indicated that HF infestation after the heat treatment led to significant regulation of major biotic response at the gene expression level, suggesting that wheat cultivar Molly appeared to implement a failed strategy to combat the insect invasion compared to that of prior heat stress. Our findings have provided a unique platform to decipher the responses of plants to heat stress and analyzed the molecular mechanism of heat-induced loss of plant resistance. The information derived from the analysis of transcriptomic dynamics of wheat plant to heat treatments and Hessian fly infestation can be used to identify gene in wheat resistance to HF through reverse genetics and develop elite wheat varieties with resistance to Hessian fly even under heat stress.

W4**59: Functional Genomics**

Soybean Tilling by Target Capture Sequencing (TbyTCS) Applications: Identification and Functional Analysis of the GmSACPD Gene Family in Soybean

Khalid Meksem, Southern Illinois University Carbondale, Carbondale, IL and Naoufal Lakhssassi, Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL

Gel-based TILLING presents limitations to study complex traits such as oil composition due to multiple genes that are present with high copy number and chromosomal segment duplications within the soybean genome. In this study, we demonstrated the effectiveness of Tilling by target capture sequencing (TbyTCS) as an alternative and novel technique for the identification of several mutants within a mutagenized population (more than 4000 EMS mutants). TbyTCS is a method that allows quick identification of mutations within a large number of genes in a TILLING population, and it is a reliable method for the characterization of several genes simultaneously within a mutagenized population. The application of high-throughput NGsequencing technologies to a large mutagenized TILLING populations has proven to be a powerful tool for functional genomics, gene network, and gene pathways engineering. Interestingly, we successfully applied this technology to soybean and identified mutations in hundreds of genes involved in oil, protein, carbohydrate, auxin, and cytokinin pathways. In addition, mutants within large number of gene families controlling different agronomical traits such us flowering, nodulation, lateral root, flooding, etc. has been identified. Here, we demonstrate the effectiveness of the TbyTCS to study the Stearoyl-acyl carrier protein desaturase (*GmSACPD*) gene family in soybean. Characterization of the mutants reveals a novel role of the *GmSACPD-D* member in plant development and unsaturated fatty acid synthesis.

W460: **Functional Genomics**
Olive Genomics
Oussama Badad, University Mohamed The Fifth, Rabat, Morocco

W461: **Functional Genomics**
Comparative Omics Analysis of Stem Solidness in Wheat and its Synteny with Closely Related Species
Hikmet Budak, Montana State University, Bozeman, MT

W462: **Functional Genomics**
Phomopsis
Shuxian Li, USDA-ARS, Crop Genetics Research Unit, Stoneville, MS

W463: **Functional Genomics**
Single-Molecule Sequencing Reveals Increased Complexity of the Transcriptome Landscape in Maize and Sorghum
Bo Wang[1], Elizabeth Tseng[2], Primo Baybayan[3], Kevin Eng[3], Michael Regulski[1], Andrew Olson[1], Kapeel Chougule[1] and Doreen Ware[1], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)PacBio, Menlo Park, CA, (3)Pacific Biosciences, Menlo Park, CA
Maize and sorghum are both important genetic models for elucidating transcriptional networks. Uncertainties about the complete structure of mRNA transcripts limit the progress of research in this system. To better understand these two organisms at the molecular level, we compared expression profiles of both protein-coding and noncoding transcripts in 11 matched tissues using single-molecule, long-read, deep RNA sequencing. Our results show that characterization of the maize and sorghum transcriptome is far from complete, and that gene expression is more complex than previously thought. On the other side, haplotype phasing of genetic variants is important for interpretation of the genome, population genetic analysis and functional genomic analysis of allelic activity, however, due to splicing variability and sequencing length limitation, phasing at isoform level is always very challenge. Here, we present the first isoforms phasing study in maize using inbred lines B73 and Ki11, as well as their reciprocal crosses (B73xKi11, Ki11xB73) using full-length single-molecule sequencing. Our results show that maize parental lines and hybrid lines display different splicing activity, and 6,847 genes can be phased through Iso-Phase in two reciprocal hybrids using embryo, endosperm and root tissues. Our study identified parental origin isoforms in maize hybrids, different novel isoforms between maize parent and hybrid lines, provides measures of haplotypic expression that increase power and accuracy in studies of allelic expression. It is the first study of phased full-length isoforms in maize, as well as in plants, which provides insights about maize and plant heterosis at allele-specific full-length transcriptional level. The approach used in this study also provide important information for many other phasing studies in different species.

W464: **Functional Genomics**
RNAseq
Song Li, Virginia Tech, Blacksburg, VA
Accurate prediction of regulatory maps in plant using DAP-seq, ATAC-seq and single cell sequencing data
Recent advances in genomic technologies such as DNA Affinity Purification Sequencing (DAP-seq) and Assay for Transposase-Accessible Chromatin using Sequencing (ATAC-seq) have generated large-scale, regulatory genomic data for the multiple plant species. To predict condition specific gene regulatory networks using these data, we developed the Condition Specific Regulatory network inference engine (ConSReg), which combines heterogeneous genomic data using sparse linear model followed by feature selection and stability selection. Using Arabidopsis as a model system, we constructed comprehensive and accurate maps of gene regulation under more than 50 experimental conditions. Our results show that ConSReg accurately predicted gene expressions with an average auROC of 0.84 across these testing datasets. Including ATAC-seq information significantly improves the performance of ConSReg across all tested datasets. We applied ConSReg to Arabidopsis single cell RNA-seq data of two root cell types (endoderims and cortext) and identified five regulators in two root cell types. Three out of the five regulators are supported by existing publications. Finally, we tested our approach in a rice gene expression dataset and were able to identify both known and novel regulatory motifs that control drought response in the rice genome. Our results demonstrated that integrating heterogeneous genomic data can provide novel insights into the regulation of condition-specific and single cell-specific gene expression.

W465: **Functional Genomics of C_4 and CAM photosynthesis**
CAM Biodesign: Engineering Crassulacean Acid Metabolism into Arabidopsis to Improve Water-Use Efficiency

Sung Don Lim[1], Won Cheol Yim[2] and **John C. Cushman**[2], (1)Kangwon National University, Gangwon-do, South Korea, (2)University of Nevada, Reno, Reno, NV
Crassulacean acid metabolism (CAM) is a specialized photosynthetic mode to increase a water-use efficiency (WUE) that exploits a temporal CO_2 pump with nocturnal CO_2 uptake and concentration to reduce photorespiration to improve the adaptability of plants to hotter and drier climates. CAM species, with their inverted stomatal behavior, display water demands that are typically 4- to 7-fold less than of comparable C_4 and C_3 photosynthesis species, respectively. Thus, introducing the CAM pathway into C_3 photosynthesis plants (CAM Biodesign) is expected to confer enhanced photosynthetic performance and WUE. Detailed functional analysis of the individual genes encoding C_4 enzymes in common ice plant including *McβCA2*, *McPPCK1*, *McPPC1*, *McNAD(P)-MDHs*, *McNAD(P)-MEs*, *McPPDK*, and McPPDK-RP of both the carboxylation and decarboxylation modules and cognate circadian clock-controlled promoters are required to reconstitute the appropriate temporal expression of the CAM pathway enzymes in the C_3 model *Arabidopsis*. Furthermore, developing an effective multi-gene assembly tool for the large number of C_4 enzyme gene cassettes is necessary to ensure proper expression of each CAM gene cassette in the target species. Current steps achieved to date for CAM Biodesign will be summarized including subcellular localization and phenotypic analysis of individual ice plant C_4 enzyme 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*.

W466: **Functional Genomics of C_4 and CAM photosynthesis**
A Phylotranscriptomic Analysis of C4 Evolution in Blepharis (Acanthaceae)
Matt Stata, University of Toronto, Toronto, ON, Canada

W467: **Functional Genomics of C_4 and CAM photosynthesis**
Using a C_3+CAM Hybrid to Elucidate Genetic Regulation of CAM
Karolina Heyduk, Yale University, New Haven, CT
Crassulacean acid metabolism, or CAM photosynthesis, is an adaptation to water limitation whereby plants use inverse stomatal opening to increase water use efficiency. CAM plants assimilate CO_2 at night, when transpiration rates are lower, then store the fixed CO_2 as an organic acid overnight. In the day, behind closed stomata, the organic acid is decarboxylated and used for Rubisco-mediated CO_2 fixation. The entire CAM phenotype therefore requires not only the CAM biochemical pathway for nocturnal CO_2 fixation and storage, but also re-wiring of other metabolic processes, including guard cell signaling and carbohydrate processing. The genomic reorganization of transcriptional cascades is critical to the evolution of the CAM phenotype, and its elucidation is aided by the comparative study of closely related species. Here, we use a C_3+CAM hybrid species, *Yucca gloriosa* (Asparagaceae), as a model for understanding genetic re-wiring for CAM photosynthesis. Using a natural diversity collection of *Y. gloriosa* plants, we measured physiological traits, such as gas exchange and leaf acidity, under both well-watered and drought stressed conditions. Individual genotypes of *Y. gloriosa* have variable responses to drought – some can upregulate the CAM pathway when stressed, whereas others respond in a more C_3-like manner by shutting their stomata. RNAseq over a 24 hour period, sampled every four hours, revealed that the variable responses to drought stress across genotypes can be detected in gene expression changes. Canonical CAM pathway genes were differentially expressed across genotypes, largely mirroring the observed physiology. Surprisingly, many circadian clock genes showed variable expression across genotypes as well, suggesting circadian response to drought stress may mediate CAM induction. Further work incorporating parental genomes will help detect whether allele-specific expression or promoter origin play a role in the re-wiring of plant metabolism for CAM.

W468: **Functional Genomics of C_4 and CAM photosynthesis**
Factors Controlling Photosynthetic Efficiency in C4 Leaves
Xinguang Zhu, Center of Excellence of Plant Molecular Sciences, Institute of Plant Physiology and Ecology, CAS, Shanghai, China

W469: **Functional Genomics of C_4 and CAM photosynthesis**
Time of Day and Network Reprogramming during Drought Induced CAM Photosynthesis in the *Sedum album* Genome
Ching Man Wai, Michigan State University, East Lansing, MI

W470: **Functional Genomics of C_4 and CAM photosynthesis**
Controls of Patterning of Kranz Anatomy in C4 Grasses
Dhinesh Kumar[1], Jeffrey Yen[2] and **Elizabeth A. Kellogg**[1], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)Plant Biology, Ithaca, NY
The C_4 pathway in plants involves spatial separation of carbon fixation and carbon reduction and thus generally requires that the carbon fixation and reduction tissues be physically close to each other. The evolutionary origin of the C_4 pathway thus required modifications of C_3 vein structure and density. While considerable progress has been made understanding the evolutionary pattern and controls of C_4 biochemistry, consideration of the controls of C_4 anatomy has lagged behind, partly reflecting incomplete understanding of vein development and patterning in general. One hypothesis is that C_4 vein development represents a heterochronic shift whereby the signals that promote vein formation persist longer in leaf development and/or that the capacity to receive and respond to the signal persists. Recent data from a number of labs show that auxin and brassinosteroids both affect patterning of veins, but these data are mostly from C_3 rather than C_4 plants. This talk will review recent literature suggesting a possible link between auxin and brassinosteroids in controlling vein patterning. A second aspect of C_4 leaf development is formation of the secondary cell wall (SCW) in bundle sheath cells. In the C_4 plant maize, the bundle sheath cells are lignified and suberized, a phenotype thought to provide a mechanism for reducing CO_2 leakage. Comparisons of gene expression in different segments of rice and maize leaves found that lignified SCWs formed later in maize than in rice, a shift thought to be associated with C_4. However, we have investigated a set of 10 grasses including both C_3 and $_{C4}$ species, and find that all form lignified SCWs earlier than maize. In addition, in several lineages of C_4

grasses, the cells that reduce carbon do not have lignified walls. We suggest that lignin deposition may not be required for a functional C_4 pathway.

W471: **Fungal Genomics**
Phytophthora Genomics
Brett Tyler, Oregon State University, Corvallis, OR

W472: **Fungal Genomics**
Transposon-Mediated Mobilisation and Horizontal Transfer of a Host-Specific Virulence Gene in Three Wheat-Pathogenic Fungi
Adam P. Taranto[1], Peter S. Solomon[2], Megan C. McDonald[2], Benjamin Schwessinger[3], Andrew W. Milgate[4] and Zhaohui Liu[5], (1)University of California, Davis, CA, (2)The Australian National University, Canberra, Australia, (3)Australian National University, Canberra, Australia, (4)NSW Dept of Primary Industries, Wagga Wagga, Australia, (5)North Dakota State University, Fargo, ND
The effector gene *ToxA* is found in three fungal pathogens of wheat (*P. nodorum*, *B. sorokiniana*, and *P. tritici-repentis*) and conveys virulence on susceptible host cultivars. Using chromosome-level assemblies for multiple isolates of each species we identified extended regions of shared identity around *ToxA* and subsequently reconstructed a composite DNA-transposon responsible for mobilising *ToxA*. We demonstrate that capture of *ToxA* by this element occurred only once in an ancestral donor species and that the element is still functional in *B. sorokiniana*. In all three species the ToxA-element is found within a large repeat-rich island (140-250kb) with shared sequence identity, this island appears to have been the unit of transfer between species. We show this genomic island to be mobile within the genome of *B. Sorokiniana* and that the absence of *ToxA* in isolates of each species corresponds to the absence of the entire island. We propose that rare inter-specific hybridisation events facilitate introgression of mobile-element islands into new populations and that these islands act as rafts from which transposon-captured effectors may be mobilised.

W473: **Fungal Genomics**
Everything but the Kitchen Sink: Combining Genetics and Genomics to Investigate *Pyrenophora Teres* f. *Teres* Effector Biology
Nathan A. Wyatt, Dept. Veterinary and Microbiological Sciences, North Dakota State University, Fargo, ND

W474: **Fungal Genomics**
Next-Generation Interaction Screening for Genome-Wide Discovery of Host Targets of Fungal Effectors
Roger Wise[1], J. Mitch Elmore[1], Valeria Velásquez-Zapata[2], Sagnik Banerjee[2], Priyanka Surana[3], Greg Fuerst[1], Meiling Liu[2] and Dan Nettleton[2], (1)Corn Insects and Crop Genetics Research, USDA-Agricultural Research Service, Ames, IA, (2)Iowa State University, Ames, IA, (3)Plant Breeding Institute - University of Sydney, Sydney, Australia
Large-scale sequencing of plant and pathogen genomes has provided unprecedented access to the genes and gene networks that underlie host-pathogen interactions. The genome of the causal agent of barley powdery mildew, *Blumeria graminis* f. sp. *hordei* (*Bgh*), encodes 500-800 candidate secreted effector proteins (CSEPs). Pathogen effectors often interact with host proteins to suppress immunity and/or manipulate host physiology to enable colonization, however, the precise mechanisms of *Bgh* CSEP-mediated virulence are largely unknown. For this reason, we have undertaken a large-scale yeast two-hybrid (Y2H) approach to establish a protein-protein interaction framework of barley defense to the powdery mildew pathogen. Traditional Y2H library screening relies on picking yeast that grow on selective media to identify prey cDNA fragments from individual colonies. Recent approaches, collectively termed next-generation interaction screening (NGIS), use deep sequencing to score the output from Y2H screens. These advancements result in a massive increase in throughput, facilitate a quantitative measure of which preys interact with each bait protein, and importantly, do not require an ordered full-length ORF library from the host of interest. We used deep RNA-Seq of barley lines containing the *Mla6* resistance gene and derived immune system mutants inoculated with *Bgh* 5874 (AVR_{a6}) to profile pathogen gene expression from penetration of host epidermal cells through establishment of haustoria. Subsequently, 100 highly-expressed CSEPs were selected based on divergent expression patterns on resistant and susceptible barley mutants and used as baits to screen a comprehensive cDNA library derived from our infection time course. We developed an informatics pipeline to quantitatively score next-generation interaction data, including the determination of whether the identified prey cDNAs are translationally in-frame with the Gal4 activation domain. Our results have uncovered candidate CSEP interactions with barley metabolic enzymes, pathogenesis-related proteins, regulators of vesicle trafficking, defense signaling, and transcription. Currently these interactions are being validated and assayed for functional significance during *Bgh*-barley interactions. To place these interactions within the context of the host cellular signaling network, we constructed a predicted barley interactome using protein-protein interactions from other species. This network is being integrated with eQTL- and temporal gene expression data to predict active signaling modules during infection.

W475: **Fungal Genomics**
Genomics of Aspergillus
Adrian Tsang, Concordia University, Montreal, QC, Canada

W476: **Fungal Genomics**
Heterochromatin Influences Genome and Chromosome Dynamics in the Wheat Pathogen *Zymoseptoria tritici*
Mareike Moller, Christian-Albrechts-Universität Kiel, Kiel, Germany, Michael Habig, Christian-Albrechts Universität Kiel, Kiel, Germany, Michael Freitag, Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR and Eva H. Stukenbrock, Christian-Albrechts Universität & Max Planck Institute for Evolutionary Biology, Kiel, Germany

The haploid genome of the pathogenic fungus *Zymoseptoria tritici* is contained on “core” and “accessory” chromosomes. While 13 core chromosomes are found in all strains, as many as eight accessory chromosomes show presence/absence variation and rearrangements among field isolates. Accessory chromosomes are transcriptionally repressed, show enrichment of repetitive elements, and enrichment with heterochromatic histone methylation marks, e.g., trimethylation of H3 lysine 9 or lysine 27 (H3K9me3, H3K27me3). The factors influencing these presence/absence polymorphisms are so far unknown. We investigated chromosome stability using experimental evolution, karyotyping, and genome sequencing. We report extremely high and variable rates of accessory chromosome loss during mitotic propagation *in vitro* and *in planta*. Spontaneous chromosome loss was observed in 2 to >50% of cells during 4 weeks of incubation. Elevating the incubation temperature greatly increases instability of accessory and even core chromosomes, causing severe rearrangements involving telomere fusion and chromosome breakage. To elucidate the role of heterochromatin on genome stability in *Z. tritici,* we deleted the genes encoding the methyltransferases responsible for H3K9me3 and H3K27me3. Many genome rearrangements and formation of new chromosomes were found in the absence of H3K9me3, accompanied by activation of transposable elements. In stark contrast, loss of H3K27me3actually increased the stability of accessory chromosomes under normal growth conditions*in vitro*, even without large scale changes in gene activity. We conclude that H3K9me3 is important for the maintenance of genome stability and H3K27me3 reduces the overall stability of accessory chromosomes, generating a “metastable” state for these quasi-essential regions of the genome.

W477: Galaxy: An Open Platform for Data Analysis and Integration

Introduction to Galaxy and the Galaxy Ecosystem

Frederik Coppens, VIB, Gent, Belgium

This talk will introduce the Galaxy Project and ecosystem, and will cover ways to access Galaxy, and the user and administrator support ecosystems.

The Galaxy Project aims to empower biological researchers to do their own data integration and analysis without the need to learn computer programming, systems administration, or command line interfaces. If you, your project, or your institution is struggling with data analysis and reproducibility challenges, then Galaxy may be a platform that can help address these issues. Galaxy is available for free on the web (there are over 90 publicly accessible server), and can also be installed locally, or on the cloud.

Galaxy is a grant-funded open-source project that is deployed in hundreds, if not thousands, of organizations around the world. Galaxy is supported by ELIXIR, the European Research Infrastructural for data in life sciences, through the ELIXIR Galaxy Community. Galaxy is one of the supported data analysis platforms in the EOSC-Life project, which will build the life science gateway to the European Open Science Cloud.

W478: Galaxy: An Open Platform for Data Analysis and Integration

A Phenomenal Workflow to Study the Metabolites Variation in Bryophytes across Seasons

Kristian Peters[1], The PhenoMeNal consortium[2], Karin Gorzolka[1], Helge Bruelheide[3] and **Steffen Neumann**[1], (1)Leibniz Institute of Plant Biochemistry, Halle, Germany, (2)Several, Several, United Kingdom, (3)Martin Luther University Halle Wittenberg, Halle, Germany

Metabolomics as a high-throughput molecular phenotyping technique is growing across all domains in the life-sciences. The data processing and analysis is often performed with many programs using conventional computing solutions but little standardisation for interoperable and reproducible research. With increasing data size this becomes intractable for desktop computers. Cloud computing allows to instantiate on-demand resources (virtual servers, networks, storage), users only pay for the time the resources are used. Microservices can run in clouds that can dynamically grow or shrink, enabling applications to be scaled.

We developed a robust and performant data analysis infrastructure that integrates all necessary components. The software tools are encapsulated as Docker containers. To automate the instantiation of this cloud-portable microservice-based system, the PhenoMeNal project developed a Virtual Research Environment (https://portal.phenomenal-h2020.eu/) to deploy on some of the largest public cloud providers, including Amazon Web Services, Microsoft Azure, Google Cloud Platform and OpenStack-based scientific and private clouds. Kubernetes (https://kubernetes.io/) is used for container orchestration in the cloud. Galaxy (https://galaxyproject.org/) is used as interface for individual tools, users can share workflows and analysis histories.

Workflows in PhenoMeNal can combine noise reduction and filtering (OpenMS), quantification, alignment and matching (XCMS), filtering features based on blank and dilution series samples (R), feature annotation (CAMERA), statistics (ANOVA, PLS-DA) and identification (MetFrag), and support NMR and fluxomics data analyses.

To showcase data analysis from data acquisition to data sharing and -publication, we here present a dataset generated from 108 samples of nine bryophyte species obtained in four seasons using an untargeted liquid chromatography coupled with mass spectrometry acquisition method (LC/MS). Using minimum information guidelines for Metabolomics studies, metadata are encoded as ISA-Tab and uploaded together with the raw data to the metabolomics repository MetaboLights with the accession MTBLS520. The workflow presented here directly connects to the data repository and performs quantification, alignment and matching with XCMS and feature annotation with CAMERA, both established Bioconductor packages.

This was followed by statistical analyses using additional R packages. A metabolite presence-absence matrix was generated to determine the differences in metabolite features between the experimental factors *species* and *season*. The presence-absence matrix was used for measuring the metabolite richness for each species and season by calculating the Shannon diversity index for each sample.

Variation partitioning was performed with distance-based redundancy analysis (dbRDA) to analyze the influence of the factors species and seasons on the metabolite profiles. In order to validate the instrument performance and to detect batch effects between the instrument runs, a quality control (QC) protocol was followed.

Using this dataset we address the current challenges when processing Eco-Metabolomics data with a reproducible and reusable computational workflow implemented in Galaxy, focusing on standard formats, data import, technical validation, feature detection, diversity analysis and multivariate statistics. Together, we achieved a complete integration of several major metabolomics software suites resulting in a turn-key workflow for mass-spectrometry-based metabolomics.

W479: Galaxy: An Open Platform for Data Analysis and Integration

Galaxy Tools for Comparative Gene Family Analysis in Plant Genomics

Eric K. Wafula[1], Greg Von Kuster[2], Joshua P. Der[3], Saravanaraj Ayyampalayam[4], Norman Wickett[5], Loren A Honaas[6], Claude W. dePamphilis[1] and Jim Leebens-Mack[4], (1)The Pennsylvania State University, University Park, PA, (2)Penn State University, University Park, PA, (3)California State University, Fullerton, Fullerton, CA, (4)University of Georgia, Athens, GA, (5)Chicago Botanic Garden, Glencoe, IL, (6)USDA-ARS Tree Fruit Research Lab, Wenatchee, WA

Galaxy PlantTribes is a collection of automated modular analysis pipelines that utilize objective classifications of complete protein sequences from sequenced genomes for comparative and evolutionary analyses of genome-scale gene families and transcriptomes. It post-processes *de novo* assembly transcripts into putative coding sequences and their corresponding amino acid translations, estimates paralogous/orthologous pairwise synonymous/non-synonymous substitution rates for a set of gene sequences, classifies gene sequences into pre-computed orthologous plant gene family clusters, and builds gene family multiple sequence alignments and their corresponding phylogenies. A user provides *de novo* assembly transcripts and Galaxy PlantTribes produces: (1) predicted coding sequences and their corresponding translations, (2) a table of pairwise synonymous/non-synonymous substitution rates for either orthologous or paralogous transcript pairs, (3) results of significant duplication components in the distribution of *Ks* (synonymous substitutions) values, (4) a summary table for transcripts classified into orthologous plant gene family clusters with their corresponding functional annotations, (4) gene family amino acid and nucleotide fasta sequences, (6) multiple sequence alignments, and (5) inferred maximum likelihood phylogenies. Optionally, a user can provide an external gene family scaffold and/or externally predicted coding sequences derived from a transcriptome assembly or gene predictions from a sequenced genome. Galaxy PlantTribes is freely available on the Galaxy main portal (https://usegalaxy.org/). In addition, the standalone version of the pipeline is available for download on GitHub (https://github.com/dePamphilis/PlantTribes) as a command-line interface for batch processing of many datasets.

W4**80: Galaxy: An Open Platform for Data Analysis and Integration**

Galaxy Metabolomics from the Ground Up

Arthur C. Eschenlauer, University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics, St. Paul, MN, Mark A. Esler, University of Minnesota, Department of Horticultural Science, Saint Paul, MN, Timothy J. Griffin, University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN and Adrian D. Hegeman, University of Minnesota, Department of Horticultural Science, St. Paul, MN

Getting started with metabolomics software in Galaxy can be as simple as using one of the established public Galaxy instances, engaging high-performance computing resources, or running a Docker container on a workstation. A research-group specific Galaxy "appliance" may offer some advantages in the areas of collaboration, agile tool innovation, access, and simplified data transfer and life-cycle; this comes at the cost of developing a sustainable way to administrate the system. We have implemented a lab Galaxy environment comprising a laboratory intranet, a file server, a virtual workstation supporting concurrent users, and several Galaxy instances; for security, only the workstation and Galaxy instance are accessible from the campus WAN. We have applied this solution to our plant metabolomics analyses and found that, with minimal instruction, users can work independently and provide feedback on usability and functionality issues in prepublication versions of Galaxy tools. We are also exploring applying the practices that we have defined to encapsulate these functionalities into an appliance that can be implemented in a broad spectrum of laboratory settings.

W4**81: Galaxy: An Open Platform for Data Analysis and Integration**

Multi-Omics with Galaxy for Diverse Biological Applications

Tim Griffin, University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN and Pratik Jagtap, University of Minnesota, Saint Paul, MN

The Galaxy bioinformatics platform offers an ideal workbench for developing and using advanced multi-omic data analysis workflows. In particular, Galaxy offers the opportunity to develop integrated workflows which combine mass spectrometry-based proteomics data with genomic and transcriptomic data, to reveal new insights into molecular mechanisms. These workflows require integration of disparate tools from across 'omic domains, as well as the flexibility to incorporate new and emerging software of high value. Galaxy provides the working environment required to develop these workflows. Over the last several years, the Galaxy for proteomics (Galaxy-P) research team has been developing multi-omic workflows which integrate MS-based proteomics data with genomic/transcriptomic data and make possible identification of protein sequence variants and unexpected protein products. This approach is called proteogenomics and is particularly well-suited for non-model plant and animal organisms of interest. The team has also focused on workflows for characterizing proteins expressed by microbial communities (microbiomes), many times living within a host animal. This approach, called metaproteomics, integrates metagenomic and proteomic data, and enables a more informed snapshot of the functional state of these communities compared to analysis by metagenomics alone. We will provide an update on these multi-omic workflows and present example applications of this software to studies relevant to plant and animal researchers. Information on access to these tools via publicly available resources will also be discussed, to promote their utilization by the research community.

W4**82: Galaxy: An Open Platform for Data Analysis and Integration**

Eukaryotic Genome Annotation with G-OnRamp

Luke Sargent, Oregon Health & Science University, Portland, OR

W4**83: Gene Expression Analysis**

Introduction to the Gene Expression Analysis Workshop

David W. Galbraith, BIO5 Institute & School of Plant Sciences, University of Arizona, Tucson, AZ

W4**84: Gene Expression Analysis**

Dynamic Transcriptional Landscape of Polyploid Plants

Avinash Sreedasyam, HudsonAlpha Institute for Biotechnology, Huntsville, AL

W485: **Gene Expression Analysis**
Single-Cell RNA-Seq for Maize Germinal Precursors
Bradlee D Nelms, Stanford University, Stanford, CA

W486: **Gene Expression Analysis**
Single-Cell Gene Expression from Arabidopsis Roots
John Schiefelbein, University of Michigan, Ann Arbor, MI
Single-cell RNA sequencing (scRNA-seq) has been used extensively to define and compare gene expression in individual cells from animal tissues, but it has not been widely applied to plants. Here, I describe our use of a commercially available droplet-based platform for high-throughput scRNA-seq to obtain more than 10,000 single-cell transcriptomes from Arabidopsis root cell protoplasts. We find that all major tissues and developmental stages of roots are represented in this single-cell transcriptome population. Further, transcriptomes corresponding to distinct cell sub-populations and rare cell types, including putative quiescent center (QC) cells, were identified. A focused analysis of transcriptomes from the epidermal cells defined individual cells progressing from meristematic through mature stages of root-hair and non-hair epidermal cell differentiation. Further, a pseudo-time analysis was used to infer the developmental trajectories for the root-hair and non-hair cell types. In addition, single-cell transcriptomes were obtained from two different root epidermal mutants, enabling a comparative analysis of gene expression at single-cell resolution and providing an unprecedented view of the impact of the mutated genes. Overall, this study demonstrates the feasibility and utility of scRNA-seq in plants and provides a first-generation gene expression map of the Arabidopsis root at single-cell resolution.

W487: **Gene Expression Analysis**
Sampling Individual Plant Cells using Laser Ablation Electrospray Ionization-Mass Spectrometry
Sylwia Stopka[1], Christopher Anderton[2], Laith Samarah[3], Rikkita Khattar[3], Dusan Velickovic[2], Jared Shaw[2], Beverly Agtuca[4], Ljiljana Pasa-Tolic[2], Akos Vertes[3] and **Gary Stacey**[5], (1)George Washington University, Washingtonq, DC, (2)Environmental Molecular Sciences Laboratory, Richland, WA, (3)George Washington University, Washington, DC, (4)Univ of Missouri, Columbia, MO, (5)University of Missouri, Columbia, MO
The ability to visualize complex metabolic behavior of living organisms *in situ* and at the single cell level represents a bioanalytical grand challenge that has yet to be fully realized. Herein, we present the combination of fiber-LAESI with fluorescence microscopy and ultrahigh resolution 21T-FTICR-MS to address these needs.To provide simultaneous spatial distributions for hundreds of metabolites in biological tissues, an LAESI ion source was coupled with a 21T FTICR-MS system. Utilizing the ultra-high mass resolving power, isotopic fine structure readout was used for metabolite annotations. Single-cell spatial resolution was achieved by delivering mid-IR laser pulses through a sharpened optical fiber, while a dual modality microscopy allowed for precise targeting of cells.
Spatial heterogeneity was tackled using LAESI coupled with the 21T-FTICR-MS. This unique instrumental configuration provides exquisite mass and spatial resolution, down to the single cell level. A dual channel microscope capable of both simultaneous brightfield and fluorescence imaging was combined with the fiber-LAESI-MS to enable selection of specific cells. For instance, cells in a soybean root nodule infected by GFP labeled soil bacteria were targeted for analysis in a heterogeneous population of infected and uninfected cells. The superior mass resolution and accuracy offered by the 21T FTICR-MS facilitated identification of differentially abundant metabolites in soybean root nodule cells infected by *wt* and *nifH-* rhizobia. Notably, this novel capability facilitated characterization of organometallic molecules due to their natural isotope patterns (e.g. ^{57}Fe isotopologues). Preliminary data indicates significantly lower abundance of heme b, disaccharide, and S-adenosylmethionine in the *nifH-* mutant.
We demonstrate that LAESI coupled with the 21T FTICR-MS holds tremendous potential for *in-situ* single cell metabolomics as illustrated by the study of plant-microbe interactions. The ability to resolve metabolites in single cells reveals a level of metabolic complexity that is normally hidden due to cell averaging seen when whole tissues, composed of many cell types, are analyzed. Hence, *In situ* single cell analysis by fiber-LAESI-MS promises new insights into cellular heterogeneity and metabolic noise.

W488: **Gene Expression Analysis**
Evolutionarily Informed Deep Learning Methods: Predicting Relative Transcript Abundance from DNA Sequence
Jacob D. Washburn, Cornell University, Ithaca, NY

W489: **Gene Expression Analysis**
Alternative Splicing adds Regulatory Complexity in Differential Ortholog Expression between Arabidopsis and Its Extremophyte Relative *Schrenkiella parvula*
Chathura Wijesinghege[1], Maheshi Dassanayake[2] and Dong-Ha Oh[2], (1)Lousiana State University, Baton Rouge, LA, (2)Louisiana State University, Baton Rouge, LA
Schrenkiella parvula, an extremophyte closely related to *Arabidopsis thaliana*, is native to hypersaline lakes in Turkey, where soils contain toxic levels of [Na^+], [K^+], [Li^+], and [BO_3^{3-}]. The genome of *S. parvula* shows a high level of synteny with that of *A. thaliana*. Despite the genome-wide synteny, we found substantial differences in the isoform structure and expression between *S. parvula* and *A. thaliana*. We hypothesized that some of these splice variants in *S. parvula* are associated with its extremophyte lifestyle adapted to multi-ion salt stress. We used Iso-Seq to obtain long read sequences of transcripts from *S. parvula* coupled with short RNA-seq read sequences to quantify isoform specific expression in response to multiple salt stresses. We identified novel transcript models previously undocumented in the genome for 11,484 *S. parvula* gene models, majority of which showed intron retention events. Additionally, we assembled 1,411 novel gene models previously unannotated in the reference genome. A comparison between *A. thaliana* and *S. parvula* orthologous gene pairs identified a significant enrichment in *S. parvula* for

isoform availability in genes related to abiotic stress responses. Differential alternative splicing combined with differential gene expression analysis revealed unique alternatively spliced variants that are expressed under stress and tissue specific conditions, implying their potential roles in the stress adapted lifestyle of *S. parvula*.

W490: **Gene Introgression**
Back to the Roots: Wild Emmer Introgressions Promote Drought Tolerance in Wheat
Zvi Peleg, The Hebrew University of Jerusalem, Rehovot, Israel

W491: **Gene Introgression**
Leveraging Alien Genetic Variation through Chromosome Engineering to Enhance Security and Safety of the Durum Wheat Crop
Ljiljana Kuzmanović, University of Tuscia, Viterbo VT, Italy

W492: **Gene Introgression**
Exploiting Introgressions to Increase the Genetic Diversity of Durum Wheat
Manel Othmeni, University of Nottingham, Nottingham, United Kingdom

W493: **Gene Introgression**
Outsourcing Rice and Wheat Genes from Wild Relatives
Antonio Costa De Oliveira[1], Luciano C. da Maia[2], Vivian Viana[3], Eduardo Venske[4], Mariana Rosa[5], Camila Pegoraro[5] and Railson Schreinert Santos[4], (1)Universidade Federal de Pelotas, Pelotas-RS, Brazil, (2)Federal University of Pelotas, Pelotas-RS, Brazil, (3)Federal University of Pelotas, Pelotas, Brazil, (4)Federal University of Pelotas, Capão do Leão, Brazil, (5)Federal University of Pelotas, Capao do Leao, Brazil
The constant increase in the world's population demands for a sustainable intensification of agriculture, where the discovery of novel genes to boost yields and adaptive characters is an important strategy. The major cereals such as rice and wheat need new sources of variation in order to promote higher genetic gains in breeding programs. Our group has been involved in International Initiatives to characterize and transfer genes into cultivated genotypes of wheat and rice. Through collaborative work with the University of Nottingham (wheat) and the University of Arizona (IOMAP), new introgressed populations, genomic sequences and genomic tools have been developed to improve wheat and rice. A backcrossing program assisted by SNP molecular markers and genomic *in situ* hybridization (GISH) was carried out, where 12 wheat lines containing introgressions from the wild relative were utilized as donors of segments to the Brazilian cultivar TBIO Sinuelo. A genetic map containing 537 markers and characterization of 236 introgressed segments were obtained. Characterization of the sub1 locus in the genomes of rice wild relatives was also performed. A deeper insight into the evolutionary origin of the SUB1 locus across the *Oryza* genus was obtained, raising the possibility of association of these genes with flooding tolerance in the outgroup grass, *L. perrieri*.

W494: **Gene Introgression**
Induced Mutation for Crop Biodiversity and Crop Improvement
Ljupcho Jankuloski, IAEA, Joint FAO/IAEA Division, Vienna, Austria

W495: **Gene Introgression**
Introgression from Diploid Wild Relatives into Tetraploid Peanut
Scott A. Jackson, University of Georgia, Athens, GA

W496: **Gene Mapping by Segregation**
Advances in Gene Mapping by Segregation
Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA, Michael Peel, USDA, Logan, UT, Deven R. See, USDA-ARS/Washington State University, Pullman, WA and Long-Xi Yu, USDA-ARS Plant Germplasm Introduction and Testing Research, PROSSER, WA
As the most widely cultivated forage legume in the world, alfalfa has been successfully developed with varieties specialized on simple traits controlled by major genes, such as winter hardiness and disease resistance. However, genetic improvements have been limited for complex traits such as forage and seed yields and forage quality, which are controlled by multiple genes. Alfalfa breeders are unable to develop real inbred lines to accumulate favored alleles and create subsequent hybrids with increased vigor due to severe inbreeding depression. Natural Selection by Inbreeding Depression (NSID) commonly terminates alfalfa fertility before eliminating deleterious alleles. Our preliminary studies suggest that this hurdle can now be overcome by jointly using Bulked Segregant Analysis (BSA) and Marker-Assisted Selection (MAS) in combination with NSID to purge deleterious alleles before they express harmful effects. Based on the genotypes of ~5,000 SNPs on the 200 accessions of the alfalfa diversity panel, we propose to divide these accessions into two heterotic galaxies for developing inbred lines. For each galaxy at each generation, DNA will be pooled for whole genome sequencing within progeny unable to germinate and progeny with the lowest and highest seed counts. Multiple BSA across galaxies and generations will be used to identify the genetic loci associated with fertility. The top-seeded plants will be genotyped for the identified loci by using targeted amplicon sequencing for MAS. The two galaxies of inbred lines are expected to be a turning point for the alfalfa industry to produce real hybrids to boost alfalfa yield and quality.

W497: **Gene Mapping by Segregation**
Pending

Rohan L. Fernando, Department of Animal Science, Iowa State University, Ames, IA

W498: **Gene Mapping by Segregation**

Dissecting Morphological Variation by Near-Isogenic Introgression Lines (NIILs) in *Brassica oleracea*

Lan-Shuan Shuang[1], Hugo E. Cuevas[2], Cornelia Lemke[1], Changsoo Kim[3], Tariq Shehzad[1] and Andrew H. Paterson[1], (1)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, (2)ARS - Tropical Agriculture Research Station, Mayaguez, PR, (3)University of Georgia, Athens, GA

The single species *Brassica oleracea* has great economic importance, remarkable morphological variation and genome complexity. To dissect the genetics of morphological diversity within the species, two backcross populations were developed by using inbred lines of cabbage (Badger Inbred) and cauliflower (Orange) as donor parents, and a rapid cycling line (TO1434) as the recurrent parent. Based on the populations, we have selected 97 and 104 introgression lines, which represent 82.09% and 78.56% of the cabbage and cauliflower genomes, respectively. Selfed-progenies of advanced backcross populations and subsets of introgression lines have been planted in the field for five and three seasons, respectively, investigating 15 leaf-, stem-, and flower-traits across the eight seasons. Among numerous phenotypes associated with DNA markers in the populations, two illustrate the value of these populations to dissect traits at fine resolution. In the cauliflower population, lines carrying introgression on the end of C8 showed stressed and dwarf phenotypes. Single marker analyses associate this phenotype with a 3.3 Mb region on C8. In the cabbage population, lines carrying introgression in a 0.9 Mb region at the top of C2 reached large plant size and were late flowering in all environments. Several positional candidate genes orthologous to flowering time genes in *Arabidopsis thaliana* are being explored by functional and evolutionary analyses.

W499: **Gene Mapping by Segregation**

Genome-Wide Association Mapping on Nutrient Uptake and other Related Traits under Direct-Seeded Cultivation using a Subset of the 3K Rice Diversity Panel

Endang M. Septiningsih[1], Sumeet Mankar[1], Nitika Sandhu[2] and Arvind Kumar[2], (1)Texas A&M University, College Station, TX, (2)International Rice Research Institute, Los Banos, Philippines

Direct-seeded rice (DSR) is a common practice for growing rice in America, and this system is also becoming increasingly popular in Asia. One of the challenges with DSR is maintaining nutrient dynamics which affect agronomic traits and ultimately yield. The present study aims to dissect the genetic control of root traits associated with nutrient uptake under DSR through genome-wide association studies (GWAS) using a subset of 315 rice accessions originating from South and South-East Asia from the sequenced 3K diversity panel. Field experiments based on an alpha lattice design with two replications were conducted at the International Rice Research Institute (IRRI), Philippines, during the dry season (DS) and wet season (WS) 2018. Various root traits and agronomic traits along with SPAD and LCC readings in booting stage, and concentration of nutrients in the plant samples, i.e. N, P, K, Fe and Zn, were measured. GWAS was performed using the GAPIT (Genome Association and Prediction Integrated Tool) package, based on the compressed mixed linear model (CMLM). Significant marker-trait associations were detected among various traits. Further analysis is underway. Understanding the genetic architecture and molecular function of macro and micro nutrients and the development of functional DNA markers will provide a strong foundation for marker-assisted breeding programs to develop nutrient-efficient high-yielding rice varieties.

W500: **Gene Mapping by Segregation**

QTL Associated with Yield Components Identified in Two Spring Wheat Mapping Populations

Jianli Chen, University of Idaho, Aberdeen, ID

W501: **Gene Mapping by Segregation**

Pending

Ismail Rabbi, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

W502: **Gene Mapping by Segregation**

Computer Vision Solutions for Deciphering Genome-to-Phenome Relationships in Plant Breeding

Todd DeZwaan, LemnaTec Corporation, RTP, NC

Phenotyping is central to the human experience of the natural world. Phenotyping drove the domestication events that are hallmarks of the Neolithic revolution, and underlies the yield improvements recognized by modern agriculture. Gregor Mendel phenotyped traits in pea plants to formulate the Laws of Inheritance describing equal segregation, independent assortment, and dominance of alleles. Increasingly sophisticated phenotyping tools are needed to recognize the trait-improvement potential of genome-wide association studies in areas such as phenotypic differentiation of complex traits, phenotypic plasticity across environments, and phenotypic validation of genomic predictions. LemnaTec fulfills this need by developing computer vision tools for high-resolution plant phenotyping in controlled-environments and fields. LemnaTec integrates industrial sensors and illumination with powerful analytical software and robotic automation to deliver high-quality digital phenotypic data for agricultural research and product development. LemnaTec multi-sensor systems measure parameters in 2D images and 3D laser scans across the wavelength spectrum including the visible range, near-infrared, hyperspectral, and PAM fluorescence. The results enable breeders to derive a comprehensive digital phenotype that describes plant growth, development, color, geometry, stress response, disease status, and much more. LemnaTec delivers a wide range of solutions for controlled-environment studies including instruments that offer complete control over temperature, humidity, CO_2 concentration, and light quality and intensity. Controlled-environment phenotyping enables year-round, high-throughput trait discovery under precise low-noise conditions. LemnaTec also delivers solutions for high-resolution non-destructive multi-sensor field phenotyping from both fixed gantries and lightweight mobile imaging modules that can be transported and deployed across multiple field locations. Finally, LemnaTec offers a soil-based solution for root phenotyping that delivers root traits used by breeders to improve plant health and productivity. As a soil-based system it enables studies that are not possible with artificial substrates and hydroponic systems including root –

root interactions, root – substrate interactions, effects of soil compaction, and effects of soil-borne microbes. This wide range of solutions provides breeders with the phenomic tools they need to keep pace with the trait improvement gains offered by genome-wide association studies.

W503: Gene Mapping by Segregation

Challenges for Polyploid Tropical Grasses Breeding: Genetic Mapping using Allele Dosage

Anete P Souza[1], Rebecca C. U. Ferreira[1], Leticia A. C. Lara[2], Lucimara Chiari[3], Sanzio Barrios[3], Cacilda do Valle[3], Jose R Valerio[3], Fabricia Z. V. Torres[3] and Antonio A. F. Garcia[2], (1)University of Campinas, Campinas, Brazil, (2)University of São Paulo (ESALQ/USP), Piracicaba - SP, Brazil, (3)Embrapa Beef Cattle, Campo Grande, Brazil

Tropical forage grasses are widely sown across the global, and especially American tropics, where they have drastically increased the efficiency of beef cattle production. *Urochloa* spp. and *Megathyrsus maximus* are most economically important forage grasses in Brazil, where they are planted on 170 Mha, accounting for 85% of sown pasture land. These forage grasses are polyploid and apomictic, which makes genetic studies challenging; therefore, the number of currently available genetic resources is limited affecting the advancement of molecular breeding. Recent advances allow generating high-density single nucleotide polymorphism (SNP) genotype data and estimating the allele dosage of SNP for mapping studies in these species. Dense GBS genetic maps have been developed from intraspecific F_1 progeny with contrasting parents to several traits for four important forage grass species: *Urochloa decumbens*, *U. humidicola*, *U. ruziziensis* and *Megathyrsus maximus*. These are the first genetic maps for tropical grasses using SNPs with allele dosage, making possible most robust genetic analysis and providing relevant genetic information. Genetic map of *U. decumbens* allowed detecting three QTL to spittlebug (*Notozulia entreriana*) resistance, an important insect pest that attacks the Brazilian pastures, providing new insights into the architecture of this trait. Our results represent an essential evolution for polyploid studies and are the first step towards possible marker-assisted selection (MAS) and genomic selection (GS) in tropical forage grass, subjects of great interest in genetic improvement programs for economically important plant species.

Funding: FAPESP, CAPES - Computational Biology Program; CNPq

W504: Gene Mapping by Segregation

Multi-Trait and Time Series Quantitative Genetic Associations in Maize and Sorghum

James C Schnable, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE

Genome-Wide Association Studies (GWAS) attempt to link phenotypic variation to specific genetic variants across the genome. High throughput phenotyping technology will accelerate the trend towards many distinct traits being characterized in the same association populations, and mining detailed medical records has done for human data. High dimensional trait datasets present both challenges and opportunities to test associations between specific genes and variation in target trait(s). Using published data for a set of 282 maize lines subjected to whole genome resequencing and 260 real world trait datasets, a Genome-Phenome Wide Association Study (GPWAS) model demonstrates its ability to identify genes more enriched in maize GO terms, experiencing greater purifying selection, and exhibiting higher recall rates than conventional GWAS with genes identified as linked to phenotyping variation in a separate population.

W505: Genome annotation resources at the EBI

Browsing Genes and Genomes with Ensembl and Ensembl Genomes

Erin Haskell, EMBL-EBI, Hinxton, United Kingdom

This workshop will include an introduction to the Ensembl browsers, demonstrate key views in browsing genomes, and show you how to use tools for accessing genomic data and analysing your own, BioMart and the Variant Effect Predictor (VEP).

Ensembl (www.ensembl.org) provides an interface and an infrastructure for accessing genomic information covering over 100 vertebrate species, including chicken, cow, pig, and sheep. Its sister project, Ensembl Genomes (www.ensemblgenomes.org), consists of five sub-portals (bacteria, fungi, invertebrate metazoa, plants, and protists) that contain data for over 1,000 eukaryotic genomes (including bread wheat and its progenitors, barley, tomato, brassicas, and many important pests and pathogens), and over 40,000 prokaryotic genomes.

All species in Ensembl and Ensembl Genomes have gene annotation and comparative genomics analyses within the taxa (excluding bacteria). For many of these genomes, we also provide annotation of variants, such as SNPs and CNVs. All these data can be accessed via our browser websites, BioMart, FTP, Perl APIs, REST API, and MySQL. Furthermore, the VEP is a powerful tool for analysing sets of genomic variants, available for all species in Ensembl and Ensembl Genomes.

W506: Genome annotation resources at the EBI

Gencode Manual Genome Annotation in Ensembl

Jane Loveland, EMBL-EBI, Cambridge, United Kingdom and On behalf of the Ensembl-HAVANA team, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambs UK.

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. In the recent past 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.

First pass manual annotation of the mouse reference genome has now been completed. We are now working on extensive QC to continue progress to determine the final protein-coding gene count, both adding gene previously unannotated in any reference geneset and removing legacy protein-coding genes that are not robustly supported in the light of currently available datasets.

The development of long transcriptomic sequencing methods such as PacBio, SLRseq and ONT is significant for gene and transcript annotation. We are incorporating new long read transcriptomics datasets into GENCODE via a manually supervised automated workflow to create and modify transcripts, allowing the manual annotators to more rapidly assess prospective gene models and leveraging computational methods to accurately identify complex regions that require manual annotation.

The Matched Annotation from NCBI and EMBL-EBI (MANE) project is a new joint initiative in collaboration with NCBI's RefSeq group. This project aims to agree a genome-wide transcript set containing one well-supported transcript (MANE_select) per protein-coding locus in human. These transcripts align perfectly to GRCh38 with complete identity (5' UTR, CDS, 3' UTR) between RefSeq (NM) and Ensembl (ENST) transcripts. A beta release of the transcripts agreed to date is available via trackhub.
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/gencode/update_trackhub/hub.txt
The annotation can be downloaded from gencodegenes.org and is available from the Ensembl and UCSC genome browsers.

W507: Genome annotation resources at the EBI

VGNC Update

Tamsin Jones, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

W508: Genome annotation resources at the EBI

Annotare; Arrayexpress and Expression Atlas: Submission, Archival and Visualisation of Functional Genomics Data at the EBI

Nancy George[1], Nuno Fonseca[1], Anja Fullgrabe[1], Laura Huerta[1], Haider Iqbal[1], Monica Jianu[1], Jon Manning[1], Pablo Moreno[1], Alfonso Munoz-Pomer[1], Lingyun Zhao[1], Alvis Brazma[1], Irene Papatheodorou[1] and Functional Genomics, (1)European Bioinformatics Institute, Hinxton, United Kingdom
To serve the functional genomics community, we have developed a number of web-based services, from data submission via Annotare to analysing and visualising gene expression data in Expression Atlas.
Annotare (www.ebi.ac.uk/fg/annotare) is a simple; intuitive web submission-tool for functional genomics datasets. The newly introduced template selection page allows users to tailor their submissions to their material and technology types. Once selected, these display the minimal technical information and metadata required to recreate the experiment, promoting data reanalysis and reproducibility. Metadata is mapped where possible to Experimental Factor Ontology (EFO) terms for search expansion and data retrieval. Upload of raw and processed data is provided either by direct upload or via FTP and Aspera. Upon submission, datasets are given a stable; citable accession; reviewed by curators and uploaded to ArrayExpress – a functional genomics data archive. The archive (www.ebi.ac.uk/arrayexpress) displays each dataset as a self-contained entity providing the experiment information; sample metadata and links to raw and processed under a single accession. Raw sequencing data is stored at the European Nucleotide Archive whilst raw microarray and processed data is stored at ArrayExpress. Submitter identity can be hidden for double-blind review and data is kept private until public release or publication via personal and reviewer logins. Related datasets within ArrayExpress and EBI databases can also be crosslinked to each other. For data discovery, searching is simplified using ontology expansion and refined using our advanced search features, allowing users to search within the sample metadata as well as the whole dataset.
Expression Atlas (www.ebi.ac.uk/gxa) and its newest component: Single Cell Expression Atlas (www.ebi.ac.uk/gxa/sc) are added-value databases where bulk and single-cell RNA-seq data are collected; annotated and reanalysed in a consistent manner via our standardised pipeline iRAP (https://nunofonseca.github.io/irap). Expression studies are derived from archives including ArrayExpress; GEO and ENA. Eligible datasets are curated to ensure accurate and comprehensive metadata that is then mapped to Experimental Factor Ontology terms. Comparison groups are identified based on experimental variables. Studies are displayed either as baseline i.e. constitutive gene expression or differential i.e changes in gene expression as a result of a perturbation. Currently we have analysed over 3,450 experiments from over 45 species; including landmark studies such as BLUEPRINT, GTEx, ENCODE, CCLE, HipSci and PCAWG and selected protein expression datasets. Expression Atlas functionality includes transcript information; visualisation of gene expression in Ensembl and Gramene genome browsers and Gene Omnibus pathway and InterPro domain enrichment analysis. All data is freely available for download via Expression Atlas and analysis results can also be obtained via the RNASeq-er API.

W509: Genome annotation resources at the EBI

Expert Curation of Protein Sequences in Uniprotkb/Swiss-Prot

Damien Lieberherr, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland and The UniProt Consortium
The UniProt KnowledgeBase (UniProtKB, https://www.uniprot.org) is a comprehensive, high-quality and freely accessible resource of protein sequences and functional information. UniProtKB combines two sections, UniProtKB/Swiss-Prot, the expert manually annotated protein sequences, and its automatically annotated complement UniProtKB/TrEMBL.
Expert curation in UniProtKB/Swiss-Prot combines the manually verified sequence with experimental evidence derived from biochemical and genetic analyses, 3D-structures, mutagenesis experiments, information about protein interactions and post-translational modifications. Besides harvesting, interpreting, standardizing and integrating data from literature and numerous resources, curators are also checking, and often correcting, gene model predictions.
UniProt has recently adopted Rhea (https://www.rhea-db.org/) as the reference vocabulary for enzyme annotation in UniProtKB, and now describes all enzymatic reactions using Rhea where possible. Rhea provides improved consistency and precision of enzyme annotation in UniProtKB and allows UniProt users to search, browse, and mine enzyme data in new ways, combining approaches from the fields of cheminformatics and bioinformatics.
Plants produce an enormous variety of natural products with extremely diverse molecular structures and activities. Linking these chemicals to their natural biosynthetic pathways in UniProtKB via Rhea will facilitate efforts to study their biology, and produce them and their derivatives at industrial scales.
Examples of curated plant pathways include anti-cancer drugs such as vincristine and vinblastine in *Catharanthus roseus*, an exotic plant found only in Madagascar, and lycopadiene, a tetraterpenoid biofuels of the microalga *Botryococcus braunii*.

W510: **Genome annotation resources at the EBI**
UniFire
Maria J Martin, European Bioinformatics Insitute, Hinxton, United Kingdom

W511: **Genome annotation resources at the EBI**
Using QuickGO to Access Gene Ontology Annotations
Sandra Orchard, EMBL-EBI, Hinxton, United Kingdom

W512: **Genome annotation resources at the EBI**
Adapting the ELIXIR Beacon Protocol to the Sharing of Plant Data
Gary Saunders, Elixir, Hinxton, United Kingdom
In this presentation I will describe the Beacon protocol and its planned extensions in order to serve the plants community. The Beacon protocol defines an open standard for genomics data discovery, developed by members of ELIXIR and the Global Alliance for Genomics & Health.
The initial version of the Beacon protocol had been developed to test the willingness and ability of international genome resources to share genomic data in a highly simplified context. The service was designed to accept specific queries in the form "Do you have any genomes with an 'A' at position 100735 on chromosome 3" and responds with "Yes" or "No."
The Beacon protocol has been designed to be:

- Simple: focus on robustness and easy implementation
- Federated: maintained by individual organizations and assembled into a network
- General-purpose: used to report on any variant collection
- Aggregative: provide a boolean (or quantitative) answer about the observation of a variant
- Privacy protecting: queries do not return information about single individuals

Recent and future versions of the Beacon protocol expand the original concept by providing a framework for querying other types of genome variation data (i.e. range queries and structural variants since v0.4) and also options for quantitative responses.
The primary Use Case for the Beacon protocol has been sensitive human data. However, ELIXIR had funded work to extend the Use Cases for the Beacon protocol into plant data. The Use Cases for human sensitive data and proprietary plant data overlap. A reference implementation of Beacon technology for plant data will be generated, extending it to integrate apricot variants data and thereafter extendable to plant data in general. Considered data standards for plant data will include Bioschemas.org and the BreedingAPI (www.brapi.org) alongside with their complementarity with GA4GH.

W513: **Genome annotation resources at the EBI**
Working with the Agricultural and Food Industries
Dominic Clark, EMBL-EBI, Cambridge, United Kingdom

W514: **Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment**
Phenomic Inference of Soybean Growth and Development
Katy Martin Rainey, Purdue University, West Lafayette, IN
There are compelling reasons to phenotype physiological traits on a large scale in soybean. First, soybeans produce few seeds, reducing opportunities for accurate phenotyping for selection and decision-making in the early generations of breeding populations. UAS imagery can overcome this limitation by providing indices for selection that are more heritable than current practices. Second, soybean is the most day length sensitive crop exhibiting remarkable developmental plasticity compared to other crops, and we need new tools to understand these phenomena. Third, soybean is the fastest growing crop globally in terms of cultivated acres and is expanding into subtropical, tropically and high-elevation growing areas where adaptation has been difficult to predict. Finally, soybean contributes significantly to overall human nutrition in terms of protein intake and global poultry production relies on soybean meal.
To addresses these opportunities, the Phenomic Inference methodology combines GWAS, quantitative genetics, mutli-environment yield trials and remote sensing into a comprehensive approach to assess applications of precision high-throughput phenotypes to field-based crop science. Ground reference and remote sensing data are collected on a panel of inbred lines unselected for yield but constrained for certain confounding traits, along with yield and phenology. Typically, a set of calibration plots is needed to validate the remote sensing prediction equations. Genetic architecture and quantitative properties, such as genetic correlation, are described as a function of development, which is an emerging area in crop science. Outputs include initial remote sensing capabilities along with genetic and phenomic information needed to tailor the applications.

W515: **Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment**
Aerial RGB and Infrared Image Analysis - Low Hanging Fruit for Phenomics
Larry C. Purcell and Avjinder Kaler, University of Arkansas, Fayetteville, AR
Crop phenotyping has not kept pace with advances in high throughput sequencing and genotyping, resulting in a phenotyping gap that has limited potential crop improvement. Unmanned aerial systems (UASs) provide a powerful means of collecting high-quality, aerial images that are informative with regards to crop performance and stress tolerance. While there is a wide range of different sensors that are being used in crop phenotyping, many unmanned aerial systems have a high-quality red-green-blue (RGB) camera as standard equipment that serves as a highly effective sensor. Phenotyping examples using the RGB camera from an UAS are based on distinguishing the green color of a crop from non-crop portions of the image (e.g., weeds or soil background). If there is a visible color distinction between the crop and the 'non-crop', it is highly likely that this can be quantified with analysis of RGB images.
Prior to the crop canopy completely covering the ground, the fraction of pixels in an image of the crop that are green relative to the total number of pixels in an image (including the soil background) is closely associated with the fraction of radiation intercepted by a crop. Canopy coverage

can be thought of as a measure of early-season photosynthetic capacity and crop competitiveness. This relatively simple measurement has been used both to identify soybean genotypes that rapidly establish a canopy and to map alleles associated with rapid canopy closure.
The intensity of greenness from RGB images can also be informative regarding differences among genotypes in plant nutrition. The intensity of greenness of individual plots can be quantified from aerial RGB images from the hue, saturation, and brightness values from each pixel and calculating the average Dark Green Color Index (DGCI) value from all pixels in a plot. The DGCI is a value ranging from 0 (light yellow) to 1 (dark green) that is closely associated with shoot nitrogen concentration. Aerial measures of DGCI are being used in genome wide association mapping (GWAM) in soybean and evaluating nitrogen fertilization needs in corn.
For evaluating genetic differences in crop drought tolerance, infrared (IR) cameras are able to detect small canopy temperature differences of about 0.05^{o}C. From the maximum allowed height for a UAS (122 m) and with a 25 mm focal length, one image is able to capture approximately 190 plots measuring 2x5 m. Under drought, genotypes that are continuing to transpire (cool canopies) are clearly distinguished from those with limited transpiration (warm canopies). Interestingly, when soil moisture is replete, there are also genotypic differences in canopy temperature. Aerial IR data is also being used for GWAM to identify genotypes and alleles associated with a cool canopy under drought in soybean.

W516: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment
Next-Generation Plant and Soil Water Sensors
Liang Dong, Iowa State University, Ames, IA
This presentation will introduce miniature water sensors for monitoring of plant and soil water status in a real-time manner. These sensors are innovative in materials, structures and manufacturing methods and have high sensitivity and large dynamic range. Coupling the measurements of the plant and soil sensors will enable continuous monitoring of water use efficiency of plants.

W517: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment
3D Mapping of Cotton Bolls in the Field
Changying Li, College of Engineering, University of Georgia, Athens, GA

W518: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment
ENVIRATRON: A Novel Controlled Environment, High-Throughput Phenotyping System at Iowa State University
Carolyn J. Lawrence-Dill, Iowa State University, Ames, IA
The ENVIRATRON project is a phenomics platform that enables researchers to monitor the performance of plants throughout their lifespan when subject to a variety of environmental conditions, including anticipated future environments. The ENVIRATRON permits researchers to incrementally alter critical variables to better simulate changing conditions that we face in the future. Most phenomics research has been two-dimensional -- looking at various attributes or traits in different taxa or genotypes under a single environmental condition. This project adds an extra dimension to the exploration of phenotype space: performance under different environmental conditions. The ENVIRATRON enables researchers to control a number of variables including temperature, day length, light intensity, humidity, CO2 levels, and water potential in the soil. This enables the system to simulate current climatic conditions in different areas of the world and future climatic scenarios. The ENVIRATRON consists of an array of plant growth chambers to create different environmental conditions. Unlike commercial plant phenomics systems, plants are not conveyed out of the growth chambers to monitor their growth performance, rather a rover with a robotically controlled arm periodically visits each chamber to image and analyze the plants. In addition to more standard RGB, florescence, near IR and IR imaging, sensors on the rover can be used to collect hyperspectral and holographic imaging and Raman spectroscopy data. The robot-assisted sensing approach enables precise location-specific data acquisition, resulting in improved sampling strategies and data quality.

W519: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment
Nanoantenna Optoelectronics for Plant Sensors
D. Keith Roper, Utah State University, Logan, UT
Sustainable use of land, water, and energy resources is imperative as evolving stresses and expanding needs put increasing demands on agricultural productivity. Selection of stress-resilient cultivars and real-time decision support systems to manage irrigation, fertilization, pests and pathogens could support sustainable agriculture. But cost, usability, and performance of current tools for precision agriculture and high throughput plant phenotyping (HTPP) limit their resolution and scalability. This prevents real-time, high-resolution, multi-scale, data-driven choices to select and manage better germplasm beyond a few plants or small plots. As a result, data from small-scale physiological studies and field-based yield assessments often do not match.
This work explores possibilities for nanoantenna optoelectronics for sensing local plant condition. Nanoantenna metasurfaces exhibit tunable effective refractive index due to coupling between near-field plasmon-active elements and far-field phase-coherent radiation. A plasmon is a quantum of plasma oscillation excited by incident electromagnetism at resonant wavelengths. Plasmon interactions are *limited to* resonance wavelengths and only occur in *close proximity* to plasmon-active nanostructures. Nanoantenna metasurfaces have tunable electromagnetic functionality due to two-dimensional (2D) structuring of suitable condensed-matter composites.
We examine nanoantenna metasurfaces using (1) multi-scale *models* to evaluate plasmon-photon coupling and design nanoantenna metasurfaces; and (2) novel methods to fabricate nanoantenna metasurfaces analyzed by optical and electron microscopy. We will present physicochemical, geometric, and electromagnetic effects on effective refractive index in nanoantenna metasurfaces. Interactions between dynamic polarizability and constructive interference suggest nanoantenna metasurfaces can provide *delocalization* of plasmon effects as well as tunability across the electromagnetic spectrum to permit far-field, broad-spectrum, plasmon enhancement. This provides innovative opportunities to characterize concentration of carbon, nitrogen and water species local to plant leaf surfaces and to capture dynamic changes in concentration as well as in physiological features of the plant.

W520: Genome to Phenome: Next Generation Sensors for Sensing Plants and Environment
The Power of Combing 3D with Multispectral Information for Phenotyping Applications

Grégoire Hummel, Phenospex, Heerlen, Netherlands
PlantEye is a high-resolution 3D laser scanner that computes a robust and validated set of morphological plant parameters fully automatically. A core feature of PlantEye is that it can be operated in full sunlight without any restrictions - crucial for plant phenotyping under field conditions or if you follow a "sensor-to-plant-concept". Phenospex has now developed a new sensor generation, which combines the actual features of PlantEye on the fly with a 4-channel multispectral camera in the range between 400 – 900nm. This unique hardware-based sensor fusion concept allows us to deliver spectral information for each data point of the plant in X, Y, Z-direction and we can compute parameters like NDVI, color index and many other vegetation indices. This new sensor generation opens a wide range of new possibilities in plant phenotyping and increases its efficiency.

W5**21: Genome Variation and Somatic Cell Breeding**
Introduction of the Genome Variation and Somatic Cell Breeding Workshop
Xiu-Qing Li, Agriculture and Agri-Food Canada, Fredericton, NB, Canada

W5**22: Genome Variation and Somatic Cell Breeding**
Inheritance of Stressed Induced Genetic Variations in Plants
Igor Kovalchuk, Department of Biology, University of Lethbridge, Lethbridge, AB, Canada

W5**23: Genome Variation and Somatic Cell Breeding**
Nucleosomes and DNA Methylation Shape Meiotic DSB Frequency in Plants
Ian R Henderson, University of Cambridge, Cambridge, United Kingdom
Meiotic recombination initiates from DNA double-strand breaks (DSBs) generated by SPO11 topoisomerase-like complexes. Meiotic DSB frequency varies extensively along eukaryotic chromosomes, with hotspots controlled by chromatin and DNA sequence. To map meiotic DSBs throughout a plant genome, we purified and sequenced *Arabidopsis*SPO11-1-oligonucleotides. SPO11-1-oligos are elevated in gene promoters, terminators and introns, driven by AT-sequence richness, which excludes nucleosomes and allows SPO11-1 access. A positive relationship was observed between SPO11-1-oligos and crossovers genome-wide, although fine-scale correlations were weaker. This may reflect the influence of interhomolog polymorphism on crossover formation, downstream of DSB formation. Although H3K4me3 is enriched in proximity to SPO11-1-oligo hotspots at gene 5'ends, H3K4me3 levels do not correlate with DSBs. Repetitive transposons are thought to be recombination-silenced during meiosis, in order to prevent non-allelic interactions and genome instability. Unexpectedly, we found high SPO11-1-oligo levels in nucleosome-depleted Helitron/Pogo/Tc1/Mariner DNA transposons, whereas retrotransposons were coldspots. High SPO11-1-oligo transposons are enriched within gene regulatory regions and in proximity to immunity genes, suggesting a role as recombination-enhancers. As transposon mobility in plant genomes is restricted by DNA methylation, we used the *met1*DNA methyltransferase mutant to investigate the role of heterochromatin in SPO11-1-oligo distributions. Epigenetic activation of meiotic DSBs in proximity to centromeres and transposons occurred in *met1*mutants, coincident with reduced nucleosome occupancy, gain of transcription and H3K4me3. Together, our work reveals a complex relationship between chromatin and meiotic DSBs within *Arabidopsis*genes and transposons, with significance for the diversity and evolution of plant genomes.

W5**24: Genome Variation and Somatic Cell Breeding**
Regulation of Transposon and Genome Instability of Plants
Xiuren Zhang, Texas A&M University, College Station, TX

W5**25: Genome Variation and Somatic Cell Breeding**
Double-Stranded Break Repair-Mediated Plastid DNA Insertion into Nucleus
Dong Wang, Nanchang University, Nanchang, China
The mitochondria and plastids of eukaryotic cells evolved from endosymbiotic prokaryotes. DNA from the endosymbionts has bombarded nuclei since the ancestral prokaryotes were engulfed by a precursor of the nucleated eukaryotic host. An experimental confirmation regarding the molecular mechanisms responsible for organelle DNA incorporation into nuclei has not been performed until the present analysis. Here we introduced double-strand DNA breaks into the nuclear genome of tobacco through inducible expression of I-SceI, and showed experimentally that tobacco chloroplast DNAs insert into nuclear genomes through double-strand DNA break repair. Microhomology-mediated linking of disparate segments of chloroplast DNA occurs frequently during healing of induced nuclear DSBs but the resulting nuclear integrants are often immediately unstable. Non-Mendelian inheritance of a selectable marker (*neo*), used to identify plastid DNA transfer, was observed in the progeny of about 50% of lines emerging from the screen. The instability of these *de novo* nuclear insertions of plastid DNA (*nupt*s) was shown to be associated with deletion not only of the *nupt* itself but also of flanking nuclear DNA within one generation of transfer. This deletion of pre-existing nuclear DNA suggests that the genetic impact of organellar DNA transfer to the nucleus is potentially far greater than previously thought.

W5**26: Genome Variation and Somatic Cell Breeding**
Landscape of Repetitive Elements in Somatically Excluded Chromosomes of the Sea Lamprey (*Petromyzon marinus*)
Vladimir Timoshevskiy, Nataliya Timoshevskaya and Jeramiah J. Smith, University of Kentucky, Lexington, KY
The sea lamprey is one of few vertebrate species that reproducibly eliminates large fractions of its genome during normal embryonic development. These elimination events result in the loss of ~20% of the lamprey's genome from all somatic cell lineages. This germline-specific DNA is lost in the form of large fragments, including entire chromosomes. However, reconstruction of eliminated regions has proven challenging due to the complexity of the lamprey karyotype and the exceedingly high repeat content of the genome.
We applied integrative approach aimed at further characterization of the large-scale structure of eliminated segments, including:
1) developing DNA-probes that selectively labels eliminated chromosomes by laser capture microdissection;

2) *in silico* identification of germline-enriched repeats;
3) determining the chromosomal location of specific repetitive sequences in germline metaphases using multicolor FISH;
4) verification of specificity to eliminated chromosomes by hybridization with lagging anaphases in whole embryos and somatic line chromosome spreads.
Our integrative approach allowed us identify 6 repetitive elements that are found exclusively on the eliminated chromosomes and resulted in the identification of 12 chromosomes that programmatically eliminated during early embryogenesis. Based on hybridization patterns of these repeats on meiotic spreads we developed cytogenetic map of chromosomes that are excluded from the somatic genome of the sea lamprey. The fidelity of germline-specific repetitive elements and their distinctive patterning in elimination anaphases is taken as evidence that these sequences might contribute to the specific targeting of chromosomes for elimination and in molecular interactions that mediate their decelerated poleward movement in chromosome elimination anaphases.

W527: Genomic features and chromosome functionality
The Genome of *Arachis hypogaea* Provide Insight into the Evolution and Domestication

Weijian Zhuang[1], Hua Chen[1], Yang Meng[2], Jianping Wang[3], Manish K Pandey[4], Chong Zhang[1], Qiang Yang[1], Tiecheng Cai[1], Ye Deng[1], Andrew H. Paterson[5], Xiyin Wang[6], Ray Ming[7], Rajeev K Varshney[4] and PGR, (1)Fujian Agriculture and Forestry University, Fuzhou, China, (2)Grandomics Biosciences, Beijing, China, (3)Agronomy Department, University of Florida, Gainesville, FL, (4)ICRISAT, Hyderabad, India, (5)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, (6)NCST and UGA, Athens, GA, (7)Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, China

W528: Genomic features and chromosome functionality
Involvement of Chloroplast Genome Copy Number and Codon Usage Preference in Crop Performance

Xiu-Qing Li, Agriculture and Agri-Food Canada, Fredericton, NB, Canada

W529: Genomic features and chromosome functionality
Walking on the Wild Side using Pangenomics for Accelerated Crop Improvement in Chickpea

Aamir W. Khan[1,2], Vanika Garg[1], Manish Roorkiwal[1], Philipp E. Bayer[2], Mahendar Thudi[1], Annapurna Chitikineni[1], Hari D Upadhyaya[1], Henry T. Nguyen[3], David Edwards[2] and Rajeev K Varshney[1], (1)ICRISAT, Hyderabad, India, (2)University of Western Australia, Perth, Australia, (3)University of Missouri, Columbia, MO

The wild accessions of chickpea have huge genetic diversity which can be exploited for increasing genetic gains in the crop. Several large scale re-sequencing efforts have been carried out to identify variants in chickpea, but the complete genetic repertoire has not been captured as these efforts involve mapping of reads on single chickpea reference genome. We report the construction of chickpea pangenome by sequencing and de novo assembly of accessions from eight different annual chickpea wild species. These accessions were sequenced at ~180X coverage using Illumina platform, generating a total of 1.14 Tbp data. The de novo assemblies for these species resulted in size varying from 512.3 Mbp to 927.0 Mbp with high N50 values. Comparisons of these assemblies identified genes exhibiting copy number variations and presence absence variations, some of which show evidence of positive selection and might have association with important agronomic traits. The chickpea pangenome will serve as a valuable genomics resource and will have broad implications in chickpea breeding.

W530: Genomic features and chromosome functionality
Rice Information GataWay (RIGW) and CRISPR-P Genome Editing Tools in Plants

Jia-Ming Song[1], Hao Liu[1], Yang Lei[1], Kabin Xie[1], Jianwei Zhang[2] and **Ling-Ling Chen**[3], (1)Huazhong Agricultural University, Wuhan, China, (2)Arizona Genomics Institute, University of Arizona, Tucson, AZ, (3)Huazhong Agriculture University, Wuhan, China

We established a comprehensive bioinformatics platform, Rice Information GataWay (RIGW), to provide GBrowse-based view of two *Oryza sativa* '*xian*' rice reference genomes Zhenshan 97 (ZS97) and Minghui 63 (MH63) and other omics data. RIGW offered homologs among '*xian*' and '*geng*' rice, and other plant species. User-friendly web interfaces were also provided to show the predicted PPIs in rice, the metabolic pathways in ZS97/MH63, CRISPR-Cas single guide RNA design tool and GO enrichment in RIGW. All the genomic sequences and annotation could be freely accessed, and useful links to other public databases were offered. RIGW is freely available at http://rice.hzau.edu.cn/.
CRISPR-P provides web services for computer-aided design of sgRNA with minimal off-target potentials, which supports sgRNA design for most of the sequenced plant genomes. CRISPR-P supports to design guide sequences for various CRISPR-Cas systems including Cpf1 and various Cas9 endonucleases. A comprehensive analysis of the guide sequence is provided, including GC content, restriction endonuclease site, microhomology sequence flanking the targeting site (microhomology score), and the secondary structure of sgRNA. Identification of sgRNA from custom sequences is also provided. CRISPR-P is freely available at http://cbi.hzau.edu.cn/CRISPR2/.

W531: Genomic features and chromosome functionality
Reanalyzing Selaginella Genomes Found Recursive Polyploidization

Xiyin Wang, North China University of Science and Technology, Watkinsville, GA and Jinpeng Wang, North China University of Science and Technology, Tangshan, China

Lycophytes and seed plants constitute the typical vascular plants. Lycophytes were thought to have no paleo-polyploidization although the event is known to be critical for the fast expansion of seed plants. Here, genomic analyses including homologous-gene dotplotting detected multiple paleo-polyploidization events, with one occurred approx.13-15 million years ago (Mya), another about 125-142 Mya, during the evolution of the genome of *Selaginella moellendroffii*, a model lycophytes. In addition, comparative analysis of reconstructed ancestral genomes of lycophytes and angiosperms suggested that lycophytes were affected by more paleo-polyploidization than seed plants. Results from the present genomic

analysis indicate that paleo-polyploidization has contributed to the successful establishment of both lineages—lycophytes and seed plants—of vascular plants.

W532: **Genomics-Assisted Breeding**
Welcome & Introduction
Rajeev K Varshney, ICRISAT, Hyderabad, India

W533: **Genomics-Assisted Breeding**
Dissection of Heterotic Genes Informs a New Rice-Improvement Strategy
Bin Han, National Center for Gene Research, Shanghai Institute of Plant Physiology and Ecology, CAS-Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China

W534: **Genomics-Assisted Breeding**
Inclusive Evaluation of the Tetraploid Wheat Germplasm based on the Svevo Durum Wheat Genome Sequence Assembly
Marco Maccaferri[1], Raj K Pasam[2], Danara Ormanbekova[3], Benjamin Kilian[4], Sara G. Milner[5], Elisabetta Mazzucotelli[6], Anna-Maria Mastrangelo[7], Elisabetta Frascaroli[8], Hakan Ozkan[9], Steven S. Xu[10], Shiaoman Chao[10], Assaf Distelfeld[11], Curtis Pozniak[12], Matthew Hayden[13], Nicola Pecchioni[14], Roberto Tuberosa[15], Reem Joukhadar[16], Luigi Cattivelli[17] and The Durum Wheat Genome Sequencing Consortium, (1)University of Bologna, DISTAL, Bologna, Italy, (2)Department of Economic Development, Jobs, Transport and Recourses, AgriBio Centre for AgriBioscience, Bundoora, VIC, 3083, Australia., Bundoora, Australia, (3)DipSA, University of Bologna, Bologna, Italy, (4)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (5)IPK Gatersleben, Stadt Seeland, Germany, (6)CREA-GB,, Fiorenzuola d' Arda, Italy, (7)CREA-MAC, SS42, 24126 Bergamo, Foggia, Italy, (8)Department of Agricultural Sciences, University of Bologna, Bologna, Italy, (9)University of Cukurova, Adana, Turkey, (10)USDA-ARS, Fargo, ND, (11)School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel, (12)University of Saskatchewan, Saskatoon, SK, Canada, (13)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia, (14)CREA - Research Centre for Cereal and Industrial Crops, Foggia, Italy, (15)DISTAL, University of Bologna, Bologna, Italy, (16)DEDJTR, AgriBio, Bundoora, Australia, (17)CREA-GB, Research Centre for Genomics and Bioinformatics,, Fiorenzuola d' Arda, Italy
Triticum turgidum genetic resources are a wide reservoir of diversity, valuable for both durum wheat and common wheat pre-breeding, as highlighted by the success of synthetic wheats. However, a comprehensive analysis of these resources is lacking.
As part of the Svevo durum wheat genome sequencing, we assembled a Global Tetraploid wheat Collection (GTC) of 1,856 accessions from 11 tetraploid wheat taxa, including wild and domesticated emmer (WEW and DEW) durum landraces (DWL) and modern durum wheat cultivars (DWC). We used the iSelect 90K SNP array anchored to the Svevo genome to investigate population structure and selection/demography signatures.
We traced: *i*) the inheritance of haplotype blocks from either the North-eastern or Southern Levantine Fertile Crescent (Turkey) WEW populations to whole DEW germplasm, with a 0.65/0.35 overall inheritance ration, *ii*) the two subsequent independent but similar star-like dispersal patterns associated to DEW and DWL evolution, with six main populations each.
Durum wheat most probably originated from the Southern-levant DEW. Ethiopian emmer and durum, *T. turanicum* and *T. carthlicum* were the most differentiated with minimal contribution to modern durum. Modern durum originated mostly from the North-African and Transcaucasian DWL populations. WEW, with the highest genome-wide diversity, provided the reference for assessing the diversity reductions associated to domestication and breeding. Numerous strong diversity depletions signals were observed primarily for the WEW-to-DEW transition. These signals progressively consolidated through domestication and breeding. Specific DEW-to-DWL signals were also observed. Diversity reduction index (*DRI*), *Fst*, haplotype-based *XP-EHH* and *hapFLK* and spatial pattern of site frequency spectrum (*XP-CLR*) metrics were considered. Frequently two or more indexes occurred in overlapping regions (selection clusters). In total, 104 pericentromeric (average size of 107.7 Mb) and 350 non-pericentromeric (average size of 11.4 Mb) clusters were identified. WEW-to-DEW and DEW-to-DWL transitions mostly involved extended pericentromeric regions tagged by *DRI* and *Fst*, while the DEW-to-DWL and DWL-to-DWC transitions were mostly associated to numerous XP-*EHH* and *XP-CLR* signals. We carried out a comprehensive projection of the published tetraploid QTLome and investigated the genome-wide QTL distribution and co-location between meta-QTLs and demography/selection signals. The usefulness of the GTC to elucidate the evolutionary patterns associated to loci, haplotype blocks and causative sequence variants for traits of breeding interest like cadmium-cumulation in grains, grain size and disease resistance loci are reported.

W535: **Genomics-Assisted Breeding**
Towards Genomics-Assisted Breeding in Walnut (*Juglans regia* L.)
Annarita Marrano[1], Charles A. Leslie[1], Gina M. Sideli[1], Hao Cheng[2] and David B. Neale[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Department of Animal Science, University of California, Davis, Davis, CA

W536: **Genomics-Assisted Breeding**
Genomics of Sustainable Breeding Progress: A Mega-Study in Wheat
Rod Snowdon, Department of Plant Breeding, Justus Liebig University, Giessen, Germany
Agricultural production must increase dramatically to sustain the growing world population, and plant breeding has been the major driver of sustained increases in crop productivity during the past 50 years. However, advanced breeding endeavours in most major crops during this period have focussed primarily on selection for high grain yield in high-input cropping systems. This raises concerns that cultivars bred for highly intensive agricultural systems may lack the adaptive capacity to cope with emerging climatic or sustainability challenges. Because high-input agriculture is frequently associated with ecological sustainability penalties, a common assumption is that high-performing modern crop varieties lack the broad genetic base required for adaptation to sub-optimal environments or reduced-input cropping systems. To address these hypotheses,

we analysed a large panel of elite European winter wheat cultivars representing 50 years of breeding progress in one of the most intensive, highest-yielding cereal production systems on the planet. We investigated long-term consequences of breeding progress in 192 European winter wheat cultivars whose release dates spanned the last 5 decades, including the most important and successful cultivars grown in western Europe during their respective period of release. The panel was grown over two years in a total of 36 environments, spanning optimal and marginal soil, water and temperature conditions and including variants comparing contrasting applications of nitrogen fertiliser and chemical plant protection. Analysis of the results in the context of long-term breeding progress provides unique insight into the influence of breeding on crop productivity and sustainability. In particular, our results emphatically contradict the popular hypothesis that intensive breeding for high performance reduces diversity in modern cultivars and thus reduces genetic potential for long-term genetic improvement. Furthermore, an accompanying genome-wide analysis of genetic diversity parameters and selection patterns in association with extensive trait data provided comprehensive insight into how intense artificial selection and breeding impacts diversity and future selection potential in the world's most important cereal crop. We developed haplotype-based methods to analyse genome-wide trait variance, providing a unique basis to identify, implement and recombine useful genetic diversity within existing elite breeding pools and promote sustainable yield progress in future.

W5**37: Genomics-Assisted Breeding**

Speed Breeding to Supercharge Our Future Crops

Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

W5**38: Genomics-Assisted Breeding**

Integration of Genetics and Genomics-Based Breeding Approaches in the CIMMYT Global Wheat Program

Susanne Dreisigacker, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

The Global Wheat Breeding Program at CIMMYT was founded to explore strategies for breeding widely adapted and highly stable wheat cultivars. Annually the program distributes nurseries and yield trials through an international collaborative network that includes more than 300 cooperators. The international nurseries and trails serve as a vehicle for the dissemination of improved wheat germplasm. The materials is used in national wheat breeding programs for subsequent crosses or tested for direct release to smallholder farmers throughout the developing world with major impact on maintaining food security.

It is well estimated that wheat yields need to grow at an annual rate of 2 to 3% in farmers' fields to meet the expected demand in 2050. To increase genetic gains and maintain a stable food supply plant breeders and geneticist need to utilize contemporary methods to enhance current breeding strategies.

Genetics and genomics-based breeding approaches have been gradually integrated into the Global Wheat Breeding Program. Informative markers are routinely used to improve a variety of traits and for the development of new products. In addition, genomic selection regarded as a useful tool to accelerate genetic gain especially for more complex traits is implemented. Achievements and prospects will be presented.

W5**39: Genomic Selection and Genome-Wide Association Studies**

US National Beef Cattle Single Step Genomic Evaluations

Matt L. Spangler, University of Nebraska - Lincoln, Lincoln, NE

Genomic selection in US beef cattle began in 2010 utilizing either a correlated trait approach or post evaluation blending. In either case, estimation of markers effects from a training population was required. For a variety of reasons, including the highly selected nature of the training populations, bias in resulting genetic predictions were often observed. In 2017, both the American Angus Association (AAA) and American Herford Association (AHA) adopted single-step methods to include genomic information into weekly genetic evaluations. International Genetic Solutions (IGS), a collaboration among 15 breed association from the US, Canada, and Australia, followed suit in early 2018. The AAA evaluation uses single-step GBLUP (ssGBLUP) via software from the University of Georgia. Both AHA and IGS employ a hybrid marker effects model (sHybrid) via the BOLT software available from Theta Solutions, LLC. Regardless of the model and software employed, the predictions of genetic merit (Expected Progeny Differences; EPD) have been shown to be more accurate and less biased than EPD resulting from two-step methods. A more tangible measure of increased accuracy is the Expected Progeny Number (EPN), or the number of offspring an individual would need to have produced to reach the same level of accuracy as simply being genotyped as a non-parent. Across breeds, the EPN for growth traits (birth, weaning, and yearling weights) ranges from 8-26, carcass traits (carcass weight, ribeye area, external fat, marbling) range from 5-12, and female fertility traits (heifer pregnancy, stayability) range from 14-25. As currently employed, a fundamental difference between these two approaches, ssGBLUP and sHybrid, is the assumption of equality among marker effects such that ssGBLUP as implemented by AAA assumes all markers (e.g., 50,000 SNP) contribute equally via a modified kinship matrix (H) and the latter assumes only a pre-identified subset of markers, with differential effects, contribute to the trait of interest. At the time of this abstract, the subset fitted by IGS and AHA includes approximately 2,200 markers. However, equivalency between these two models (ssGBLUP and sHyrbid) is possible depending on the assumptions made relative to marker effects. In example, if all markers were fitted under the assumptions of a BayesC model where the mixing proportion (pi) equaled 0, then sHyrbid would be equivalent to ssGBLUP. Moreover, fitting a weighted genomic relationship matrix (G) to form the augmented kinship matrix H in ssGBLUP is possible. Future efforts, in both cases, will likely need to focus on the utilization of causal variants as they are made available and strategies that optimize cost of data collection (genotypes and phenotypes) and resulting gains in prediction accuracy.

W5**40: Genomic Selection and Genome-Wide Association Studies**

Exploration of the Impact of Genetic Architecture on the Performance of GWAS Models Quantifying Epistasis and Genomic Selection Models that include Peak-Associated Markers as Fixed-Effect Covariates

Alexander E. Lipka, University of Illinois, Urbana, IL

Statistical approaches for genome-wide association studies (GWASs) and genomic selection (GS) have facilitated the identification of genomic loci associated with agronomically important traits while controlling for false positives and the use of genome-wide marker data to accurately

predict trait values. Nevertheless, statistical models that reflect the multifaceted contributions of loci throughout the genome have a potential to facilitate unprecedented quantification of the genetic architecture underlying various traits and increase prediction accuracies. Here, the performance of two statistical models are evaluated for traits with contrasting genetic architectures simulated using marker data in maize, sorghum, and humans. The results suggest that the first model, designed to quantify the simultaneous contribution of additive and two-way epistatic loci, is capable of identifying and distinguishing between simulated additive and epistatic quantitative trait nucleotides (QTNs). These simulation studies also reveal that the second model, which includes peak-associated markers from a GWAS of a training set as fixed-effect covariates in an RR-BLUP GS model, typically decreases prediction accuracy, increases the variability of prediction accuracy across replicate traits, and is capable of increasing the bias of predictions compared to a standard RR-BLUP GS model.

W541: Genomic Selection and Genome-Wide Association Studies
PanGWAS: GWAS of the Pan-Genome Provides New Insights into *P. trichocarpa* Phenotypic Variation
David Kainer, Oak Ridge National Laboratory, Oak Ridge, TN

Most GWAS are performed using SNPs obtained from reads aligned to a reference genome of the same or related species. However, it is becoming clear that a reference genome assembled from one individual may not adequately represent the genomic feature space of the species. A large number of genes may be present or absent in any individual. The overall species-wide gene space can be represented as the Pan-genome of the species, containing core genes found in most or all individuals, and distributed genes that are found in fewer individuals and may be rare. We have assembled the pan-genome of *Populus trichocarpa* using DNA and RNAseq data from almost 1000 trees, resulting in a new marker dataset representing the presence or absence variation (PAV) of putative genes across the population. Here I present the results of association between these PAV markers and phenotypes assayed in the same population, such as leaf metabolites and wood chemistry.

W542: Genomic Selection and Genome-Wide Association Studies
MVP and Hiblup for Efficient Genome-Wide Association Study and Genomic Prediction/Selection
Lilin Yin[1], Haohao Zhang[2], Xinyun Li[1], Xiaohui Yuan[2], Shuhong Zhao[1] and **Xiaolei Liu**[1], (1)Key Laboratory of Agricultural Animal Genetics, Breeding, and Reproduction, Ministry of Education, Huazhong Agricultural University, Wuhan, China, (2)School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China

Genome-Wide Association Study (GWAS) and Genomic Prediction/Selection (GP/GS) have been widely used for detecting candidate genes that affect agricultural economic traits and predicting genetic value of each individual, as well as risk of human diseases. With the rapid decreasing of genotyping cost, big data has been used in GWAS and GP/GS. In order to meet the computation challenges, here we present **MVP** and **HIBLUP** for efficient GWAS and GP/GS, respectively.

MVP can accept genotype data in Plink binary, VCF, and Hapmap formats, and it implements efficient algorithms for population structure evaluation, parallel-accelerated association tests by general linear model, mixed linear model, and FarmCPU model, and provides high-quality customer design figures of GWAS related information. The MVP source code, user manual, and example datasets are freely available at https://github.com/XiaoleiLiuBio/MVP.

A series of statistical models under BLUP (Best Linear Unbiased Prediction) framework were developed based on the information provided by different species with different reproduction modes, such as Genomic BLUP and Single Step BLUP. Here we introduce **HIBLUP**, an R package that provides estimated genetic value of each individual by maximizing the usage of information from pedigree records, genome, and phenotypic observations, as well as all process-related functions, such as construction of relationship matrix and estimation of variance components. Our experience indicates that a combination of **H**E Regression algorithm and Average **I**nformation (**HI**) algorithm provides efficient and robust variance component estimation. We also discussed the computational complexity of the BLUP procedures and the optimized the design for efficient handling of practical human, plant, and animal data, respectively, and we believe that HIBLUP will facilitate the research for human geneticist and as well as plant animal breeders.

W543: Genomic Selection and Genome-Wide Association Studies
Establishing a Unified Framework for Genomic Selection and Genome-Wide Association Study of Complex Traits
Jianming Yu, Xianran Li and Tingting Guo, Iowa State University, Ames, IA

Integrated analysis of genotype by environment can reveal the pattern and mechanistic interplay underlying the observed phenotype dynamics. A critical question needs to be answered to enhance our ability to conduct genomic and environmental analysis of varied phenotypic plasticity observed in natural field conditions: How to uncover patterns at different levels to facilitate complex trait dissection and performance prediction. In this talk, I will describe multiple patterns underlying the observed phenotype, and introduce the integrated analytical framework CERIS-JGRA, or Critical Environmental Regressor through Informed Search - Joint Genomic Regression Analysis. Findings from E-enabled GWAS and GS in multiple plant species will be shown.

I will relate this research to the historic debate between the biometrical concept of G x E_b from R. A. Fisher and the developmental concept of G x E_d from Lancelot Hogben, and the "synthetic" work by Clausen-Keck-Hiesey, and the newly proposed "Omnigenic Model" for complex traits.

W544: Genomics of Crop Ecosystem Services
Improving the Longevity of Perennial Wheat for Better Ecosystem Services
Shuwen Wang, The Land Institute, Salina, KS

Perennial wheat, a hybrid species between annual wheat (*Triticum ssp*.) and perennial wheatgrass (*Thinopyrum ssp*.), has the potential of increasing food security and ecosystem services. However, nearly all the perennial wheat cultivars/breeding lines developed in the past decades are short-lived which prevents perennial wheat from being a viable crop. We observed that perennial wheat required vernalization in the 1st year but lost the requirement starting in the 2nd year. The plants kept flowering, which reduced the rate of summer and winter survival. We are attempting new approaches to increasing the level of longevity. Through genome editing, we have obtained edited plants for a vernalization gene which enhanced longevity-related traits. Many secondary tillers produced after the primary tillers had ripened; the plants stayed green for longer time; and a few plants had become rhizomatous. Through hybridization of winter durum wheat with intermediate wheatgrass, we have obtained

progeny plants that could live in the field for 2 years with most plants survived. The plants responded to vernalization in a manner more like wheatgrass. Using 34,000 chromosome-specific markers we are investigating a breeding population to track the inheritance of wheat and wheatgrass chromosomes and to identify which chromosomes are important to the regulation of longevity. The strategy for the breeding of long-lived perennial wheat will be discussed.

W545: Genomics of Crop Ecosystem Services
Genomics of Crop Microbiomes
Jason G Wallace, University of Georgia, Athens, GA

The microbiome of an agricultural field is an invisible but important component of healthy ecosystem functioning. Crop microbiomes can affect carbon sequestration, nutrient acquisition, disease resistance, and toxin removal. Despite these known benefits, we know very little about how crop microbiomes are formed, how they affect plants (and vice-versa), and what we can do to influence them in specific ways. Genomic research is currently identifying specific mechanisms for how crops and microbes form these communities and how they interact, including community assembly and host-microbe specificity, although a great deal of work remains to be done. Harnessing these microbial populations will be a major challenge for 21st century agriculture, but doing so holds great potential for improving both the efficiency and sustainability of global agriculture.

W546: Genomics of Crop Ecosystem Services
TBA
Yaniv Brandvain, University of Minnesota, St Paul, MN

W547: Genomics of Crop Ecosystem Services
Toward Ratooning/Perennial Sorghums
Andrew H. Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

W548: Genomics of Genebanks
Quantifying Genetic Diversity in Cultivated Beet (*Beta vulgaris*) using a Pooled Population Sequencing Strategy
Paul Galewski, Michigan State University, East Lansing, MI and Mitch McGrath, USDA-ARS Sugarbeet & Bean Research Unit, East Lansing, MI

Beta vulgaris L. (beet) is a species complex composed of several distinct crop types, including sugar beet, fodder beet, table beet and leaf beet/chard. Development of crop type lineages appears to have resulted from interactions between selection, drift, geneflow, recombination and the sorting of ancestral polymorphism. Beets are generally heterozygous and contain self-incompatibility mechanisms. Therefore, reproducing and maintaining the genetic constitution of a single individual for genetic and phenotypic analysis is a challenge. Beet populations are the fundamental unit of improvement and contain the evolutionary and adaptive potential of the species. This research summarizes several experiments which explored pooled population genomic data to survey the organization and distribution of genetic diversity within cultivated *B. vulgaris*. Using hierarchical approaches to examine diversity in individual populations, crop type lineages and the species as a whole, we observed clear patterns in genome-wide variation reflecting the relative importance of admixture, selection and reproductive isolation in the development of *B. vulgaris*crop type lineages. Genome differentiation along chromosomes was plotted using 2pq, Fst, and lineage specific variation for each crop type. Regions identified as differentiated were further investigated for signals of selection and drift, and their underlying genomic features (e.g. genes, CDS, promoters) were prioritized by their predicted role in controlling phenotypic variation (e.g. root enlargement, biomass production, patterning of root tissues and cell types, and the accumulation of sucrose). Admixture was evaluated along chromosomes using correlations in allele frequency and quantity of shared variation between lineages. Interestingly, this revealed a single origin for genes controlling important agronomic characters. Integrating selection, drift, and admixture into a putative demographic history of beet provides evidence for how specific genes have influenced the development of beet crop types.

W549: Genomics of Genebanks
Characterization of a Sunflower Collection using Genomic Prediction
Alexandra Duhnen, Marie-Claude Boniface, Nicolas Pouilly and Brigitte Mangin, Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France

The french National Institute of Agricultural Research (INRA) manages a Biological Resource Center (BRC) for sunflower. Projects are conducted in partnership with private companies to maintain, characterize and develop sunflower genetic resources. Our project aims at obtaining genomic prediction of traits of agricultural interest for a large set of lines from the INRA sunflower collection.

Our approach exploits the available phenotypic data for a core collection to obtain predicted values for unobserved lines, owing to genotyping of both sets of lines for a common set of markers (about 300 lines from core collection and about 800 unobserved lines were genotyped with an AXIOM 51k array). This idea was already successfully tested on sorghum genetic resources and presented in Yu, X. et al. (2016).

We are particularly interested in obtaining good quality predictions for traits of tolerance to abiotic stresses because their observation is particularly long and costly. In fact, observation for these traits requires observation of genotypes in multiple environments and determination of stress level for each environment (Mangin, B. et al., 2017).

Using resampling methods to estimate prediction accuracy, we obtained a high prediction accuracy for oil content (average correlation of about 0.72 between predicted and observed values) and a moderate prediction accuracy for yield (about 0.37). Our current estimators of prediction accuracy are generally lower for traits of tolerance to abiotic stresses.

References:

Mangin, B., Casadebaig, P., Cadic, E., Blanchet, N., Boniface, M. C., Carrère, S., ... & Langlade, N. B. (2017). Genetic control of plasticity of oil yield for combined abiotic stresses using a joint approach of crop modelling and genome‐wide association. Plant, Cell & Environment, 40(10), 2276-2291.

Yu, X., Li, X., Guo, T., Zhu, C., Wu, Y., Mitchell, S. E., ... & Yu, J. (2016). Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nature Plants, 2:16150.

W550: Genomics of Genebanks

Redefining Ways of Efficient Management and Proper Utilization of Genetic Resources from Genebank Collection in NGS Era

Sandip Mallikarjun Kale[1], Albert Wilhelm Schulthess[1], Axel Himmelbach[1], Martin Mascher[1], Jochen Christoph Reif[1] and Nils Stein[2], (1)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Stadt Seeland, Germany

Germplasm, both as a raw or as a source of variation is a primary requirement of crop improvement programs. Initially, the crop genebanks were established with the aim of conservation of genetic variation present within the genetic resources. The emphasis was then shifted towards proper management and utilization of germplasm materials and so information about germplasm became equally important as the germplasm itself. Till date, “passport” data which store information about origin of a germplasm, its taxonomic status and some easy to major traits, is the only information available about majority of germplasm. Although such dataset can be a useful resource, its reliability is often a major concern to the researchers. Further, the genetic resources activities are being coordinated on global scale. Therefore, there is need to redefine the ways of management of genetic resources for their smooth exchange and proper utilization in global crop improvement programs. The genetic profiling of each accession with environmentally neutral markers could be the best complementary option. Further, in the current NGS era, it is now possible to generate sequence information sufficient not only for fingerprinting, but also for identification of duplicates, assessing diversity and for performing genome-wide association studies. With this goal, we have genotyped 6000 out of 9700 winter wheat accessions from IPK Gatersleben genebank using Genotyping-by-sequencing technique and showed its usability for assessing genetic diversity and also for genome-wide association study for agronomically important traits in European wheat collection. This information would be a valuable resource for wheat breeding research.

W551: Genomics of Genebanks

Domestication as Starting-Point, End-Point, or Transition?

Paul Gepts[1], Andrea Ariani[2] and Jorge C. Berny Mier y Teran[1], (1)University of California, Davis, CA, (2)Bayer CropScience N.V., Gent, Belgium

Domestication is usually considered as the starting-point of agriculture. Selection acting on a crop's wild ancestor led to phenotypic changes, called the domestication syndrome, accompanied by changes in gene frequencies. In this presentation, we will argue that we need to revise this vision to treat domestication as a transition or end-point rather than a starting-point, based primarily on our data from common bean. The wild progenitor of common bean *(Phaseolus vulgaris*) has an extraordinarily broad distribution, from northern Mexico to northwestern Argentina, with some significant gaps. Recent research by colleagues and our group have clarified the process and timing of range expansion events that led to this extended distribution from the origin of the species in Mesoamerica. Additional studies have also examined the changes in the prevailing climates to which this extended distribution was exposed and the effects of this expansion on the two domesticated gene pools (Andean vs. Mesoamerican) of common bean and their current distribution throughout the world. Thus, events that took place well before domestication are equally, if not more important, in determining characteristics of domesticates, including levels of genetic diversity, genetic relationships, and adaptation. Data from other plants, whether crops or wild taxa, also show the importance of human interactions with wild plants prior to domestication in affecting the domestication process.

W552: Genomics of Non-Classical Model Animals

When the Brain Goes Diving: A Genome and Transcriptome Approach to the Hypoxia Tolerance of Brains of Whales and Seals

Thorsten Burmester, Universität Hamburg, Zoologisches Institut, Hamburg, Germany

The sufficient supply of oxygen is essential for life. Diving mammals such as seals and whales tolerate repeated and prolonged deficiency of oxygen. Their brains survive low oxygen levels that would be fatal to most other mammals, including humans. In vitro, neurons of the hooded seal (*Cystophora cristata*) endure several hours without oxygen, as well as low glucose and high lactate levels. Diving mammals offer the unique opportunity to study the molecular adaptations of the mammalian brain to hypoxia and reoxygenation.

We apply comparative genomics and transcriptomics to identify genes and proteins involved in the dive-adaptation of the brain and elucidate their specific functions. For example, RNA-seq analyses using transcriptomes from the visual cortices of selected whale and seal species showed that higher expression of genes involved in metabolic processes and stress, while there was a lower expression of genes involved in neuronal signaling and protein synthesis. Diving mammals had significantly higher mRNA levels of the stress-genes clusterin and 100B. Experiments using stably transfected cell lines confirmed the ability of these genes to confer tolerance against hypoxia and reactive oxygen species (ROS). The results provide novel insights into how the brain survives hypoxia and hypoxia/reoxygenation, and will improve the understanding of energy metabolism in the mammalian brain.

W553: Genomics of Non-Classical Model Animals

Returning to the Sea and Beating Cancer: Comparative Genomics of Whales and the Evolution of Cancer Suppression

Marc Tollis, School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, Jooke Robbins, Center for Coastal Studies, Provincetown, MA, Per J. Palsbøll, University of Groningen, Groningen, Netherlands and Carlo C. Maley, Arizona State University, Tempe, AZ

Cetaceans include all whales, dolphins, and porpoises, and are highly specialized aquatic mammals that include the largest animals that have ever lived. With ~1000X more cells than a human, gigantic and long lived whales are theoretically susceptible to a higher lifetime risk of cancer. However, the largest animals do not get more cancer than humans – an observation known as Peto’s Paradox. To investigate this paradox, and uncover the genomic bases of many cetacean-specific adaptations, we provide a genome assembly for the humpback whale (*Megaptera*

novaeangliae) and leverage existing genomic resources in a study that is novel in scope for cetaceans. We address questions about the evolutionary timescale of the diversification of modern whales – particularly during the Miocene and earlier. The complex demographic history of North Atlantic humpback whales is likely related to ancient population structure and perturbations in climate during the Pleistocene. We find that accelerated evolution in cetacean genomes include gene duplications and positive selection in many cancer-relevant pathways including cell signaling, cell cycle control, cell proliferation, and apoptosis, as well as in many tumor suppressor genes that are causally implicated in human cancers. Our results suggest that continued conservation efforts of these enigmatic ocean giants will contribute not only to their persistence on Earth but also to our understanding of how multicellular life has combatted cancer through the eons.

W554: Genomics of Non-Classical Model Animals
Safeguard the Genome for Healthy Aging: Lessons from the Subterranean Blind Mole Rat, Spalax
Imad Shams, Vered Domankevich, Amani Odeh and Irena Manov, University of Haifa, Haifa, Israel
The blind mole rat (*Spalax*) is a wild solitary long-lived rodent (~20 years) that tolerates environmental hypoxia and resists cancer, which implies molecular adaptations to prevent genomic instability underlying cancer and aging. Here, we confirm that *Spalax* cells resist genotoxic insults (i.e. UV, ionizing radiation), accumulate less genotoxic lesions, and maintain enhanced repair capacity. Since persistent DNA damage response (DDR) triggers senescence, we also addressed cellular senescence program in *Spalax* cells. Cellular senescence is an important program designed to stop the division of damaged cells. Yet such cells also express an inflammatory signature, the so-called senescence-associated secretory phenotype (SASP). Accumulating with aging, senescent cells induce chronic inflammation and support cancer-promoting microenvironment. In this context we investigated whether cellular senescence in *Spalax* cells is associated with inflammatory responses known in human and other animals. In contrast to mouse and human, senescent *Spalax* cells did not accumulate DNA damage and showed undetectable expression of key SASP factors IL1α and IL6, indicating the uncoupling of SASP from cellular senescence as a unique feature of *Spalax* senescent cells. Our results strongly support that this species has evolved efficient mechanisms to maintain DNA integrity and to avoid age-related maladies as prerequisites of survival and fitness under the stressful conditions in its subterranean habitat.

W555: Genomics of Non-Classical Model Animals
Giraffe Genome Reveals Clues to its Unique Anatomy and Turbocharged Cardiovascular System
Douglas Cavener, Lan Wu-Cavener and Chelsea Hudson, Pennsylvania State University, University Park, PA
We sequenced the genomes of Masai giraffe (*Giraffa camelopardalis tippelskirchii*) and okapi (*Okapia johnstoni)* and performed comparative genomic analyses with other mammalian species in order to identify specific genetic changes in giraffe that might be responsible for giraffe's extraordinary stature and associated cardiovascular adaptations (Nature Comm. 2016). Among the approximate 19,000 genes we identified 70 genes that had multiple signs of adaptation. A large fraction of these genes were well known regulators of skeletal, cardiovascular, and neurological development and/or physiology. We are experimentally testing the role of one of these genes, FGFRL1, that is known to play an essential role of skeletal and cardiovascular development in humans and mice. To further explore the genetic regulatory differences in giraffe we are conducting a gene expression analysis among giraffe, okapi, cattle, sheep, and goat.

W556: Genomics of Non-Classical Model Animals
Backward Genomics: Turtle and Bird Genomes Provide Clues to Dinosaur Genomics
Denis Larkin, Royal Veterinary College, University of London, London, United Kingdom

W557: Genomics of Phytoremediators, Metal Accumulators and Relatives
Genomics of Endophyte Assisted Phytoremediation
Sharon L. Doty[1], Andrea Firrincieli[1], Adam Deutschbauer[2], Pierre M. Joubert[1] and Robert Tournay[1], (1)University of Washington, Seattle, WA, (2)Lawrence Berkeley Natl Lab, Berkeley, CA
Endophytes, the microorganisms within plants, can improve phytoremediation success directly, by detoxifying pollutants, and indirectly by promoting root growth, reducing general stress responses, and overall improving plant health. *Enterobacter* sp. strain PDN3 was isolated from hybrid poplar collected from a trichloroethylene (TCE)-contaminated site. TCE, a solvent and degreaser, is one of the most common environmental pollutants. Strain PDN3 grew well on high concentrations of TCE and aerobically dechlorinated TCE without the addition of co-metabolite inducers. In an effort to determine the candidate genes involved in TCE degradation by PDN3, an extensive genome analysis was performed. Interestingly, despite the presence of nine different genes encoding for monooxygenases, none of these encode for proteins known to be involved in TCE degradation, meaning that TCE degradation in PDN3 likely occurs through a novel non-canonical dechlorination pathway. The genomic analysis confirmed the earlier report that no TOM (toluene monooxygenase) genes could be amplified by PCR and none of the typical aerobic TCE degradation intermediates could be detected by GC-TOF-MS. Heavy metals are common co-contaminants of polluted soils and are usually found as trace elements in non-contaminated soils (89). PDN3 carries genes which are part of copper, cadmium, arsenate, tellurite and hexavalent chromium detoxification systems, suggesting that it may be an effective inoculant for phytoremediation of mixed polluted sites. Random barcoded TnSeq experiments are underway to determine the genes required by PDN3 to grow in media containing TCE and arsenic.

W558: Genomics of Phytoremediators, Metal Accumulators and Relatives
Role of Na^+ Accumulation and Na^+ Exclusion in Salinity Tolerance of Plants
Mark A. Tester, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
Forty percent of the world's food is produced under irrigation, and this is directly threatened by over-exploitation and changes in the global environment. One way to address this threat is to develop systems for increasing our ability to use lower quality water, in particular saline water. Low cost partial desalination of brackish water, use of saline water for cooling greenhouses, and increases in the salinity tolerance of crops can all contribute to the development of this new agricultural system.

We have been using forward genetic approaches for discovery of genes related to salinity tolerance in barley and tomatoes. Rather than studying salinity tolerance as a trait in itself, we dissect salinity tolerance into a series of components that are hypothesized to contribute to overall salinity tolerance (following the paradigm of Munns & Tester, 2008).
For example, one component of tolerance of most crop plants to soil salinity that has been classically considered to be important is the ability to maintain low concentrations of Na^+ in the leaves, and much analysis of this aspect has been done. A major site for the control of shoot Na^+ accumulation is at the plasma membrane of the mature stele of the root, controlled by genes encoding HKT1 Na^+ transporter proteins. However, yield benefits from Na^+ exclusion appear to mainly occur under high salinities, with little benefit being seen at moderate salinities – where most agriculture occurs. As such, the role of Na^+ exclusion and Na^+ accumulation in salinity tolerance will be critically examined in this talk. The energetics of Na^+ transport will also be discussed.

W559: Genomics of Phytoremediators, Metal Accumulators and Relatives
Heavy Metal Sensing and Signaling Mechanisms in Plants
Mather A Khan, Norma Castro-Guerrero, Nga T Nguyen, Sam McInturf and David Mendoza-Cozatl, University of Missouri, Columbia, MO
Plants and seeds are the main dietary source of essential micronutrients (Iron, Zinc, Copper, Manganese) and also the main entry point for non-essential elements (Cadmium, Arsenic) into the food chain. Under low Fe or in the presence of Cd, plants change their transcriptional programs but the nature of the signals and sensors and are unknown. Understanding sensing and allocation in plants will help developing crops with higher nutritional value and minimal accumulation of non-essential elements.
In Arabidopsis, OPT3 has recently been identified as a component of the systemic network mediating Fe deficiency responses. *opt3-2* mutants shows a constitutive Fe-deficiency response and over-accumulate Fe in roots and leaves. Using RNA-seq, we demonstrated that *opt3-2* roots display an activation of the major networks mediating Fe uptake. However, markers for Fe excess are exclusively induced in leaves, suggesting that Fe excess is properly sensed in leaves and that OPT3 mediates the shoot-to-root signaling critical to prevent an Fe overload. In addition, we have found that the leaf vasculature responds more rapidly than roots to changes in Fe availability, suggesting that the vasculature is the primary site for sensing the Fe status of the whole plant. Our current experiments, including high-throughput protein-DNA interaction together with gene network analyses of leaf specific RNA-seq data, are directed towards the identification of the transcriptional networks that coordinate the response to changes in nutrient availability and crosstalk between different nutrients.

W560: Genomics of Phytoremediators, Metal Accumulators and Relatives
The Genomics of Adaptation to Serpentine Soils in *Mimulus guttatus*
Jessica Selby, Bayer, Woodland, CA, Kristin Lee, Columbia University, New York, NY, Graham Coop, UC Davis, Davis, CA and John H. Willis, Duke University, Durham, NC
Some of the most striking examples of the power of natural selection to shape biological diversity involve adaptation of plants to extreme soil environments such as serpentine soils. The unique chemical and physical properties of serpentine soils prove a difficult substrate for plant growth and most species are excluded from these habitats. However, some species, such as *Mimulus guttatus,* have repeatedly adapted to serpentine soils, providing an opportunity to investigate whether populations have adapted via the same or different genetic mechanisms. In this study we use both QTL mapping and whole genome re-sequencing of native populations to characterize the genetic basis of repeated adaptation to serpentine soils in *M. guttatus*. Common garden studies show that populations of *M. guttatus* are locally adapted to their native soil habitat with plants from non-serpentine populations unable to survive past the juvenile stage when planted in serpentine soils. We map these survival differences in eleven different F2 populations generated from crosses between lines from unique serpentine populations from throughout California and Oregon crossed to a common non-serpentine line. We find that a previously identified QTL known to control survival in the field as well as tolerance to low Ca:Mg has major effects on survival in these widespread populations. By re-sequencing pooled samples from these same serpentine populations along with nearby non-serpentine populations we identify additional loci contributing to serpentine adaptation that appear to be widely shared. Furthermore, the population resequencing data enables identification of specific candidate genes, including a promising candidate within the QTL region. Many of these genes have functions related to Ca homeostasis, ion transport and stress tolerance and future works aims to characterize the mechanisms underlying serpentine tolerance in *M. guttatus.*

W561: Genomics of Phytoremediators, Metal Accumulators and Relatives
Assessing the Degree of Convergent Evolution and Gene Flow in Wild Outcrossing Arabidopsis Relatives on Serpentine and Metalliferous Soils
Levi Yant, University of Nottingham, Norwich, United Kingdom
It is a tempting hypothesis that parallel adaptation events to the same environmental challenge should result in genetic changes of similar or identical effects, depending on underlying fitness landscapes. However, systematic testing of this hypothesis is scarce. Our group in collaboration with others has performed several studies to begin to test this hypothesis in wild outcrossing Arabidopsis relatives that have adapted to similar or identical metalliferous soil habitats. Here I describe results from two of these studies: one on serpentines and the other on mines. In both cases we use a population-based evolutionary genomic approach coupled with elemental profiling to assess how Arabidopsis arenosa and either A. lyrata (serpentine) or A. halleri (mines) adapted to highly challenging, metal-rich habitats. We first demonstrate that serpentine- or mine-adapted plants exhibit dramatically altered elemental accumulation levels in common conditions relative to matched nonmetalliferous controls, and then resequence 61 (mine contrasts) or 64 (serpentine contrasts) individuals from 16 populations to perform high density genome scans for signatures of selective sweeps. We find evidence for highly localized selective sweeps that point to polygenic, multitrait bases for adaptation to both site types. We identify a handful of candidate alleles for mediating convergent evolution between A. arenosa and A. lyrata on serpentine, and fewer alleles convergent between A. arenosa and A. halleri on mines, although several conspicuous exceptions stand out. This overall modest overlap between independent colonizations of different sites within and between species suggests the number of evolutionary strategies suited to overcome the challenges of serpentine or mine adaptation are not severely limited. Additionally, in several cases, quantitative tests of introgression indicate that some alleles exhibiting marked selective sweep signatures appear to have been introgressed from A. lyrata. This finding suggests that migrant alleles may have facilitated adaptation of A. arenosa to this environment. This finding, in combination with other

population genomic studies from our group suggests that this may not be an uncommon pattern. Our data suggest that species-specific metal handling and other biological features could explain a low degree of convergence between species. The parallel establishment of plant populations on metalliferous soils involves a modest degree of convergence, which will likely be more pervasive across sites purposely chosen for maximal similarity in soil composition.

W562: Genomics of Phytoremediators, Metal Accumulators and Relatives

Extremophyte Arabidopsis Relatives Thrive in Soils Contaminated with Toxic Concentrations of Sodium and Lithium

Kieu-Nga T Tran, Dong-Ha Oh and Maheshi Dassanayake, Louisiana State University, Baton Rouge, LA

Recent availability of genomes for naturally salt-adapted extremophyte wild-relatives of Arabidopsis enables investigating how plants may have evolved modified or novel molecular networks to cope with the salt stress which, in turn, will offer an opportunity to adopt naturally existing stress response mechanism for the development of stress-resilient crops. *Schrenkiella parvula, one of the most salt-tolerant species in Brassicaceae, is a* native plant to the shores of the hypersaline lake Tuz in Turkey, and can complete its life cycle in the presence of high salt concentrations of Na^+, K^+, and Li^+, that are lethal to most plants. Despite the different evolutionary paths and adaptive strategies, the genomes of Arabidopsis and *S. parvula* show high macrosynteny and over ~70% of gene models can be identified as orthologous pairs. Using comparative transcriptomic approaches, I have analyzed a time-dependent transcriptomic response to Na^+ and Li^+ stress; developed metabolomic and ionomic profiles for each of the transcriptomic sample to obtain a global view of how each plant responds to different stress concentrations at different stress durations; and assessed the physiological responses to Na^+ and Li^+ stress. I plan to present the key pathways that are unique to the extremophyte as well as the differences from the canonical stress response pathways known to model plants. Some of the stress responsive pathways in Arabidopsis that gets highly induced upon stress are already expressed at high levels prior to the stress implying stress preparedness in the extremophyte. Ionomic profiles reveal the ability of the extremophyte to maintain low levels of Na^+ and Li^+ in shoots and possibly continue to uptake macronutrients and micronutrients under salt stress for a longer duration compared to the stress sensitive model, Arabidopsis.

W563: Genomics of Plant Development

Regulation of Barley Leaf Patterning and Axillary Branching

Gary J. Muehlbauer, University of Minnesota, St. Paul, MN

Regulation of lateral organ and ligule development share some of the same genes. In barley, tillers (lateral branches) are derived from axillary meristems at the crown of the plant, and the ligule is an epidermal flap of tissue that divides the sheath and blade. Many genes that regulate lateral organ development and leaf patterning are referred to as boundary genes because they mark the boundary of a new organ. For example, the barley *UNICULME4* (*CUL4*) gene regulates tiller and ligule development, and CUL4 transcripts are detected in the leaf axil and ligule, indicating CUL4 acts at boundaries. To further examine the genetic control of leaf and tiller development, my lab has conducted laser capture microdissection RNA-SEQ of the ligule and axillary meristem in two wildtype genotypes and identified genes that are expressed in both ligule and axillary meristem tissues. In addition, we are studying three classes of barley mutants including: (1) those that exhibit few tillers and a liguleless phenotype (*CUL4* and *ELIGULUM-A*), (2) those that exhibit few tillers and a wildtype ligule boundary (*UNICULM2*), and (3) those that exhibit wildtype tiller number and a liguleless phenotype (*HvLIGULESS1*). We showed that *ELIGULUM-A* and *HvLIGULESS1* encode an unknown protein and a squamosa promoter binding protein, respectively. RNA in situ hybridizations showed that *ELIGULUM-A* and *HvLIGULESS1* play a role in establishing boundaries. The genetic interactions between *ELIGULUM-A*, *HvLIGULESS1* and *CUL4* will be discussed.

W564: Genomics of Plant Development

Tilling by Target Capture Sequencing as a Tool for Functional Gene Analyses in Soybean

Leandro G. Neves, RAPiD Genomics, Gainesville, FL

W565: Genomics of Plant Development

Application of TILLING and BSA-Seq Approach to Map Prominent Main Stem and Leaf Spot Resistance in Groundnut

Annapurna Chitikineni[1], Hui Wang[2], Gaurav Agarwal[2], Aamir W Khan[1], Manish K Pandey[1], Peggy Ozias-Akins[2], Corley Holbrook[3], Rajeev K Varshney[1] and Baozhu Guo[2], (1)ICRISAT, Hyderabad, India, (2)University of Georgia, Tifton, GA, (3)United States Department of Agriculture - Agricultural Research Service, Tifton, GA

Targeting Induced Local Lesions in Genomes (TILLING) is considered a powerful reverse genetics approach for functional genomics studies. However, because of availability of low-cost and high-throughput sequencing technology, it has become possible to sequence TILLING lines and identify SNPs associated with genes responsible for traits. One TILLING population has been developed in the “Tifrunner” genotype of groundnut, an economically important oilseed crop grown in tropical and warm temperate regions of the world. The TILLING population has shown phenotypic variation for several traits including resistance to leaf spots and the features of prominent main stem. A total of 25 lines comprising of 16 susceptible and 9 resistant lines for leaf spots, and 11 lines with presence and 14 lines with absence of the prominent main stem from the TILLING population were sequenced on Illumina HiSeq 2500 and a total of 745.8 Gb sequencing data has been generated. These sequence data are being analyzed to identify structural variations including SNPs and INDELs across the lines with Tifrunner. In parallel, two mapping populations from these TILLING lines namely T47-7 (resistant to leaf spots) x T33-3 (susceptible to leaf spots) and T90-1 (presence of stem) x T71-2 (absence of stem) are being developed. It is planned to phenotype the segregating progenies and also sequence the extreme bulksA of segregating progenies for these traits. We anticipate identification of candidate genes and SNPs for these important traits by deploying the BSA-Seq approach in groundnut in due course.

W566: Genomics of Plant Development

Whole Genome Re-Sequencing Reveals the Impact of Copy Number Variants of the *Rhg4* Gene on Broad-Based Resistance to Soybean Cyst Nematode

Gunvant Patil[1], Naoufal Lakhssassi[2], Tri D. Vuong[3], Jinrong Wan[4], Mariola Klepadlo[3], Adrian O. Stec[5], Babu Valliyodan[3], Tarek Hewezi[6], Robert M. Stupar[5], Henry T. Nguyen[3] and Khalid Meksem[7], (1)University of Minnesota, St Paul, MN, (2)Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL, (3)University of Missouri, Columbia, MO, (4)University of Missouri & National Center for Soybean Biotechnology, Columbia, MO, (5)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (6)University of Tennessee, Knoxville, TN, USA,, Knoxville, TN, (7)Southern Illinois University Carbondale, Carbondale, IL

Soybean cyst nematode (SCN) is the most devastating plant-parasitic nematodes worldwide. The majority of the commercial soybean varieties with SCN resistance was derived from PI88788. The effectiveness of resistance derived from PI88788 is breaking down due to narrow genetic background and SCN population's shift. PI88788 requires mainly the *rhg1-b* locus, while 'Peking' requires *rhg1-a* and *Rhg4* for SCN resistance,. In the present study, whole-genome sequencing of 106 soybean lines was used to define the *Rhg* haplotypes and investigate their responses to the SCN HG-Types. The analysis showed a comprehensive profile of SNPs and copy number variations (CNV) at these loci. CNV of *rhg1* (GmSNAP18) was confirmed to contribute towards resistance only in lines derived from PI88788 and "Cloud". At least 5.6 copies of the PI88788-type *rhg1* are required to confer SCN resistance, regardless of the *Rhg4* (*GmSHMT08*) haplotype. However, when the *GmSNAP18* copies dropped below 5.6, a "Peking"-type *GmSHMT08* haplotype was required to ensure SCN resistance. This points to a novel mechanism of epistasis between *GmSNAP18* and *GmSHMT08* involving minimum requirements for copy number. The presence of more *Rhg4* copies confers resistance to multiple SCN races. Moreover, transcript abundance of the *GmSHMT08* in root tissue correlates with more copies of the *Rhg4* locus, reinforcing SCN resistance. Finally, haplotype analysis of the *GmSHMT08* and *GmSNAP18* promoters inferred additional levels of the resistance mechanism. This is the first report revealing the genetic basis of broad-based resistance to SCN and providing new insight into epistasis, haplotype-compatibility, CNV, promoter variation, and its impact on broad-based disease resistance in plants.

W567: **Genomics of Plant Development**

Going Beardless: Identification of the B1 Awn Inhibitor Gene from Durum and Bread Wheat Populations

J. Allan Feurtado[1], Daiqing Huang[1], Qian Zheng[1], Tancey Melchkart[1], Yasmina Bekkaoui[1], David J. F. Konkin[2], Venkatesh Bollina[1], Sateesh Kagale[1], Martial Martucci[3], Frank M. You[4], Curt McCartney[3], Martha Clarke[5], Nikolai M. Adamski[5], Catherine Chinoy[5], Paul Nicholson[5] and Adrian J Cutler[1], (1)National Research Council Canada, Saskatoon, SK, Canada, (2)National Research Council of Canada, Saskatoon, SK, Canada, (3)Agriculture and Agri-Food Canada, Morden, MB, Canada, (4)Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (5)John Innes Centre, Department of Crop Genetics, Norwich, United Kingdom

W568: **Genomics of Tissue Regeneration in Plants and Animals**

The Role of Cell Plasticity during Pattern Formation in the Axolotl Limb Regenerate

Catherine McCusker, College of Science and Mathematics, Boston, MA

W569: **Genomics of Tissue Regeneration in Plants and Animals**

High throughout Transcriptomics of a *Solanum* Introgression Population to Unravel Novel Determinants of the Regeneration Process

Samuel Bocobza, The Institute of Plant Sciences, ARO, Rishon Lezion, Israel

Regeneration of whole plants from tissue explants is a key technology for crop improvement and plant biotechnology. Nevertheless, while this technology can be highly advantageous, it is not applicable to many recalcitrant plant species. In order to identify novel genetic factors that limit a biological process, such as plant regeneration, a differential gene expression analysis may be taken. However, the resolution such an analysis greatly depends on the number of samples used in this experiment, and small scale experiments are likely to yield unclear results. Thus, to increase the resolution of differential expression analysis and discern a limited number of candidate genes, a high-throughput transcriptomic approach involving hundreds of samples is highly profitable. To this end, I will present a novel 3' end sequencing method and its applications. This method significantly reduces costs and allows the characterization of hundreds of samples' transcriptomes in a single experiment. Furthermore, this technique allows the identification of gene heterogeneity among the samples analyzed, and also facilitates the isolation of genetic hot spots that control a given biological process. This high-throughput approach can be applied to nearly any biological question in order to determine the genetic factors that direct a given biological process.

W570: **Genomics of Tissue Regeneration in Plants and Animals**

Unlocking the Mysteries of Stem Cells

Robert A.J. Signer, Division of Regenerative Medicine, Moores Cancer Center, La Jolla, CA

W571: **Genomics of Tissue Regeneration in Plants and Animals**

Breaking and Entering: Molecular Insights into Fertilization and Viral Infection of Gametophytes and Seeds

Ben Rimon, The Institute of Plant Sciences, ARO, The Volcani Center, POB 15159, Rishon Lezion, Israel

W572: **Graft Genetics and Genomics**

Characterization of Epigenetic Variations in *Brassica* Grafting Chimera and their Heritability in the Progeny

Liping Chen, Zhejiang University, Hangzhou, China

As an effective means of vegetative propagation, plant grafting is widely employed to improve tolerance to stresses or diseases, increase yield, and promote vigor. However, phenotypic variations acquired by plant grafting have been observed in a number of studies, and the mechanism and the heritability of grafting-induced variation in progeny is not well understood. In our present study, *in vitro* shoot apical meristem (SAM)

grafting between tuber mustard and red cabbage was performed and produced two interspecific periclinal chimeras, TCC and TTC (where the origin of the outer, middle, and inner cell layers, respectively, of the SAM is designated by a 'T' for tuber mustard and 'C' for red cabbage). Phenotypic variations, which mainly showed in leaf shape and wax content, were observed in selfed progeny GSn (GS = grafting-selfing, n = generations) of TTC and asexual progeny r-CCCn (r = regenerated,) regenerated from TCC. Here the heritability of phenotypic variation and its association with epigenetic variation, mainly in methylation modification and sRNA fluctuation, in the sexual and asexual progenies of chimeras were investigated. The results showed that the communication between heterologous cells of different cell layers results changes in the number and variety of small RNAs. Moreover, we observed global DNA methylation modifications in the sexual and asexual progenies of chimeras compared with their corresponding parents, mainly in the repeat elements, and the acquired DNA methylation changes could be both heritable and reversible. Interestingly, we determined that grafting mobile and triggered siRNAs were responsible for directing and maintaining the DNA methylation modifications, especially in repeat elements.

W573: **Graft Genetics and Genomics**
Mobile Florigenic Signals in Woody Plants: Flowering LOCUS T, Sugar, and Phytohormones
Guo-Qing Song, Michigan State University, East Lansing, MI

W574: **Graft Genetics and Genomics**
Molecular Mechanism of *Nicotiana* Interfamily Grafting
Michitaka Notaguchi, Graduate School of Bio-agricultural Sciences, Nagoya University, Nagoya, Japan
Plant grafting has been an important technique in agriculture to propagate clones and to obtain benefits of certain rootstocks. However, graft-incompatibility has limited the technique. Recently, we found that a genus, *Nicotiana*, showed an extreme capability of grafting. *Nicotiana* can be grafted with a wide range of vascular plants. To investigate the molecular basis of *Nicotiana* interfamily grafting, we performed time-course transcriptome analysis on grafting region. By comparing transcriptome of interfamily grafting of *Nicotiana* and ones of the other plants, we identified characteristic upregulation of several genes related to cell wall modification. A virus induced gene silencing system was applied to test the function of identified genes in grafting. We also observed morphology change at graft boundaries of *Nicotiana* interfamily grafting where cell wall was partially digested resulting in tissue adjoining of grafted plants. We will present these data and discuss about the characteristic features of *Nicotiana* interfamily grafting.

W575: **Graft Genetics and Genomics**
Elevated Auxin Content and Reduced Cytokinin Level in Rootstock Improve Grafting Success
Yi Li[1], Wei Li[2], Chu Fang[3], Longmei Zhai[1], Richard McAvoy[1], Ziniu Deng[4] and Janice Zale[5], (1)University of Connecticut, Storrs, CT, (2)CHina Ag University, Beijing, China, (3)University of Connecticut, Storrs,, CT, (4)Hunan Ag University, Changsha, China, (5)University of Florida - IFAS, Gainesville, FL
Many rootstock plants suffer from undesirable lateral bud outgrowth, low grafting success rates or poor rooting. We have used a root-predominant gene promoter (SbUGT) to drive the expression of a tryptophan-2-monooxygenase gene (iaaM) to increase auxin levels in tobacco. The transgenic plants, when used as a model rootstock, displayed inhibited lateral bud outgrowth, enhanced grafting success rate and improved root initiation. However, root elongation and biomass of SbUGT::iaaM transgenic plants were reduced compared to those of wild-type plants. In contrast, expression of the SbUGT::CKX (CKX: a cytokinin degradation gene enhanced root elongation and biomass, which could neutralize the negative effects of auxin overproduction. Our results demonstrate that expression of both the iaaM and CKX genes in roots of rootstock inhibits lateral bud release from rootstock, improves grafting success rates and enhances root initiation and biomass, which may provide a useful tool for grafting.

W576: **Graft Genetics and Genomics**
Rootstock Effects on Shoot System Phenotypes in Grafted Grapevines
Zoë Migicovsky[1], Arianna Bozzolo[2], Daniel Chitwood[3], Anne Fennell[4], Zachary Harris[5], Laura L. Klein[6], Laszlo Kovacs[7], Misha Kwasniewski[8], Mao Li[9], Jason Londo[10], Qin Ma[4], Adam McDermaid[4] and Allison Miller[6], (1)Dalhousie University, Truro, NS, Canada, (2)University of Missouri, Colombia, MO, (3)Michigan State University, East Lansing, MI, (4)South Dakota State University, Brookings, SD, (5)Saint Louis University, St. Louis, MO, (6)Saint Louis University / Danforth Plant Science Center, St. Louis, MO, (7)Missouri State University, Springfield, MO, (8)University of Missouri, Columbia, MO, (9)Donald Danforth Plant Science Center, St. Louis, MO, (10)USDA-ARS Grape Genetics Research Unit, Geneva, NY
Grapevine is an excellent model for understanding how rootstocks can impact shoot systems phenotypes due to the available genomic resources and ability to grow across diverse environments. We examined an experimental vineyard in Mount Vernon, Missouri which includes a locally important scion ('Chambourcin') own-rooted as well as grafted onto three different rootstocks (SO4, 1103P and 3309C). The vineyard also includes 3 different irrigation treatments. From 2013-2016, we assessed different kinds of phenotypic variation including leaf ion concentrations, viticulture measurements such as pruning weight, and GC-MS using a targeted panel of metabolites in berries and wine. We also examined rootstock-induced changes in gene expression in the scion using RNA-seq. Each phenotype was studied for 1 to 3 years. We found distinct and significant effects of rootstock and irrigation on the phenotypes examined. Current work underway expands sampling to include additional phenotypes, samples, and time points across three years.

W577: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**
Gramene Database Update: Genome & Pathways Updates
Pankaj Jaiswal, Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR

W578: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**

Towards the Completion of a Set of Platinum Standard Reference Genome Sequences (PSRefSeqs) that bridge the Genetic Diversity of Cultivated Rice & its Wild Relatives

Rod A. Wing, Arizona Genomics Institute, University of Arizona, Tucson, AZ

The International *Oryza* Map Alignment Project (I*O*MAP) is interested in discovering and utilizing standing natural variation from the genus *Oryza* to breed new varieties of rice that are higher yielding and more nutritious, but have less of an environmental footprint (i.e. rice that requires less water, fertilizer & pesticides, can grow on marginal lands and have reduced greenhouse gas emissions) – a.k.a. "Green Super Rice". A critical component for the success of this project is to have access to platinum standard reference genomes (PSRefSeqs) that represent the genetic diversity of cultivated rice and its wild relatives. Such genomes will serve as a pan-*Oryza*¬genome that can be used as a template to map resequencing data from rice germplasm banks across the globe aimed at capturing virtually all genetic variation that exist across the *Oryza*sphere. Such information can then be used in GWAS studies and genomic selection strategies, coupled with high throughput phenotyping, and gene editing to breed the crops of the future to help solve the 2050 10-billion people question – i.e. "How do we feed our world without destroying our world". In this workshop, I will present the status of our efforts to generate PSRefSeqs that bridge the *Oryza* genus.

W579: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**
Whole Genome Assembly and Annotation of the Maize NAM Founders

Jianing Liu and R. Kelly Dawe, University of Georgia, Athens, GA

Maize is not only an important food crop but also an important model organism for genetics studies. To the present day, maize is grown globally and has adapted to a range of environments. The genome architecture of maize has evolved along with the expansion of its habitat, and high levels of genetic divergence are present among inbreds. However, with only one genome of maize inbred B73 available as reference, the study of genome variation among maize inbreds has been impeded due to limited resources. In this study, we are sequencing and assembling the genomes of B73 and 25 other maize inbreds, which were chosen to represent broad maize diversity and known as NAM founder lines. We will incorporate PacBio sequencing, BioNano optical mapping, Illumina-sequencing to construct de novo sequence assemblies and genetic maps will be employed for final scaffolding. RNA-seq from six tissues will be used to annotate the genomes, and released with browser support to Gramene, MaizeCODE and NCBI. Pangenomes will be synthesized with comparative genomic analysis to index the total collection of genes present all 26 sequenced lines. We will also characterize structural variation including copy number variation and rearrangements to test whether these structural polymorphisms contribute to agronomic traits or local adaptation.

W580: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**
Scoring Alternative Transcripts

Sarina Awatramani[1], Pragati Muthukumar[2], Andrew Olson[3] and Doreen Ware[3], (1)UCLA, Los Angeles, CA, (2)CSHL, Cold Spring Harbor, NY, (3)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

The maize B73 v4 reference genome annotations were based on a variety of transcriptome evidence including a large number of full length cDNAs sequenced with PacBio Iso-Seq. While these annotations have been a great advancement, the abundance of transcript isoforms in many gene loci present some obstacles. One challenge is that it is difficult to know which transcript models are most relevant in a given experimental context. Furthermore, it complicates the choice of a canonical transcript for comparative gene family analysis, because the canonical transcript, conventionally defined as the longest transcript with the longest open reading frame, often has a dubious in-frame retained intron. Therefore, we developed additional metrics based on gene expression and InterPro domains to rank and filter the many transcript models. Implementing these methods will simplify experimental work and improve downstream analyses.

W581: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**
Involving the Research Community in Biocuration of Genes and Pathways

Sushma Naithani, Justin Preece, Parul Gupta and Pankaj Jaiswal, Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR

Researchers and database users can play a crucial role in advancing public genomic resources and benefit from active engagement. Acquiring Big Data literacy and biocuration skills are important incentives for researchers, and especially could help young researchers to think more critically about their research projects and get acquainted with relevant bioinformatics tools and data resources. Plant Reactome (http://plantreactome.gramene.org) curators tested strategies for introducing researchers to pathway curation tools, harnessing biologists' expertise in curating plant pathways, and developing a network of community biocurators. We will discuss the strategy, workflow, and outcomes of our on-site and online efforts towards involving researchers and database users in plant genes and pathway curation. The Plant Reactome database is funded by an NSF award (#IOS-1127112) to the Gramene project. It is produced with intellectual and infrastructure support provided by the Human Reactome award (NIH: P41 HG003751, ENFIN LSHG-CT-2005-518254), Ontario Research Fund, and European Bioinformatics Institute (EBI) Industry Programme).

W582: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**
Dive: Publication Pipeline Integration with Automated Biological Entity Detection and Validation Service

Amit Gupta, TEXAS ADVANCED COMPUTING CENTER, Austin, TX and Weijia Xu, Texas Advanced Computing Center, Austin, TX

W583: **Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities**
Prototyping Community Curation Approaches for Improving Maize Genome Annotation

Rajdeep Khangura[1], Cristina F. Marco[2], Marcela K. Tello-Ruiz[3], Hao Wu[4], Junpeng Zhan[5], Michelle C. Stitzer[6], Pengfei Qiao[7], Rachael Wasikowski[8], Fei-Man Hsu[9], Sirjan Sapkota[10], Michael Okoro[11], Andrew Olson[3], Kapeel Chougule[3], Doreen Ware[3] and David Micklos[3], (1)Department of Biochemistry, Purdue University, West Lafayette, IN, (2)DNA Learning Center, Cold Spring

Harbor Laboratory, Cold Spring Harbor, NY, (3)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (4)Department of Genetics, Developmental and Cell Biology, Iowa State University, Ames, IA, USA, Ames, IA, (5)School of Plant Sciences, University of Arizona, Tucson, AZ, (6)Department of Plant Sciences, University of California, Davis, CA, (7)School of Integrative Plant Science, Cornell University, Ithaca, NY, (8)Department of Biological Sciences, University of Toledo, Toledo, OH, (9)Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan, (10)Institute for Translational Genomics, Clemson University, Clemson, SC, (11)Center of Genomics and Systems Biology, New York University, New York, NY
Genome annotation is the most critical component of a genome assembly. It is the process through which the structural and functional knowledge of a genome is associated to its biological features, such as genes and transposons. The first draft of the maize B73 genome was put together in 2009, yet its initial assembly was fragmented. With the improvement in sequencing chemistry, the latest genome assembly (B73RefGen_V4) of maize offered a much complete picture of the genic and intergenic landscape. Automated genome annotation pipelines are very efficient at predicting the gene models, but even those algorithms make mistakes. So, can we improve the genome annotation pipeline and make it more accurate? In December of 2017, the NSF-funded MaizeCode and Gramene projects organized the First Maize Genome Annotation Jamboree to manually curate the V4 gene models suspected to be erroneous by MAKER-P quality metrics. Participants (graduate students and postdocs) in this workshop curated five gene families using the Apollo genome annotation tool, and ~90% of the genes flagged by AED and QI2 annotation quality criterion in MAKER-P were found to have discrepancies in gene models. With this initial training, participants were tasked to manually triage and curate subsets of the 419 so called “maize classical genes” and report their progress and pitfalls during bi-weekly conference calls. Gene models in the order of thousand were visited using Gramene’s gene tree visualization tool to flag gene models that show a clear discrepancy from its orthologues in eudicots and monocots. At least three participants visited the same gene model independently, and genes that were flagged by 2 of 3 participants were analyzed through the Apollo platform. If the gene model seemed erroneous, the correct gene model was reconstructed by the participants using the available transcriptomic and protein evidence. As of now, a subset of the gene models that were improved are being experimentally validated using RT-PCR amplification from targeted plant tissues. Initial results indicate user annotated gene models to be correct. This data is informing the development of new workflows to automate the curation process and further improve the gene prediction for not only B73 RefGen_V4 but also the maize pan-genome.

W584: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities
Visualisation and Analysis of Plant Gene Expression Functional Genomics Data in Expression Atlas
Nancy George[1], Nuno Fonseca[1], Anja Fullgrabe[1], Laura Huerta[1], Haider Iqbal[1], Monica Jianu[1], Jon Manning[1], Suhaib Mohammed[1], Pablo Moreno[1], Alfonso Munoz-Pomer[1], Alvis Brazma[1], Lingyun Zhao[1], Irene Papatheodorou[1] and Functional Genomics , (1)European Bioinformatics Institute, Hinxton, United Kingdom
Expression Atlas (www.ebi.ac.uk/gxa) and its newest component Single Cell Expression Atlas (www.ebi.ac.uk/gxa/sc) are value added databases and web services that collect, annotate and consistently re-analyse and display gene expression data from both bulk and single cell RNA-seq studies. Eligible datasets; e.g. minimum biological replicates, are sourced from various archives, including ArrayExpress; GEO and SRA/ENA/DDBJ. Once selected, samples are curated to comply with minimal metadata standards and mapped to Experimental Factor Ontology (EFO) terms. This facilitates comparison across datasets and easy data retrieval. Once curated, comparisons are decided based on experimental variables.
Data analysis for both bulk and single cell data RNA-seq data is performed using our standardized pipeline iRAP (https://nunofonseca.github.io/irap) while our microarray pipeline use standard open source tools. To explore the data, Expression Atlas can be search for specific gene(s) or different biological conditions, assisted by ontology search expansion. Data is represented either in baseline context, e.g. normal gene expression across different tissues or differential e.g. changes in gene expression in response to drug treatment or other perturbation. Within datasets, users can explore transcript quantification; biological replicate variation; gene co-expression and visualize gene expression in the Gramene genome browser. Furthermore we also display enrichment of gene ontology terms and pathways from the REACTOME database.
Currently, there are approx 3,450 experiments across 59 species, of which almost 800 are plant experiments across 36 different species. All data in Expression Atlas is free for use; reanalysis and download and all analysis results are available programmatically via our RNASeq-er API (https://www.ebi.ac.uk/fg/rnaseq/api/).

W585: Grape Genome Initiative
The Role of ERF6L in Grapevine
Haley S. Toups, University of Nevada, Reno, Reno, NV

W586: Grape Genome Initiative
***Vitis riparia* Draft Genome Assembly and Analysis**
Michael Robben, Sagar Patel, Dilmini Alahakoon and Anne Fennell, South Dakota State University, Brookings, SD
Vitis riparia, a native North American species, is used globally in rootstocks and scion breeding and has multiple abiotic and biotic stress tolerance traits unique to the species. Here we report a draft assembly of the *V. riparia* genome with an approximate size of 495 Mb. Using short read RNA-seq libraries, 40,019 coding sequences were predicted from this assembly and the annotation was compared to *V. vinifera*. GBS markers from an F2 population resulting from a cross of *V. riparia* x ‘Seyval’ were used to explore genes underlying quantitative and qualitative traits.

W587: Grape Genome Initiative
Developing a Versatile System for Functional Genomic Studies in Grapevine
Satyanarayana Gouthu[1], Pat Iocco-Correna[2], Mark Thomas[2], Hiro Nonogaki[3] and **Laurent G. Deluc**[4], (1)Oregon State University, Corvallis, OR, (2)CSIRO, Urrhae, Australia, (3)Department of Horticulture/OSU, Corvallis, OR, (4)Dept. of Horticulture, Oregon State University, Corvallis, OR

Understanding the relationship between gene function and economically important traits is an essential component for genetic improvement of grapevine. Functional genomics may require the use of tractable system for genetic engineering and a versatile expression system for the temporal and spatial contextualization of gene expression. Here, we report on the use of the microvine system combined to an inducible over-expression system (Plant Gene Switch System, PGSS) containing a switch construct (AGE vector) induced by a chemical ligand methoxyfenozide (MOF). For the generation of silencing lines, we used the artificial microRNAs vectors technology containing the *AtMIR390a*-based amiRNA construct that was sub-cloned to the PGSS. Generation of the over-expressing and silencing lines for two candidate genes (*VitviNCED2* [VIT_10s0003g03750]and *VitviARF4* [VIT_06s0004g03130]) likely involved the control of ripening initiation was used to describe the features of the system. In vitro *VitviNCED2* over-expressing lines were exposed to MOF; the expression system was turned on within 72 hours and resulted in a 15-fold increase of the transcript abundance in comparison with un-induced conditions. The efficacy of the amiRNA system to silence *VitviARF4*was assessed from *Agrobacterium*-infiltrated leaves of young microvines. We observed a 40% reduction in the transcript abundance of the target mRNAs after 6 days.The use of the same system (microvine + PGSS) for CRISPR interference and activation, along with multiplex gene silencing experiments on the microvine, is also currently explored.

W588: Grape Genome Initiative

High-Fidelity RHampSeq Primers Targeting the Core Genome enables shared Exploration of Highly Diverse and Heterozygous Species

Cheng Zou, Cornell University, Ithaca, NY
Cheng Zou, Cornell University

Genotyping via high-throughput sequencing can be challenging in highly diverse and heterozygous species, but much progress has been made to solve the problems of mapping sequencing reads accurately, correcting heterozygote undercalling, resolving multi-allelic loci, and phasing alleles. In previous QTL mapping using genotyping-by-sequencing (GBS) in 17 *Vitis* families, GBS SNP marker-trait predictions from one family were not transferable to other families, so we developed a pipeline for core genome haplotype markers via amplicon sequencing. Using PN40024 as the reference genome, we constructed a pan-genome from nine *de novo* assembled genomes, representing germplasm from six wild species and three *Vitis vinifera* cultivars. The *Vitis* core genome was positively correlated with gene density, but negatively correlated with transposable element density. In core genome primer design, we required collinearity of the core-genome and moderate polymorphism of the target regions, avoiding genetic variation within primer sequences. Primers for 2000 sites were genotyped via high-fidelity amplicon sequencing with a novel RNase H2-dependent PCR system from Integrated DNA Technologies (IDT), known as rhAmpSeq. Validating this strategy in one F2 and three F1 families, 92% of markers were amplified in all four families, and 83% of markers are informative for the consensus genetic map construction. The genetic map built from the markers is highly consistent. Most importantly, the most significant marker related to flower sex in two population are the same, which indicate that not only random markers are transferable, but also the functional markers are transferable. This genotyping strategy offers elevated marker transferability, increased specificity and evenness of genotyping, tolerance of highly polymorphic target regions, and straightforward applicability to other highly diverse and heterozygous species.

Title: **High-fidelity rhAmpSeq™ primers targeting the core genome enables shared exploration of highly diverse and heterozygous species**

Submitter's E-mail Address: cz355@cornell.edu

Session Selection: Grape genome Initiative workshop

W589: Grape Genome Initiative

Genomic Diversity Among Wine Grape Cultivars and Clones

Andrea Minio[1], Amanda Vondras[1], Melanie Massonnet[1], Yongfeng Zhou[2], Edwin Alberto Solares[2], Rosa Figueroa-Balderas[1], Danelle Seymour[2], Grant R. Cramer[3], Massimo Delledonne[4], Brandon S. Gaut[2] and **Dario Cantu**[1], (1)University of California Davis, Davis, CA, (2)University of California, Irvine, Irvine, CA, (3)Dept of Biochemistry & Molecular Biology, University of Nevada, Reno, Reno, NV, (4)University of Verona, Verona, Italy

Substantial differences in gene content exist between wine grape cultivars. We sequenced and assembled eight genomes using a combination of single molecule real time sequencing, proximity-based ligation with HiC, and optical maps to study structural diversity among grape genomes and its impact on gene content. An average of 70% of each assembly was phased into two haplotypes, revealing heterozygous loci and haplotype structure. An average of 57,884 protein coding genes were predicted in each diploid assembly using a combination of *ab initio* and evidence-based approaches. Variability in gene content between grape cultivars was analyzed by pairwise-comparisons of all predicted transcriptomes. An average of 1,275 genes and 2.4% of the gene space per genotype was not shared between pairs of cultivars. Nearly 51,000 genes were present in all genomes, a total of 3,521 genes was shared by more than two genotypes, but not all, and a total of 349 genes was specific to individual genotypes. Long reads were used to analyze structural variation among the assemblies to determine the impact of structural variation on gene content. In addition to comparing cultivars, structural variations among grape clones were explored by resequencing using short reads of 16 selections of Zinfandel, among which variations could only have arisen during vegetative propagation or meristematic culture. The analyses confirmed that fifteen of the sixteen clones likely arose from a common ancestral plant and suggest that one of the selections is likely not a Zinfandel clone, contradicting earlier marker-based analyses. Overall, 8,179 deletions (median size = 47 bp), 4,039 insertions (median size = 24 bp), 981 duplications (median size ~ 35 kb), and 166 inversions (median size ~70.4 kb) were identified among the selections. Of these, 3,269 structural variants intersected 2,515 genes.

W590: Grape Genome Initiative

Modulation of Grapevine Berry Development and Chemical Composition by Solar Irradiance

Noam Reshef[1], Nurit Agam[1] and **Aaron Fait**[2], (1)Ben Gurion University of the Negev, Blaustein Insts for Desert Research, Sde Boqer campus, Israel, (2)Ben-Gurion University of the Negev, Blaustein Insts for Desert Research, Sde Boqer campus, Israel

In arid regions, where solar irradiance levels are high, insolation may become a limiting factor for plant development. Absorption of excess light energy by plant tissues may increase their temperature, impact metabolic processes, and form reactive oxygen species. Compared to leaves, fruits

have a limited ability to regulate their temperature and dissipate this excess of energy. Their defense mechanisms therefore involve the accumulation of compounds that filter solar irradiance and protect against oxidation and dehydration. As a result, the effect of solar irradiance and fruit temperature on the composition of wine grapes has received much attention in recent years. However the complexity of the metabolic processes involved in grape response to light, and the lack of accurate micrometeorological measurements in field trials, have, thus far, hindered the predictability of grape compositional consequences given a set of meteorological conditions. We unraveled the dynamics of metabolite changes in the grape associated with the modification of solar irradiance conditions. We integrated high- resolution micrometeorological measurements with seasonal, spatial, and diurnal analyses of the grape metabolic profile. We found that grape chemical composition readily responded to changes in solar irradiance, involving the preferential accumulation of sugar alcohols, amino acids, and flavonols, at the expense of sucrose, malate, flavan-3-ols and anthocyanins, respectively. Our results also revealed that the variability in grape composition within a cluster is a function of the variability of solar irradiance and the light-response curve of the specific metabolite group. Last, via the analysis of diurnal metabolite changes, we found an indirect effect of the sun path on fruit sucrose levels, affecting the translocation of assimilates to the fruit. Conversely, the grape insolation pattern directly affected sugar metabolism and the levels of amino acids and phenylpropanoids in the fruit. Our study shows that the spatiotemporal solar regime in a vineyard is a major factor driving spatial variability and diurnal fluctuations in fruit composition. Taken together, we conclude that precise solar irradiance management is a central element of improving fruit quality and homogeneity and mitigating the detrimental consequences of warm and arid conditions.

W591: Grasslands (Lolium Genome Initiative)
Characterization of a *Lolium multiflorum* Diploid Assembly

Dario Copetti, ETH Zurich, Zurich, Switzerland, Joel-Elias Kuon, ETH Zurich, Zürich, Switzerland, Steven Yates, Swiss Federal Institute of Technology, Zuerich, Switzerland, Roland Kölliker, Molecular Plant Breeding, Institute for Agricultural Sciences, ETH Zurich, Zurich, Switzerland and Bruno Struder, ETH Zürich, Zurich, Switzerland

Grasses of the genera *Lolium* and *Festuca* are the main feed sources for a sustainable livestock production due to their high palatability and biomass production. Since decades, their importance for the agriculture of temperate regions led to the development of new varieties through traditional breeding programs. However, newer crop improvement methods such as genomic selection could benefit from a high quality reference genome assembly. In the past, attempts in delivering such a dataset have struggled due to the complexity of the genome and the high heterozygosity of individual genotypes. We sequenced an individual of the *L. multiflorum* (Italian ryegrass) cv. Rabiosa, producing a highly contiguous (N50 of 3 Mb) and complete assembly (97% of the BUSCO gene models). Due to the high heterozygosity of the line, the assembly (4.5 Gb) resulted to be as large as the diploid genome, and presented the sequence of both alleles in separate scaffolds. About ~70,000 gene models were identified, and the repeat content approached 80%. The comparison of a representative allelic region showed an extensive amount of intergenic sequence variation, supporting the high dynamicity of grass genomes. Compared to the available *Lolium* genome assemblies, the Rabiosa assembly improves contiguity by >40-fold, contains both haplotypes of the diploid parent, and assigns most of the sequence to chromosomes.

The availability of a complete and highly-contiguous genome assembly of Italian ryegrass paves the road to the exploitation of the forage crops genetic resources by means of modern genomic platforms.

W592: Grasslands (Lolium Genome Initiative)
Genomic Prediction: Using Multi-Trait Approaches to Improve Predictive Ability for Nutritive Quality Traits in Perennial Ryegrass

Sai Krishna Arojju, Mingshu Cao, Brent Barrett and Marty Faville, AgResearch, Grasslands Research Centre, Palmerston North, New Zealand

Genomic prediction (GP) is a molecular breeding tool which enables selection of superior genotypes from a pool of candidates, enabling acceleration of genetic gain by reducing the length of a breeding cycle, increasing selection intensity, or reducing phenotyping costs. In perennial ryegrass, breeding has primarily focused on improving yield, persistency and quality traits. Over the last two decades, interest in quality traits has grown as research has demonstrated their impact on ruminant performance. Gains have been limited as many nutritive quality traits are often expensive and time-consuming to measure, and affected by environmental or life-cycle variation. This makes them ideal targets for improvement through GP. In perennial ryegrass, GP models have primarily been single-trait; however, recent studies in wheat and sorghum have shown that two or more correlated traits can be used in multi-trait GP models to improve predictive ability for the primary trait. Here, we compare and contrast the relative performance of single-trait and multi-trait GP models, built using phenotypic data from a multi-site trial of a ryegrass training set (n = 517) in combination with genotyping-by-sequencing SNP data. Based on cross-validations, we demonstrated that predictive ability for water soluble carbohydrate level in leaves was improved by 103% and 113% when using digestibility or fibre, respectively, as secondary traits in a multi-trait model.

W593: Grasslands (Lolium Genome Initiative)
Epigenetics of Drought Responses in Barley and Perennial Ryegrass

Marta Malinowska[1], Istvan Nagy[1], Anja Karine Ruud[1], Pernille Bjarup Hansen[1], Simon Svane[2], Cornelis Wagemaker[3], Lene Krusell[4], Birger Eriksen[4], Ahmed Jahoor[5], Jens Jensen[6], Dario Fè[7], Ingo Lenk[8], Morten Greve[8], Lars Bonde Eriksen[9], Kristian Thorup-Kristensen[2], Christian Sig Jensen[8] and Torben Asp[10], (1)Aarhus University, Slagelse, Denmark, (2)Copenhagen University, Taastrup, Denmark, (3)Radboud University, Nijmegen, Netherlands, (4)Sejet Plant Breeding, Horsens, Denmark, (5)Swedish University of Agricultural Sciences, Alnarp, Sweden, (6)Nordic Seed, Odder, Denmark, (7)DLF A/S, Store Heddinge, Denmark, (8)DLF Seeds A/S, Store Heddinge, Denmark, (9)SEGES, Aarhus, Denmark, (10)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark

Epigenetic regulations are of paramount importance for almost every aspect of plant life including cell regulation, differentiation, and transposable elements control. Due to its reversible manner DNA methylation is thought to form a vast source of phenotypic plasticity in

response to environmental stimuli. When genetic diversity within species is low, the epigenetic variation could become an essential resource for optimising plant performance.
Lolium parenne and *Hordeum vulgare*, two model species of vital importance to Danish agriculture were grown in the RadiMax facility designed to identify varieties with improved resource utilisation from deep soil layers. A total of 75 spring barley and 269 ryegrass lines consisting of both breeding material and commercial varieties were exposed to a water-deficiency stress gradient.
Aiming to identify genomic regions with differential methylation states in response to water deficiency stress, Illumina reduced-representation genome sequencing on bisulfite-treated DNA from leaves was carried out. An important goal of the work was to develop models for prediction of quality traits that can integrate multi-omic profiles (including DNA sequence and methylation profile).
DNA methylation was compared across lines and treatments for each species. The analysis allowed detection of context-specific DNA methylation at the single-base resolution, as well as analysis of the extent of methylation at each site. Overall, methylation levels were significantly affected by the treatment applied, and numerous differentially methylated regions were identified, many of which were associated with coding regions. Methylation information was used to predict different quality traits in ryegrass and barley using multi-layered Bayesian regressions. The predictive correlation reached between 0.5 and 0.75. Depending on the trait epigenetic variation accounted for up to 40% of the phenotypic variance.
The results provide insights into the epigenetic status of sites in the barley and ryegrass genomes and suggest its role in abiotic stress adaptation in the agriculturally important species. Understanding the nature of epigenome as a potent source of diversity for agronomical traits may support further strategies to incorporate epigenetics in crop breeding programs. DNA-methylation based trait-prediction models offer increased predictive power, gain in prediction accuracy when combining multi-omics, and it may help to enhance understanding of complex traits when epigenetics is under examination.

W594: Grasslands (Lolium Genome Initiative)
Molecular Breeding for Freezing Tolerance in Lowland Switchgrass
Michael Casler, USDA Dairy Forage Research Center, Madison, WI
Development of switchgrass as a sustainable biofuel crop for temperate climates requires combining two traits that do not exist together within naturally occurring germplasm: late flowering and freezing tolerance. One strategy is to select for freezing tolerance within late-flowering lowland germplasm from subtropical regions of the USA. Genetic variability exists for winter survivorship within many natural populations, but selection pressures vary widely with winter severity and snow cover. Selection based on phenotype has been successful, but depends on having good luck and good timing. Additionally, the frequency of freezing tolerant plants within most lowland populations is well under 1%. This study reports on quantitative trait loci (QTL) discovery in a four-way linkage mapping population and on a broad sampling of natural populations from the southern USA and Mexico, as well as the prognosis of genomic selection for winter survivorship. Six QTL were identified within the four-way linkage mapping population, corresponding to six of the 18 chromosomes. Conversely, the broader sample of 39 diverse populations resulted in 21 potential QTL regions on 11 of 18 chromosomes, four of which corresponded to those QTL identified within the four-way cross. Many of these putative QTL were identified in multiple populations across a broad geographic area. Furthermore, several QTL were proximal to candidate gene homologs implicated in freezing tolerance of other species. Genomic prediction accuracies were population dependent, but expected to be in the range of r=0.77 to 0.87. Genomic selection approaches should be highly effective and greatly reduce the time required to develop adequate winter hardiness within late-flowering lowland populations of switchgrass, as well as reduce the need to rely on appropriate winter weather for phenotypic screening. Furthermore, genomic predictions were moderately effective at predicting future germplasm exploration and collection sites, with accuracy of r=0.62 to 0.80. These results can be used to target future germplasm collection trips that will expand the genetic diversity within gene pools being developed for use in temperate regions of North America.

W595: Grasslands (Lolium Genome Initiative)
Bacterial Endophytes for Sustainable Agriculture
Kerrie Farrar, Aberystwyth University, Aberystwyth, United Kingdom
Miscanthus and other perennial energy crops are being developed for renewable energy and bioproducts as substitutes for fossil fuels. They are largely undomesticated and must produce high annual biomass yields on low-quality land without inputs such as water, fertiliser or pesticides. Bacterial endophytes, which live in plant tissues, interact with aspects of plant growth and development and can play a role in host plant resilience to multiple abiotic and biotic stresses. We have isolated diverse bacterial endophytes from Miscanthus seed and mature plant tissues, as well as from plants growing under abiotic stresses such as salinity and heavy metal contamination. Genomic and functional analyses of the endophyte collections are underway to identify strains for sustainable agriculture applications and to address fundamental questions about plant-microbe relationships.

W596: Host-Microbe Interactions
Olga Francino: Nanopore Sequencing in Microbiome Studies
Olga Francino, Autonomous University of Barcelona, Barcelona, Spain

W597: Host-Microbe Interactions
Genomic and Transcriptomic Characterisation of Campylobacter Resistance in Broilers
Androniki Psifidi[1], Lisa Rothwell[2], Diego Robledo[2], Abi Bremner[2], Andreas Kranis[3], Georgios Banos[4], Richard Bailey[3], Mark S Fife[5], Paul M. Hocking[2], Kellie Watson[2], Pete Kaiser[2], Santiago Avendano[3] and Mark Stevens[2], (1)Royal Veterinary College, University of London, Hatfield, United Kingdom, (2)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (3)Aviagen Ltd, Midlothian, United Kingdom, (4)The Roslin Institute and Scotlands' Rural College, Midlothian, Scotland, (5)The Pirbright Institute, UK, Newbury, United Kingdom
Campylobacter is the leading cause of human foodborne diarrhea. The main source of infection is consumption or handling of contaminated poultry meat. While there is a range of effective biosecurity strategies at farm level, there are no effective vaccines and inhibitors. Moreover,

little is known about the genetic basis of *Campylobacter* colonisation. To identify genomic regions influencing *Campylobacter* load we performed a GWAS study and determined whether eQTLs, allelic-imbalance and significant gene-expression differences are present in broilers with different *Campylobacter* load. Caecal contents were collected from 3,000 broilers and the number of viable Campylobacter per gram was determined. All the birds were genotyped with the 600K SNP-array (Affymetrix). Heritability of the trait was modest (h^2=0.11). GWAS and RHM analyses identified four QTLs on chromosomes 14, 16, 19 and 26. Twenty-three birds were selected for RNA-sequencing based on their genotype and phenotype. RNA was extracted from the caecal tonsils. Three genes located within the QTL region on chromosome 16 were differentially expressed (*BFIV21* and two *BTN-like*). We identified strong cis-QTLs located within the Major-Histocompatibility-Complex (MHC) region on chromosome 16 (the QTL explained 60% of the genetic variance, log allelic-fold-change 2.03), suggesting the presence of cis-acting mutations in *BFIV21* and *BG1* genes. Further, we identified one trans-acting element located within the *CCL4* gene which may regulate 41 genes, including interferon alpha. Based on these results the MHC region seems to play an important role in *Campylobacter* load in chickens, and we have identified three strong candidate genes underlying two load-associated QTLs.

W599: Host-Microbe Interactions

A Systems Approach to Understanding *Phytophthora* infection in Forest Systems

Natalie Graham[1], Preeti Panda[1], Rebecca McDougal[1], Mireia Gomez-Gallego[1], Stan Bellgard[2], Chantal Probst[2], Ian Horner[3], Rosie Bradshaw[4], Simren Brar[4], Laura Raymond[1], Stefan Hill[1], Peter Scott[1], Emily Telfer[1] and Nari Williams[1], (1)Scion, Rotorua, New Zealand, (2)Manaaki Whenua - Landcare Research, Auckland, New Zealand, (3)Plant and Food Research, Havelock North, New Zealand, (4)Massey University, Palmerston North, New Zealand

New Zealand, like many countries in the world, is facing an increasing challenge of dealing with biosecurity pathogens impacting long-term perennial horticulture, plantation forests and natural eco-systems. In recent years, several species within the genus *Phytophthora* have emerged as significant threats to plant systems of importance to New Zealand. Key pathogen species of *Phytophthora pluvialis, P. agathidicida, P. cactorum, P. cinnamomi, P. kernoviae* and *P. multivora* are all having an impact on one or more of these plant hosts, with their interactions contributing further to the stress on tree systems. Taking a genus-wide approach to disease breeding, management and research, we are investigating the underlying mechanisms of pathogenicity, epidemiology and resistance to these pathogens across three contrasting host tree species – radiata pine, kauri and apple.

Integrating genomic, field and controlled inoculation studies we are aiming to inform genomic selection using a range of phenotypic measures to better understand the underlying mechanisms of pathogenicity and resistance, and build breeding and selection programmes for each of these tree species. A summary of the progress across this six year programme will be presented and discussed highlighting the key areas of progress and challenges to date.

W600: Hybridization, heterosis and balancing selection

Heterosis in Maize - Theoretical Models and Empirical Observations

Shawn Kaeppler, Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI, Mike White, University of Wisconsin-Madison, Madison, WI and Natalia de Leon, Department of Agronomy UW, Madison, WI, WA

Heterosis is a phenomenon in which the hybrid of two parent inbreds or populations performs above the mid-parent and, in some cases, substantially above the better parent. Heterosis is manifested across the plant and animal kingdoms and has been recognized as a driving biological force behind the maize hybrid seed industry. The record maize hybrid yield is 542 bu/A, with the hybrid produced from parents each yielding on the order of 1/3 of that amount. Productivity of hybrids has been attributed both to optimized performance as well as enhanced stability across environments. The search for the basis of heterosis has continued for nearly 100 years. A critical question is whether heterosis is a composite outcome of many genes with varying effect size and gene action, or whether there is a gene-agnostic mechanism that boosts performance in a hybrid. The observation that the detection and magnitude of heterosis varies for different traits within hybrids supports that the primary basis is a network of genes as opposed to a general physiological vigor effect. Recent discovery in maize of extensive presence/absence variation and deleterious alleles enriched in low recombination regions of the genome is consistent with the dominance/pseudo-overdominance hypothesis. However, continued observation of hybrid performance above the sum of the inbreds indicates that epistasis and other uncharacterized effects must also play a role. Availability of commercial inbred lines with expired patent and Plant Variety Protection Act certificates provides insights into commercial inbred germplasm in the pre-transgenic era. Hybrid yield performance was certainly a key driver in product development, but plant health, stress tolerance, phenology, and parental inbred performance also weighed heavily in a selection process impacted substantially by the crop biology and logistics of experimental seed production. Rather than an increase generally in heterosis across a broad germplasm base, the result of this multi-sector breeding experiment was divergence of complementary genomic regions driven by evaluation of breeding materials hybridized with a limited number of key founder lines and their highly related progenies.

W601: Hybridization, heterosis and balancing selection

The Role of Balancing Selection and Other Forces in the Evolution of Weedy Rice

Jie Qiu, Washington University in St. Louis, St. Louis, MO, Lingfeng Mao, Zhejiang university, Hangzhou, China, Yongliang Lu, China National Rice Research Institute, Hangzhou, China, Longjiang Fan, Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, Hangzhou, China and Kenneth Olsen, Washington University, St. Louis, MO

De-domestication is a unique evolutionary process by which domesticated crops are converted into 'wild predecessor like' forms. Weedy rice (*Oryza sativa* f. *spontanea*) is an excellent model to dissect the molecular processes underlying de-domestication. Here, we analyse the genomes of 155 weedy and 76 locally cultivated rice accessions from four representative regions in China that were sequenced to an average 18.2× coverage. Phylogenetic and demographic analyses indicate that Chinese weedy rice was de-domesticated independently from cultivated rice and experienced a strong genetic bottleneck. Although evolving from multiple origins, critical genes underlying convergent evolution of different weedy types can be found. Allele frequency analyses suggest that standing variations and new mutations contribute differently to *japonica* and *indica* weedy rice. We identify a Mb-scale genomic region present in weedy rice but not cultivated rice genomes that shows evidence of balancing selection, thereby suggesting that there might be more complexity inherent to the process of de-domestication.

W602: **Hybridization, heterosis and balancing selection**

Genomic Selection of Crossing Partners on Basis of the Expected Mean and Variance - a New Fast Resource-Efficient Breeders Tool

Tanja Osthushenrich[1], Matthias Frisch[2] and Eva Herzog[2], (1)Justus Liebig University of Giessen, Gießen, Germany, (2)Justus Liebig University, Giessen, Germany

The presentation is placed in the research area of plant breeding with the special focus cross prediction and prediction of genetic variance. Predicting the genetic variance of crosses based on genomic data is currently of great interest in the animal and plant community. This is underlined by a range of recent publications in this field (e.g. Mohammadi et al. 2015, Bonk et al. 2016, Lado et al. 2017, Han et al. 2017, Lehermeier et al. 2017). The study presents a resource-efficient tool for breeders to select parental lines within a line or hybrid breeding program to distinguish between the most promising crosses that could be made. Due to the fact of limited field capacity not all crosses can be evaluated in field experiments. Therefore, it is valuable to pre-select the most promising parental combinations based on their expected mean and variance to form the subsequent breeding pool and achieve a higher selection gain. The recombination of elite breeding material often results in crosses with similar means, therefore the variance gains in importance as a selection criterion. We derived formulas to estimate the parameters based on the expected gametic disequilibrium between two loci and can be used for typical mating systems like single seed decent and double haploid lines. The estimation is based on marker effects that are predicted with genome-wide prediction models. Therefore, our approach is an extension of genomic selection that can be obtained with less effort since genomic selection is gaining ground as a tool in breeding programs where genotypic and phenotypic data is routinely available. Our derived formulas were tested on a published maize data set and compared with the simulation approach PopVar. The analytical results for means and variances are highly correlated to simulation results and in times of big data management have a promising speed advantage. From this, we conclude that the presented methods are useful for increasing response to selection of hybrid and line breeding programs by extending genomic selection approaches to the selection of crossing partners, optimizing or reducing resource use for phenotyping and maintaining genetic diversity in breeding programs.

W603: **Hybridization, heterosis and balancing selection**

Statistical Models for Heterosis Analysis

Shizhong Xu, Department of Botany & Plant Sciences, University of California, Riverside, CA and Meiyue Wang, University of California, Riversie, Riverside, CA

Understanding the genetic basis of heterosis is a long-standing challenge in biology. A new statistical model is developed to map quantitative trait loci for heterosis. The new model is different from traditional models of heterosis in three aspects: (1) In addition to the dominance effect, the new model contains a confounded effect of epistasis (dominance x dominance - additive x additive) and does not involve the additive x dominance effect; (2) The new model has a correlated error structure while errors of the traditional model are independent; (3) The new model is more powerful than the traditional model in detection of dominance effect QTL. The conclusions are verified by QTL mapping of heterosis for yield (YD) and 1000-grain weight (KGW) of a hybrid rice population.

W604: **Hybridization, heterosis and balancing selection**

The Relevance of Dominance to Genomic Selection in Breeding Clonally Propagated Plant Species

Christian R. Werner[1], R. Chris Gaynor[1], Dan Sargent[2], Gregor Gorjanc[1] and John M. Hickey[1], (1)University of Edinburgh, The Roslin Institute, Edinburgh, United Kingdom, (2)Driscoll's Genetics Limited, Watsonville, CA

Genomic selection (GS) has the potential to fundamentally revolutionize breeding for quantitative traits in plants. It can enable better use of resources in breeding programs and increases in genetic gain per unit time. We investigated three different strategies to implement GS for a trait representing yield in clonally propagated plant species that exhibit diploid genome structure or diploid-like recombination during meiosis, such as a strawberry. Stochastic simulations were used to evaluate genetic gain, changes in genetic variance and inbreeding over 40 years of breeding. The simulated individuals were highly heterozygous, and different dominance coefficients were applied in order to examine the impact of non-additive genetic effects on the accuracy of GS. The three scenarios using GS were compared to a conventional breeding program based solely on phenotypic selection. Cost effectiveness was analysed by constraining all scenarios to approximately equal annual operating costs. While our results clearly illustrate the usefulness of GS to optimize breeding for clonally propagated varieties, the potential of the different GS strategies examined in this study was highly dependent on the ratio between dominance and additive genetic effect size.

W605: **Hybridization, heterosis and balancing selection**

Deleterious Variants and Genomic Prediction

Peter L. Morrell, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

W606: **Increasing Genetic Gains for Food Security in the Developing World: A New Crop Improvement Innovation Lab and Linkages to a Global Vision**

Introducing a Vision for Crop Improvement Toward Global Food Security

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

W607: **Increasing Genetic Gains for Food Security in the Developing World: A New Crop Improvement Innovation Lab and Linkages to a Global Vision**

Improving Sorghum Adaptation with a Genomics-Enabled Breeding Network

Geoffrey P. Morris, Fanna Maina and Jacques Faye, Department of Agronomy, Kansas State University, Manhattan, KS

W608: **Increasing Genetic Gains for Food Security in the Developing World: A New Crop Improvement Innovation Lab and Linkages to a Global Vision**
Improving Wheat: CGIAR-Innovation Lab Collaboration and Impact
Hans-Joachim Braun, CIMMYT, Mexico D.F., Mexico

W609: **Increasing Genetic Gains for Food Security in the Developing World: A New Crop Improvement Innovation Lab and Linkages to a Global Vision**
Crops to End Hunger
Robert Bertram, US Agency for International Development, Washington, DC

W610: **Increasing Genetic Gains for Food Security in the Developing World: A New Crop Improvement Innovation Lab and Linkages to a Global Vision**
Breeding Insight: Bringing a Unified Informatics and Genomics Platform to Specialty Crops and Animals
Edward S. Buckler, USDA-ARS, Ithaca, NY

W611: **Increasing Genetic Gains for Food Security in the Developing World: A New Crop Improvement Innovation Lab and Linkages to a Global Vision**
Notice of Funding Opportunity: A Crop Improvement Innovation Lab
Daniel Bailey, U.S. Agency for International Development, Washington, DC

W612: **Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**
The Breeding Management System: Progress on the Multi-User Institutional System and Directions Towards a Service Oriented Cloud Application
Mariano Crimi, CIMMYT, Int., Texcoco, EM, Mexico
The need for continuing change is intrinsic to the nature of software. Once a software application is deployed, it invariably faces pressures from its environment (user feedback, emerging technologies, new processes, evolving requirements, etc) and must adapt to adapt to remain relevant. The BMS (Breeding Management System) is an example of such a pathway of adaptation. As a core product of the Integrated Breeding Platform (IBP), the BMS was initially designed as a single-user stand-alone desktop application. In the course of time it has gone through challenging transition to a web based institutional multi user system. This evolution entailed complex changes in the way we managed, conceived, designed and built the product. Our development team had to actively work against the entropy that arises from such profound changes by actively working to minimize complexity through code refactoring and reduction of technical debt and protecting against it through extended unit test coverage.
But software is never finished: As a widely deployed solution, we know that the environment will continue to put evolutionary pressure that the BMS will have to adapt to; if we cannot predict this pressure a priori, we can at least prepare for it by building an architecture that is amenable to change. We see the adoption of the principles of Service Oriented Architecture (SOA) as an opportunity to achieve this by reducing coupling and complexity (which reduces the cost of maintenance and increases adaptability) while also unlocking the potential of leveraging modules and services being built by other teams and initiatives.

W613: **Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**
The Genomic Open-Source Breeding Informatics Initiative: Developing a Global Community of Knowledge Surrounding the Use of Markers in Genomic Selection
Elizabeth Jones[1], Kelly Robbins[2], Star Yanxin Gao[2], Yaw A. Nti-Addae[1], Kate A. Dreher[3], Michael Olsen[4], Susanne Dreisigacker[3], Rajeev K Varshney[5], Abhishek Rathore[6], Joshua N. Cobb[7], Ramil Mauleon[7] and Susan McCouch[2], (1)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (2)Cornell University, Ithaca, NY, (3)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (4)CIMMYT, Nairobi, Kenya, (5)ICRISAT, Hyderabad, India, (6)International Crops Research Institute For The Semi-Arid Tropics, Patancheru, India, (7)International Rice Research Institute, Los Baños, Philippines
The Genomic Open-Source Breeding informatics initiative (GOBii) is a Bill and Melinda Gates funded project with the mission to implement genomic and marker assisted selection as part of routine breeding programs for staple crops at CGIAR centers. We believe that much of the gains that have been achieved by major ag-biotech companies can also be achieved in these centers through adopting data management systems and bioinformatics pipelines that aid breeding decisions. Our challenge is to implement genomics data management and connect to breeding data management and analysis tools being developed as part of diverse projects and within different organizational structures. To achieve our goal, we have employed a global team of data curators, developers, molecular breeders and system administrators based at Cornell University, The Boyce Thompson Institute and at each of our collaborating CGIAR centers; CIMMYT, ICRISAT and IRRI. This team have experience and skill sets that cross multiple data management and curation projects with backgrounds in industry and academia and together can collaborate to find best solutions for use cases gathered in their own environments. We are partnering synergistically with adjacent data management and genotyping projects to prevent feature redundancy and promote the use of data management systems, and we are aligned with the Excellence in Breeding Program to ensure united approaches and goals. Our CGIAR center partners have now become the experts in managing and analyzing genomics and genotyping data and are training their own communities in using these systems. Together we have built Genomic Selection pipelines in Galaxy, data QC tools with Diversity Arrays Technologies and marker-assisted backcrossing, forward breeding and pedigree verification with the James Hutton Institute, based on use cases collected by our CGIAR molecular breeders. Together we are building a global community of knowledge surrounding best practices for implementing marker-assisted and genomic selection at CGIAR centers.

W614: **Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**
BMS Implementation across Multiple Crops: Progress and Challenges
Genevieve Thompson, Agricultural Research Council, Pretoria, South Africa

W615: **Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**
Ongoing BMS Adoption for Three Major Breeding Programs at EMBRAPA
Flavio Breseghello[1], Helton S. Pereira[1], João Luis da Silva Filho[1] and Julian Pietragalla[2], (1)Embrapa, Santo Antônio de Goiás, Brazil, (2)Integrated Breeding Platform, Texcoco, Mexico
Embrapa is adopting BMS as the information system for its rice, common bean and cotton breeding programs. Those are long-standing national programs, with breeding locations and testing fields in different sites throughout the country. Hundreds of crosses are done each year, in different subprograms and breeding pools. Despite its size and duration, until 2018 those programs relied entirely on spreadsheets for data handling. BMS deployment is in progress for the season 18/19 and the first benefits have been already perceived, especially in germplasm management and trial preparation. After full adoption, we expect great operational gains in all phases of breeding, from season planning until statistical analyses, with less time spent in routine operations. In rice, we are using BMS to turn a cross book, with more than 20,000 crosses, into pedigrees with up to ten generations. Genealogies are further extended by connecting with the BMS international rice germplasm database. From this information, we should be able to compute kinship coefficients for all the germplasm. We also have phenotypic data from 3,000+ replicated field trials, which will be uploaded to BMS. Those data will allow "head-to-head' queries and computing breeding values through mixed models. The most relevant parental lines are being genotyped and SNP data will be uploaded in a second phase. With the integration of all those sources of information, we expect to have a better understanding of our germplasm, accelerate genetic gain and achieve a seamless connection between plant breeding and genetic research.

W616: **Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**
Simplifying the Implementation of Genotype-Based Breeding through Automatic SNP Scoring
Michel Ragot, DeltaGee / Nouvelle France Genetics, Minnetonka, MN

W617: **Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding**
Accelerating Genetic Gains in Crop Breeding in Sub-Saharan Africa and Southern Asia by Improving Access and Affordability to High Throughput Genotyping
Eng Hwa Ng[1], Rajaguru Bohar[2], Milcah Kigoni[3], Rajeev K Varshney[2], Michael Olsen[4], Patrik Stolt[5] and Petra Vanroggen[5], (1)Excellence in Breeding, Los Baños, Philippines, (2)ICRISAT, Hyderabad, India, (3)ICRISAT, Nairobi, Kenya, (4)CIMMYT, Nairobi, Kenya, (5)Intertek, Alnarp, Sweden
High Throughput Genotyping (**HTPG**) Project is a new initiative funded by Bill and Melinda Gates Foundation to provide low cost genotyping to all CGIAR and NARs users via outsourcing. The project is led by ICRISAT and it is one of the flagship projects under module 3 (genotyping and sequencing) of the newly conceived Excellence in Breeding Platform (**EiB**). Under the HTPG service agreement, more than 40 public institutions globally (16+ species) are listed as key users for the SNPLine genotyping platform. Primary objective of the project is to broker access of the latest genotyping platform to a wide user group at a reduced cost via sample aggregation. For $1.50 to $2.00 per sample, user can have DNA extraction, DNA QC, 10 KASP marker genotyping and post-genotyping data curation within a two weeks turnaround time. The project is also working closely with various stakeholders within the EiB platform to integrate genotyping results with breeding management systems and molecular database. The aim is to enable timely and seamless decision making in breeding programs to accelerate the rate of genetic gain. HTPG project is unique in the sense that it brings together all the latest developments in molecular markers globally under one unified platform and enables timely access to all parties, private sectors included, to facilitate greater interdisciplinary collaboration for a stronger breeding networks. In 2019, HTPG project will revise its existing pricing structure to provide more competitive options for users seeking greater flexibility in marker selection and at a higher density than the existing 10 SNPs setup, specifically targeting applications such as QC and background recovery in breeding programs. Building on the current HTPG mode of operation, a higher density genotyping service agreement (1000- 5000 density) is under negotiation and expected to become available in 2019.

W618: **International Cotton Genome Initiative (ICGI)**
The Gossypol Biosynthetic Enzymes are Diversified Phylogenetically and Dispersed in Cotton Genome
Xiao-Ya Chen, National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai, China

W619: **International Cotton Genome Initiative (ICGI)**
Reference-Grade Genome Assemblies Facilitate Exploitation of Favorable Genetic Variations underlying the Development of Superior Cotton Fibers
Xianlong Zhang, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

W620: **International Cotton Genome Initiative (ICGI)**
Targeted Systems Analysis of Immune Receptor Genes in Cotton Response to Fusarium Wilt
Ping He, Texas A&M University, College Station, TX

W621: **International Cotton Genome Initiative (ICGI)**

New Roles of Transcriptional and Translational Regulation in Early Stages of Cotton Fiber Development

Atsumi Ando[1], Ryan Kirkbride[1], Qingxin Song[1], David M. Stelly[2], Don C. Jones[3] and Z. Jeffrey Chen[1], (1)The University of Texas at Austin, Austin, TX, (2)Texas A&M University, College Station, TX, (3)Cotton Incorporated, Cary, NC

Cotton is the largest renewable source of textile fiber. Approximately 25% of protodermal cells on the cotton ovule surface (seed coat) differentiate into fiber cells. Each fiber is a singular cell, which undergoes rapid cell elongation and cellulose biosynthesis to make it one of the longest types of plant cell, reaching up to six centimeters in length. Rapid cell growth is associated with dynamic molecular and physiological changes, but the mechanism for fiber cell initiation is poorly understood. Here we report transcriptome dynamics between fiber and epidermal cells during fiber cell initiation in two cultivated cotton species and a fiber-less mutant. RNA-seq analysis was performed using the fiber and epidermal cells that were isolated by modified Laser-Capture Microdissection (LCM) techniques. By comparing thousands of differentially expressed genes, we found enrichment of the highly expressed genes in Gene Ontology (GO) groups of ribosomal biosynthesis in the fiber cells and of DNA replication and chromatin in the epidermal cells. Furthermore, *in vitro* cultured cotton ovules, when exposing to chemical inhibitors for ribosome biosynthesis and cell cycle regulation, developed shorter and longer fiber cells, respectively. Our data suggest that rapid cell divisions during the differentiation of epidermal cells require DNA replication and chromatin activities, while in the fiber cells, the cell cycle arrest could induce ribosome biosynthesis, which promotes protein synthesis that is required for rapid cell elongation and expansion.

W622: **International Cotton Genome Initiative (ICGI)**

Toward the Construction of Chromosome Introgression Population from *Gossypium arboreum* in *Gossypium hirsutum*

Baoliang Zhou, Nanjing Agricultural University, Nanjing, China

Gossypium arboreum, a diploid cultivated cotton species (2n = 26, AA) native to Asia, possesses invaluable characteristics unavailable in the tetraploid cultivated cotton gene pool, such as resistance to pests (*Apolygus lucorum* and *Rotylenchulus reniformis*) and diseases (caused by *Verticillium dahliae*, *Fusarium oxysporum vasinfectum* and cotton leaf curl virus) and tolerance to abiotic stresses (drought and heat). However, it is quite difficult to transfer favorable traits into Upland cotton through conventional methods due to the cross-incompatibility of *G. hirsutum* (2n = 52, AADD) and *G. arboreum,* resulting to very a few genetic resources of the diploid-cultivated species successfully used in cotton breeding for a long time. Here, using *G. hirsutum* acc TM-1 as recipient parent and *G. arboreum* as donor parent, we improved an embryo rescue technique to overcome the cross-incompatibility between these two parents for developing chromosome introgression lines and transferring favorable genes from *G. arboreum* into *G. hirsutum.* Eight putative hybrids were successfully obtained and were subsequently treated with colchicine solution to double their chromosomes. The results demonstrate that four putative hybrid plants were successfully chromosome-doubled and become amphiploid (2n = 78, AADDAA), which were confirmed by cytological observation and self-fertilization. The synthetic amphiploid was employed to backcross with TM-1 and produced 102 BC_1F_1 individuals (2n = 65, AADDA). A total of 499 ILs, with an average segment length of 6.27 Mb, spanning 1185.54 Mb and covering 80.67% of the genome from *G. arboreum*. These ILs contain numerous variants, such as boll sizes, boll numbers, leaf shapes, flower petal spots and colours. These ILs can be used as tools for mining, isolation, characterization and cloning of *G. arboreum*-specific desirable genes for future cotton breeding.

W623: **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

David D. Fang, Cotton Fiber Biosciences Research Unit, USDA-ARS-SRRC, New Orleans, LA

W624: **International Goat Genome Consortium**

Introduction

Gwenola Tosser, INRA GenPhySE, CASTANET-TOLOSAN, France

W625: **International Goat Genome Consortium**

Contiguity of the Goat Genome Assembly Enables Comparative Analyses and Gene Discovery within the Repetitive Immune-Related Regions

John C. Schwartz, The Pirbright Institute, Woking, United Kingdom

Highly repetitive regions of the genome are notoriously difficult to assemble in whole genome sequencing attempts resulting in gene complexes that are either heavily disrupted with many sequence gaps, incorrectly assembled, or missing a large amount of sequence. Many such regions encode genes that are integral to immune system functions, are under strong selective pressures, and as such among the most quickly evolving regions of the genome. In the recent high-quality goat assembly (ARS1) we have characterized six immune-related gene clusters that are largely intact on single contigs: the three antibody loci, the natural killer complex (NKC), and the leukocyte receptor complex (LRC). Comparative analyses have identified a novel immunoglobulin-like receptor gene family across non-primate and non-rodent species and the independent expansion of the killer cell immunoglobulin-like receptors (KIR) in sheep and goats compared to cattle. These accurate gene models have enabled us to examine transcription revealing a diverse gene expression profile across different immune and non-immune tissue compartments. These results will help inform future species-specific studies that investigate the role of these receptors in livestock health and disease.

W626: **International Goat Genome Consortium**

Accuracy of Direct Imputation of 50k SNP-Chip Data to Sequence Level Genotypes in French Dairy Goats

Estelle Talouarn, INRA GENPHYSE, CASTANET TOLOSAN, France

Goats were domesticated 10,500 years ago with the aim of supplying milk, meat and fibers. Since then, breeds have specialized and adapted to their local environment developing specific genetic profiles. The VarGoats project is an international resequencing program which aims at covering at best the genetic diversity of the *Capra* species. To date the sequence data available include 16 wild types and 578 *Capra hircus* of 65 different breeds. Variant calling led to the identification of a total of 105,772,894 variants on the 29 autosomal chromosomes. For French Alpine and Saanen individuals the concordance with 50k genotypes was checked. Quality checks were applied to sequence variants using various

indicators: QUAL, GQ, DP, MAF and position. Mean concordance rate was 97.77% and ranged from 83.09% to 99.98%. Imputation was tested on the 27,979,856 filtered variants using FImpute software. Pedigree was provided and imputation was performed in a within-breed leave-one-out scenario. Imputation quality was checked for 4 individuals on all chromosomes and for every sequenced animal on chromosome 29. Mean concordance rate for chromosome 29 ranged from 88.44 to 92.82% and from 92.29 to 94.95% in Alpine and Saanen breeds respectively. On all chromosomes average correlation between true and imputed sequence were 0.86 and 0.83 in Alpine and Saanen respectively. Imputation will be applied to a population of 3,618 French Alpine and Saanen goats genotyped with the CaprineSNP50 BeadChip. The next step will be to compare association studies using 50K data or SNPs imputed from the sequence data. This research provides insights on how to implement a solid quality check, and imputation that will ensure the quality of subsequent analyses.

W627: **International Goat Genome Consortium**
Completion and Haplotype Phasing of the Goat Genome
Benjamin D. Rosen, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

W628: **International Goat Genome Consortium**
Genome-Wide Characterization of Selection Signatures in Ugandan Goat Breeds
Robert Onzima, Animal Breeding and Genomics Centre, Wageningen University and Research (WUR), Wageningen, Netherlands

W629: **International Goat Genome Consortium**
Genome-Wide Scan of Copy Number Variations (CNVs) in Ethiopian and Asian Indigenous Goat Populations
Getinet Mekuriaw Tarekegn[1,2], Yuexing Wang[3,4], Yuan Chai[3,5], Bin Liu[5,6], Jian-Lin Han[7], Joram M. Mwacharo[8], Dessie Tadelle[9], Appolinaire Djikeng[10], Mwai Okeyo[11], Josephine Birungi[12] and Wenguang Zhang[3], (1)Department of Animal Production and Technology, Bahir Dar University, Bahir Dar, Ethiopia, (2)Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, (3)College of Animal Sciences, Inner Mongolia Agricultural University, Hohhot, China, (4)Peking Jabrehoo Med Tech., Ltd., Beijing, China, (5)Key Laboratory of Animal Science, Breeding and Reproduction of the Inner Mongolia Autonomous Region, Hohhot, China, (6)Nei Mongol BioNew Technology Co.Ltd, Hohhot, China, (7)Chinese Academy of Agricultural Sciences-International Livestock Research Institute Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Sciences, Beijing, China, (8)International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia, (9)International Livestock Research Institute, Addis Ababa, Ethiopia, (10)Centre for Tropical Livestock Genetics and Health, The University of Edinburgh, Edinburgh, United Kingdom, (11)International Livestock Research Institute, Nairobi, Kenya, (12)BecA-ILRI Hub, Nairobi, Kenya
Copy number variations are major sources of genomic structural variations which affect phenotype of the animals. We carried out genome-wide CNVs analysis of 1302 goats collected from Ethiopia, China, Pakistan and Bangladesh aiming to detect CNVs in the genome. The study animals were grouped based on the admixture output into six: Ethiopian, Cashmere, Cashmere-Ibex hybrid, Bengal and Pakistan goat populations. We identified 4666 CNVs which overlapped into 446 CNVRs in all groups. The overlapping regions spanned on 110.03 Mb (3.76%) of the genome. The highest CNVs and CNVRs were observed in Ethiopian goats (2516 and 147, respectively) and the lowest number of CNVs (184) and CNVRs (30) in Bengal and Cashmere goats, respectively. Unlike the trend observed in CNVs, most of the CNVRs detected in the groups were loss CNVRs except for Cashmere. In the combined data set, 198 CNVRs were identified which spanned throughout the autosomal chromosomes except chromosome 28. There was no overlapping CNVs in chromosome 28. The highest number of CNVRs were detected in chromosome 1&6 (15 CNVRs each). The enrichment analysis of 198 CNVRs revealed 877 annotated genes involved in various biological pathways which include responses to protein stimulus, nutrients, extracellular stimulus, organic substances, defense response, etc. Genes identified in three CNVRs of all goat groups involve in olfactory receptor, ribosomal and Zinc Finger Protein pathways. Overall, this study provides comprehensive insights on distribution of the CNVs in the genome and their functions. This may lead to include the variants in genomic selection strategies.

W630: **International Goat Genome Consortium**
The African Goat Improvement Network Update - Then, Now, and Future Directions
M.J. Woodward-Greene[1], Benjamin D. Rosen[2], Keabetswe Tebogo Ncube[3], Gábor Mészáros[4], Henry Aaron Mulindwa[5], Helen N. Nakimbugwe[6], Wilson Nandolo[7], Wilson Sostern Kaumbata[7], Brian L. Sayre[8], Steven G. Schroeder[2], Gordon Spangler[2], Timothy Gondwe[7], Tad S. Sonstegard[9], Jeffrey T. Silverstein[10], Paul Boettcher[11], Liveness J Banda[7], Max F. Rothschild[12], Farai C. Muchadeyi[3], Johann Soelkner[13] and Curtis P. VanTassell[2], (1)National Agricultural Library, USDA-ARS, Beltsville, MD, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)Agricultural Research Council, Pretoria, South Africa, (4)Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria, (5)National Agricultural Research Organization, National Livestock Resources Institute, Tororo, Uganda, (6)Ministry of Agriculture Irrigation and Water Development, ., Uganda, (7)Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi, (8)Virginia State University, Petersburg, VA, (9)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN, (10)USDA-Agricultural Research Service, Beltsville, MD, (11)Animal Genetics Resources Branch, Animal Production Division, Food and Agriculture Organization of the UN (FAO), Rome, Italy, (12)Department of Animal Science, Iowa State University, Ames, IA, (13)University of Natural Resources and Life Sciences, Vienna, Vienna, Austria
The African Goat Improvement Network (AGIN), a product of the USAID - USDA Feed the Future Livestock Improvement Project, is a collaborative group of scientists focused on improving African indigenous goat production since 2011 with a long-term goal of developing cost-effective genomics strategies to improve productivity without sacrificing adaptation. AGIN also supports use and development of mobile data recording and transfer to collect and track phenotypes. AGIN engages farmers directly through Community-Based Breeding Programs (CBBPs) for their input to identify important production traits and production system challenges, as well as to collect data on phenotypes and production

results. AGIN experiences demonstrate that a CBBP works when communication with the farmers is effective and engages them not only as full partners, but as leaders in their CBBP. A stable and supportive government unopposed to non-nucleus breeding programs is also essential. Impact of AGIN is seen in the development of improved goat herds in communities, as earned doctoral degrees conferred in Africa and elsewhere, in participation of members in international research collaborations, by development of photo-based tools and methods for phenotype data collection, and, most notably, in the development of a cheap, rapid, and accurate method to build a goat reference genome. A brief AGIN update, highlighting successes, products, and future plans for long-term independence and sustainability will be shared, including a demo of the recently enhanced ADAPTMap digital phenotyping software.

W631: International Goat Genome Consortium
Conclusion
Gwenola Tosser, INRA GenPhySE, CASTANET-TOLOSAN, France

W632: International Phytomedomics and Nutriomics Consortium (ICPN) 1
Plant Genomics and Breeding for Health and Nutrition Security
Chittaranjan Kole, Department of Atomic Energy, New Delhi, India

W633: International Phytomedomics and Nutriomics Consortium (ICPN) 1
Draft Genome Sequence of Bitter Gourd (*Momordica charantia*) as Vegetable Crop and Medicinal Plant
Hideo Matsumura, Shinshu University, Ueda, Japan

Bitter gourd (*Momordica charantia*) is one of Cucurbitaceae species and known as an important vegetable in tropical and sub-tropical regions, including Asia, Africa and South America. Also, bitter gourd is a medicinal plant, which has been used in traditional Chinese medicine. We analyzed the draft genome sequence of a monoecious bitter gourd inbred line in Okinawa, OHB3-1. Through Illumina sequencing of paired-end and mate-pair libraries (37Gb in total) and *de novo* assembly, scaffolds of 285.5 Mb in length were generated. In this draft genome sequence, 45,859 protein-coding gene loci were identified, and transposable elements accounted for 15.3% of the whole genome. Based on sequences of conserved genes, synteny mapping and phylogenetic analysis were carried out among biter gourd, watermelon (*Citrullus lanatus*), cucumber (*Cucumis sativus*) and melon (*Cucumis melo*). According this comparative analysis, genes related to medicinal, like putative trypsin-inhibitor and ribosome-inactivating proteins, were uniquely multiplicated in bitter gourd genome. Additionally, 1,507 polymorphic loci were identified by RAD-seq analysis between two bitter gourd lines, and genotyping of these marker loci in their F_2 progeny resulted in a linkage map, comprising 11 linkage groups equivalent to the chromosome number in bitter gourd. By anchoring RAD tag markers, 255 scaffolds were assigned to the linkage map. Recently, we also analyzed bitter gourd genome sequence using PacBio sequencing technology and an improved version of assembly could be obtained. These draft genome sequences and linkage map of bitter gourd will allow to elucidating its pharmacological effect and promoting its genetic improvement.

W634: International Phytomedomics and Nutriomics Consortium (ICPN) 1
Systematic Analysis of microRNAs and Phased Small Interfering RNAs in *Salvia miltiorrhiza*
Hong Zhou, Jiang Li, Caili Li and Shanfa Lu, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

MicroRNAs (miRNAs) and phased small interfering RNAs (phasiRNAs) are two primary categories of endogenous small non-coding RNAs in plants. They play critical functions in multiple biological processes, including development, secondary metabolism, diseases and stress responses. However, genome-wide characterization of small non-coding RNAs has not been carried out in *Salvia miltiorrhiza* Bunge, a well-known traditional Chinese medicinal plant with significant medicinal and economic value. In order to comprehensively identify and characterize miRNAs and phasiRNAs in *S. miltiorrhiza*, six small RNA libraries from mature roots, young roots, stems, mature leaves, young leaves and flowers, and one degradome library from the mixed tissues were constructed. High-throughput small RNA sequencing generated approximately 70 million clean reads from the sRNA libraries and 10 million clean reads from the degradome library. Genome-wide analysis of sRNAs in *S. miltiorrhiza* identified 155 conserved and 48 novel miRNAs, which are members of 82 families. Analysis of degradome data revealed 156 targets for conserved miRNAs, of which, 75 are conserved among various plant species, whereas the others are non-conserved. Further experimental validation showed that *SmNINV3* and *SmNINV4*, two non-conserved miRNA targets, were indeed cleaved by smi-miR399. It suggests that conserved miRNAs may play species-specific regulatory roles in *S. miltiorrhiza*. In addition to the targets for conserved miRNAs, 21 genes were found to be regulated by novel miRNAs. It includes three miRN1-targeted *mTERFs*, two miRN2-targeted *MADSs*, and two miRN3-taregted *PPRs*. The regulation of these targets by novel miRNAs further reveals functional specificity of miRNAs in *S. miltiorrhiza*. Cleavage of *mTERFs*, *PPRs* and other twelve targets by miRN1, miRN3 and members of five conserved miRNA families, including *MIR390*, *MIR828*, *MIR393*, *MIR482/2118* and *MIR1510*, may trigger the generation of phasiRNAs. The phasiRNA-generating pathways of miR390-*TAS3*-tasiARFs, miR828-*TAS4*-siRNA81(-), miR393-*F-BOX*-phasiRNA and miR482/2118/1510-*NB-LRR*-phasiRNA were conserved in various plant species, whereas the miRN1-*mTERF*-phasiRNA pathway could be *S. miltiorrhiza* or lineage-specific. It suggests that miRNAs, phasiRNAs and target genes form an intricate regulatory network in *S. miltiorrhiza*. The results provide a genome-wide view of *S. miltiorrhiza* miRNAs and phasiRNAs.

W635: International Phytomedomics and Nutriomics Consortium (ICPN) 1
Herbgenomics: The Development and Future of TCM-Omics
Jiang Xu, China Academy of Chinese Medical Sciences, BeiJing, China

Traditional herbal medicines, such as plant- and fungi-based remedies, have beenused for more than 5,000 years. However, the genetic background, the agricultural traits, and the medicinal quality of most traditional herbs are poorly understood. With rapid advances in high throughput sequencing technologies and greatly reduced costs, a new discipline called “herbgenomics” has emerged. Researchers are now systematically categorizing medicinal herbs by sequencing, assembling, and annotating their genomes, and by analyzing their genes’ functions.

Genomic information, together with transcriptomic, proteomic, and metabolomic data, can therefore be used to predict secondary metabolite biosynthetic pathways and their regulation, triggering a revolution in discovery-based research aiming to understand the genetics and biology of herbs. Herbgenomics provides an effective platform to support the chemical and biological analyses of complex herbal products that may contain more than one active component. Therefore, it is now being applied to many areas of herbrelated biological research to help understand the quality of traditional medicines and for molecular herb identification through the establishment of an herbal gene bank. Moreover, functional herbal genomics can contribute to model herb research platforms, geoherbal research, genomics-assisted herb breeding, and herbal synthetic biology, all of which are important for securing the sourcing of TCMs and their active compounds in the future.

W636: International Phytomedomics and Nutriomics Consortium (ICPN) 1

Toward Understanding the Role of Polyphenol Oxidases in Phenolic Acid Biosynthesis in *Salvia miltiorrhiza*

XiaoXiao Qiu[1], Caili Li[1], Dongqiao Li[2] and Shanfa Lu[1], (1)Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, (2)Institute of Medicinal Plant Development, Beijing, China

Polyphenol oxidases (PPOs) are a class of copper-containing oxidases that oxidize phenols to quinones. Previous studies indicated that PPOs are involved in phenolic acid biosynthesis in *Salvia miltiorrhiza*, a well-known material of traditional Chinese medicine and an emerging model system for medicinal plant biology. Phenolic acids, such as salvianolic acid A and B, are a class of pharmocologically active compounds in *S. miltiorrhiza*. However, the molecular mechanism of PPO-involved phenolic acid biosynthesis has not been fully elucidated. Here we report the identification and characterization of 19 *S. miltiorrhiza PPO* (*SmPPO*) genes. The *PPO* gene family in *S. miltiorrhiza* is the largest compared with that in other plant species with whole genome sequence available. Analysis of gene structures and protein conserved domains and motifs showed that *SmPPOs* are very similar to the *PPOs* in other plant species. It is consistent with that all PPOs are capable of oxidizing phenols to quinones. *SmPPOs* showed differential expression in different plant tissues and the majority were responsive to MeJA, yeast extract and Ag+ treatments. It suggests functional redundancy and divergence of SmPPOs. In addition, eight *SmPPOs* were predominantly expressed in the phloem and xylem of *S. miltiorrhiza* roots, where the pharmacologically active salvianolic acids were actively biosynthesized. It indicates the involvement of SmPPOs in salvianolic acid biosynthesis. Analysis of small RNA sequences and degradome data of *S. miltiorrhiza* revealed a new miRNA family, termed *MIR12112*, which specifically targets PPO transcripts for cleavage. Systematic analysis of whole genome sequences and transcriptome data allowed us to identify 76 *MIR12112* precursors from Lamiales plants. It suggests that *MIR12112s* are specific to Lamiales plants, which usually produce high levels of natural phenolic compounds. To further elucidate the functions of SmPPOs and the regulatory roles of Smi-miR12112 in phenolic acid biosynthesis, we are producing transgenic plants with overexpressed or silenced *PPO* and *MIR12112* genes.

W637: International Phytomedomics and Nutriomics Consortium (ICPN) 1

Reaping the Benefits of Nutriomics for Human Health - Chinese Cabbage

SuRyun Choi, Jana Jeevan Rameneni, Sushil Satish Chhapekar, Seongmin Hong and **Yong Pyo Lim**, Chungnam National University, Daejeon, South Korea

Chinese cabbage (*B. rapa*) is considered as an essential nutri-vegetable with an abundance of health benefiting compounds. The functional genomics and molecular breeding researches have been rapidly expanded to understand metabolic pathways and related gene functions on important functional compounds. Advancement of nutriomics related to functional foods has become an imperative subject for global human health. We analyzed and identified the accessions which are enriched in functional compounds such as glucosinolates, anthocyanins, vitamins, β-carotene, total sugars, lutein, flavonol, Fe, Ca, etc. Further, we have generated double haploid (DH) lines through microspore culture to investigate various aspects of nutrigenetics and nutriomics of these inbred lines. For glucosinolates, we have performed a conventional QTL analysis using F2/3 mapping population of *B. rapa* combined with genome-wide association approach by using natural population to identify the genomic region and genes regulating glucosinolate biosynthesis in *B. rapa* crops. The advanced analysis is being carried out to explore the identified candidate genes related to glucosinolates enrichment. Similarly anthocyanin, the predominant flavonoids in red/purple crops, were tested for its inhibitory effects in cultured endothelial cells and hyperlipidemic apolipoprotein E-deficient mice using anthocyanin-rich extract from red Chinese cabbage and found to reduce the risk of vascular inflammatory diseases. Furthermore, we generated mapping population from red and green Chinese cabbage and QTL mapping coupled with genotyping by sequencing (GBS) approach was used to identify genomic loci associated for anthocyanin biosynthesis. The transcriptome sequencing of both parents along with QTL mapping revealed a potential candidate genes/transcription factors regulating anthocyanin biosynthesis. Our overall results will be a foundation for future studies on nutrition breeding of enriched varieties of *B. rapa* and their subspecies for human health.

W638: International Phytomedomics and Nutriomics Consortium (ICPN) 1

Understanding and Modifying the Nutritional and Oil Quality Architecture to Breed Nutrition-Rich and High Oil Quality Peanuts

Manish K Pandey[1], Hari Kishan Sudini[1], Arun Kumar Pandey[1], Yaduru Shasidhar[1], Jake Fountain[2], Surendra S Manohar[1], Spurthi N Nayak[3], Pooja Soni[1], Rakesh Kumar[1], Murali T Variath[1], Sarita Pandey[1], Pasupuleti Janila[1], Sandip K Bera[4], Boshou Liao[5], Baozhu Guo[6] and Rajeev K Varshney[1], (1)ICRISAT, Hyderabad, India, (2)University of Georgia, Tifton, GA, (3)University of Agricultural Sciences-Dharwad, Dharwad, India, (4)ICAR-Directorate of Groundnut Research (DGR), Junagadh, India, (5)Oil Crops Research Institute, Wuhan, China, (6)USDA-ARS, Crop Protection and Management Research Unit, Tifton, GA

Peanut or groundnut (*Arachis hypogaea*), grown and consumed in several Asian and African countries in addition to Americas, plays an important role in providing daily nutritional requirement for large population of the world. Aflatoxin contamination and allergens are the major quality and food safety concerns across globe which adversely impact the global peanut trade and commerce. On the other hand, high oleic acid is an industry preferred trait for imparting increased shelf life to peanut-based products. Through precise phenotyping, genomics, transcriptomics and molecular breeding approaches, we are developing better understanding of these traits, conducting trait mapping and candidate gene discovery, and deploying molecular breeding for developing improved peanut varieties. For example, transcriptome analysis have identified several important candidate genes and pathways for three different types of resistance mechanisms of aflatoxin contamination namely *in vitro* seed colonization (IVSC), pre-harvest aflatoxin contamination (PAC), and aflatoxin production (AP). Further, genetic analysis of multi-parent

advanced generation intercross (MAGIC) and genome-wide association study (GWAS) on a diverse association mapping panel are likely to provide associated genomic regions and candidate genes for aflatoxin contamination. Development and deployment of precise ELISA-based methods for quantifying five major and important peanut allergens (Ara h 1, Ara h 2, Ara h 3, Ara h 6 and Ara h 8) have led to the identification of several hypoallergenic lines. Subsequently sequence/GWAS analysis is likely to identify the alleles responsible for making peanut, hypo or hyper allergenic. Allele-specific genetic markers were successfully deployed for developing several high oleic molecular breeding lines in multiple genetic backgrounds. Many of these lines are in final year of testing in India and are most likely to get released in 2019 for cultivation. Identification and development of improved peanut lines with combination of these nutritionally important and oil quality traits are likely to enhance the consumption and international trade of peanut.

W640: International Phytomedomics and Nutriomics Consortium (ICPN) 2
TBA
Kathleen Hefferon, University of Toronto, Toronto, ON, Canada

W641: International Phytomedomics and Nutriomics Consortium (ICPN) 2
Sequence Variation within *DcOr* is Associated with Carotenoid Accumulation in Carrot
Philipp W. Simon[1], Kevin Michael Coe[2], Shelby Ellison[3], Douglas Senalik[3] and Julie Dawson[4], (1)USDA, ARS and University of Wisconsin, Madison, Madison, WI, (2)University of Wisconsin-Madison, Madison, WI, (3)USDA, ARS, Madison, WI, (4)University of Wisconsin, Madison, WI
Carrot (*Daucus carota*) is one of the richest sources of the vitamin A precursor b-carotene in the human diet. Two genes, *Y* and Y_2 have been previously identified to be responsible for the majority of carotenoid accumulation in carrot roots. *Y* conditions all carotenoid accumulation in carrot roots, and one allele present in orange and yellow carrots harbors a 212 nt insertion in the gene. Y_2 is known to condition the accumulation of b- and a-carotene in carrot roots. The identity of Y_2 is unknown, but Y_2 has been fine-mapped to a 650-kb region on Chromosome 7. Recently, the *Or* gene was identified by a genome-wide association study (GWAS) to also be significantly associated with carotenoid accumulation in carrots roots. Molecular studies of *Or* in other plants, such as Arabidopsis, melon, and cauliflower have revealed mutations that result in increased sequestration of b-carotene in tissues that are normally non-photosynthetic. During plant growth, *Or* stimulates chromoplast biogenesis, thereby creating a sink for carotenoids to accumulate. Additionally, OR has been shown to stabilize PHYTOENE SYNTHASE (PSY), the rate limited enzyme in the carotenoid biosynthetic pathway. It is our hypothesis that during carrot domestication, a mutated *Or* allele was selected, alongside *Y* and Y_2, for its unique ability to increase carotenoid accumulation in root tissue. Causal mutations in *Or* are currently being identified using a panel of wild and domesticated resequenced plant introductions (PIs). Additionally, patterns of *Or* expression will be analyzed in a mapping population of carrots identified to be fixed for *Y* and Y_2 but still segregating for orange and yellow root color.

W642: International Phytomedomics and Nutriomics Consortium (ICPN) 2
Whole Genome Sequencing in Two Medical Plants, *Atractylodes lancea* and *Ephedra sinica*
Kenta Shirasawa[1], Andrea Ghelfi[1], Hajime Hiyama[2], Takahiro Tsusaka[2], Hideki Hirakawa[1] and **Sachiko Isobe**[1], (1)Kazusa DNA Research Institute, Kisarazu, Japan, (2)Tsumura & Co., Inashiki, Japan
Atractylodes lancea and *Ephedra sinica* has been used as medical plants especially in East Asia. In Japan, it is used as Kampo medicine, and most of materials has been derived from the plant naturally grown in China. However, because of depletion of natural resources, breeding and field cultivation has been just stared recently based of gemplasm collections kept in Tsumura & Co. The genetic variation within the collection are presumed large because no selection has never been done. 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. Approximately 100X coverage Illumina pairedend (PE) and mate pair (MP, 2 Kb and 5 Kb insertions) reads were obtained, and assembly was done by using SOAPdenovo 2 and Platanus. In *A. lancea,* scaffolding was done by SSPACE with the MP reads. PacBio assembly with 39.7 X sequences was also done by using FALCON. In parallel, scaffolding and gap-filling was done in *E. sinica* with MP and 5.5 x PacBio reads by using OPERA LG. Iso-Seq was also performed in both *A. lancea* and *E. sinica* for gene prediction and annotation. The assembled genome and gene sequences were expected to use as references in diversity analysis.

W643: International Phytomedomics and Nutriomics Consortium (ICPN) 2
Asparaptine: A Natural Inhibitor of Angiotensin-Converting Enzyme in *Asparagus officinalis*
Ryo Nakabayashi, RIKEN Center for Sustainable Resource Science, Chuo-ku, Japan
Asparaptine consisting of L-arginine and asparagusic acid was discovered in *Asparagus officinalis* by ultrahigh-resolution metabolomics for sulfur-containing metabolites (S-omics) (Nakabayashi et al., 2013, 2016, 2017). Interestingly, asparaptine showed the inhibitory activity against angiotensin-converting enzyme *in vitro*, while L-arginine and asparagusic acid showed no activity (Nakabayashi et al., 2015). This suggested that the whole structure of asparaptine is important for the activity. This metabolite was applied to a hypertensive mouse (average around 140 mmHg) to investigate the effect of lowering blood pressure *in vivo,* lowering around 25 mmHg after two hours of a dose (50 mg/kg). After two days, the blood pressure was back to around 140 mmHg. To identify its biosynthetic genes, integrated metabolomics was performed in *Asparagus* samples. Imaging mass spectrometry and metabolome analysis using liquid chromatography-tandem mass spectrometry identified that asparaptine mainly localizes around meristem in *Asparagus* spear. Transcriptome analysis was performed to obtain differential expressing genes among the samples. With the trend of accumulation on asparaptine, candidate biosynthetic genes were narrowed down for phytochemical genomics. The candidates will be tested to identify the function for the asparaptine biosynthesis.
Nakabayashi, R., Sawada, Y., Yamada, Y., Suzuki, M., Yokota-Hirai, M., Sakurai, T., Saito, K. (2013). Combination of Liquid Chromatography-Fourier Transform-Ion Cyclotron Resonance-Mass Spectrometry with ^{13}C-labeling for Chemical Assignment of Sulfur-containing Metabolites in Onion Bulbs. Anal. Chem., 85, 1310-1315.

Nakabayashi, R., Yang, Z., Nishizawa, T., Mori, T., Saito K. (2015). Top-down targeted metabolomics reveals a sulfur-containing metabolite with inhibitory activity against angiotensin-converting enzyme in *Asparagus officinalis*. J. Nat. Prod. 78, 1179-1183.
Nakabayashi, R., Sawada, Y., Aoyagi, M., Yamada, Y., Yokota-Hirai, M., Sakurai, T., Kamoi, T., Rowan, D.D., Saito, K. Chemical Assignment of Structural Isomers of Sulfur-Containing Metabolites in Garlic by Liquid Chromatography−Fourier Transform Ion Cyclotron Resonance−Mass Spectrometry. (2016). J. Nutr. 146, 397S-402S.
Nakabayashi, R. and Saito, K. (2017). Ultrahigh resolution metabolomics for S-containing metabolites. Curr. Opin. Biotechnol. 43, 8-16.

W646: International Phytomedomics and Nutriomics Consortium (ICPN) 2
Mapping of Yield and Quality QTLs in Bitter Gourd through Genotyping by Sequencing

Tusar Kanti Behera, ICAR-Indian Agriculture Research Institute, New Delhi, India

Bitter gourd (*Momordica charantia* L.; 2n=22; ~339Mb) is an economically important vegetable as well as medicinal crop in tropical countries. Fruits and seeds of bitter gourd are consumed at immature stage and both possess medicinal properties. The genetic map was constructed using $F_{2:3}$ mapping populations derived from the cross DBGy-201 × Pusa Do Mousami through genotyping by sequencing (GBS). A total of 2,013 high quality SNP markers derived from GBS were binned on 20 linkage groups. The average distance between markers was 1.16 cM and total 73 QTLs were identified using composite interval mapping (CIM) for 18 traits those mapped on 20 linkage groups. The gynoecy (*gy-1*) trait was mapped with SNP marker TP_54890 with a distance of 3.04 cM on LG-12 whereas ridgeness and tubercles were mapped on LG-4 and LG-13 respectively. Two major QTLs were identified for node at first pistillate flower appearance and both QTLs explained 25.97% of phenotypic variation. Five major additive QTLs were identified for days to first pistillate flower appearance and explained highest of 58.75% phenotypic variation. Four major QTLs were identified for sex ratio and together explained 65.60% of phenotypic variation. Two major additive QTLs were identified for number of primary branches and both explained 61.35% of phenotypic variation. Three major QTLs were identified for fruit length and together explained 38.37% of phenotypic variation. Major QTLs mapped and identified in present study have key role in assisting molecular breeding for quality traits, as well as for future genome editing projects.

W647: International Sheep Genomics Consortium
The Ovine FAANG Project Update

Brenda M. Murdoch, Department of Animal and Veterinary Science, University of Idaho, Moscow, ID

W648: International Sheep Genomics Consortium
The SheepGenomesDB Update

Hans D. Daetwyler, La Trobe University, Bundoora, Australia

W649: International Sheep Genomics Consortium
Comprehensive Transcriptional Profiling of the Gastrointestinal Tract of Ruminants from Birth to Adulthood

Emily Clark, Stephen J Bush, Mary E. B. McCulloch, Charity Muriuki, Mazdak Salavati, Gemma Davis, Iseabail L. Farquhar, Zofia Lisowski, Alan L. Archibald and David A. Hume, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom

One of the most significant physiological challenges to neonatal and juvenile ruminants is the development and establishment of the rumen. Using a subset of Illumina RNA-Seq data from our high-resolution atlas of gene expression in sheep (*Ovis aries*) we have provided the first comprehensive characterisation of transcription of the entire gastrointestinal (GI) tract during the transition from pre-ruminant to ruminant. The multi-dimensional dataset comprises 164 tissue samples from Texel x Scottish Blackface sheep at four different time points (birth, one week, 8 weeks and adult).

Using network cluster analyses we illustrate how the complexity of the GI tract is reflected in tissue- and developmental stage- specific differences in gene expression. The most significant transcriptional differences between neonatal and adult sheep were observed in the rumen complex. This was reflected both in principal component analysis of expression patterns in the wider dataset and a set of macrophage-associated transcripts. Comparative analysis of gene expression in age-matched sheep and goats revealed species-specific differences in genes involved in immunity and metabolism.

This study improves our understanding of the transcriptomic mechanisms involved in the transition from pre-ruminant to ruminant. The results form a basis for future studies linking gene expression with microbial colonisation of the developing GI tract and will contribute towards identifying genes that control immunity and metabolism in early development, which could be utilised to improve ruminant efficiency and productivity. We are now performing similar analysis on the time series of reproductive and fetal tissues from the sheep gene expression atlas dataset.

W650: International Sheep Genomics Consortium
High Density Genome-Wide Association with Domestic Sheep Monocyte Count

Ryan D. Oliveira, Washington State University, Pullman, WA

W651: International Sheep Genomics Consortium
The Genomic Architecture of South African Mutton, Pelt and Dual Purpose Sheep Breeds Relative to Global Sheep Populations

Farai C. Muchadeyi, Agricultural Research Council, Pretoria, South Africa and Edgar F. Dzomba, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Sheep *(Ovis aries)*are globally an important livestock species through which products such as mutton, wool and pelts are obtained. Breed development resulted in specialized animals according to products (primary trait of economic importance), as well as adaptive and functional

traits, leading to a diverse array of phenotypically distinct breeds. The current study used the Illumina OvineSNP50 BeadChip genotypes to evaluate genetic diversity, population genetic structure and divergence between South African sheep breeds. Genotype data from 400 animals from 13 breeds developed for mutton, pelt, and mutton and wool dual-purpose from South Africa were obtained. In addition, Nguni sheep samples were included as a representative of unimproved indigenous breeds that are reared in smallholder farming areas. We hypothesized that breed history and genomic architecture are aligned to breeding goals and production systems of the various sheep breeds of South Africa. We also juxtaposed the South African breeds to global populations consisting of 623 animals comprising six breeds from other African countries, and two Asian and eight European breeds to gain more insight into the South African breeds' genetic diversity relative to worldwide populations. Across breeds, genetic diversity ranged from H_O = 0.62±0.01 (Dohne Merino) to H_E= 0.74± 0.02 (Namaquar Afrikaner) with an overall mean of 0.65±0.04. The African and Asian populations were the most inbred populations with F_{IS}ranging from 0.17±0.05 in Grey Swakara sheep to 0.34±0.07 in the Namaqua Afrikaner. The South African Dohne Merino (F_{IS}= 0.03±0.01), SA Merino (F_{IS}= 0.05±0.04) and Afrino (F_{IS}= 0.09±0.02) and other global Merino breeds were the least inbred. The first principal component explained 27.7% of the variation and grouped the Merino and Merino derived breeds as a single cluster separated from (i) a cluster of Dorset Horn and Australian Poll Dorset, (ii) the Karakul, Karakas and Swakara cluster, (iii) the Dorpers, (iv) a cluster of Nguni and Black head Persian and (v) a mixed cluster with Ethiopian Menzi, Red Maasai, Ronderip Afrikaner and Namaqua Afrikaner. The second principal component explained approximately 19.7% of the total variation and clustered the breeds according to their function and historical origin splitting the different Merino specialized breeds and distinguishing the Nguni, Namaqua Afrikaner, Blackhead Persian, Afrino and Dorper breeds. The optimal cluster K = 20 for ADMIXTURE revealed various sources of within and amongst breed genomic variation associated with purpose, adaptation and history of the breeds. The Blackhead Persian, Nguni and Namaqua Afrikaner breeds differed significantly from other breeds particularly with the South African Mutton Merino and Dorset Horn. Further analysis included population pairwise and per marker $F_{ST}$and Runs of Homozygosity, which reported breed differentiating SNPs in genomic regions associated with growth, adaptation and reproduction. The results gave insight into the current status of the sheep genetic resources of South Africa as well as information to guide breed conservation and improvement.
Key words: Sheep, SNP genotypes, breed diversity, population structure, South Africa

W652: **International Sheep Genomics Consortium**

The Genetic Basis of Respiratory Disease in New Zealand Lambs

Kathryn McRae[1], Hayley J. Baird[2], Suzanne J. Rowe[1], Matthew J. Bixley[1], Ken G. Dodds[1] and Shannon M. Clarke[1], (1)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand

Respiratory disease is an important issue for sheep production in New Zealand and internationally, leading to reduced growth rate and a predisposition to pleurisy. The majority of the economic burden is through slower growth and reduced carcass value at slaughter. Historically, farm management practices and vaccine development have been the main focus for the prevention of pneumonia. The objective of this research was to 1) develop a visual scoring system for ovine pneumonia at the processing plant, 2) estimate the heritability of pneumonic lesions and pleurisy at slaughter, and investigate the genetic relationship with key production traits, and 3) identify regions of the genome associated with respiratory disease in sheep. The lungs of over 13,000 lambs from pedigree-recorded flocks have been scored for the presence and severity of pneumonic lesions at slaughter, using a Consolidated Pneumonia Score (CPS). On average 29% of lambs had pneumonic lesions, with 7% showing severe lesions. The incidence of pleurisy, as scored by the processing plant, was 6%. Heritability estimates for pneumonic lesions and pleurisy adjusted for heteroscedasticity (CPSa and PLEURa) were 0.07 (± 0.02) and 0.02 (± 0.01), using pedigree records, and 0.16 (± 0.03) and 0.05 (± 0.02) when using the genomic relationship matrix from a subset (N ≈ 3500) of genotyped animals. There was a significant genetic correlation between CPSa and faecal egg count (0.30 ± 0.13). Animals with pneumonic lesions had grown faster from birth to weaning, and slower from weaning to slaughter than animals without lesions. Genome-wide association analyses identified several genomic regions associated with pneumonic lesions and pleurisy in NZ lambs. These regions contained genes involved in DNA damage response and the innate immune response, including several SNPs within genes that have previously been associated with the respiratory system in cattle, pigs, rats, and mice. This indicates that there may be common pathways that underlie the response to invasion by respiratory pathogens in multiple species.

W653: **International Sheep Genomics Consortium**

Genome-Wide Variations and Signatures for Fat Deposition and Tail Morphology in Ethiopian Indigenous Sheep

Joram M. Mwacharo, International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia

Variations in body weight and body fat distribution are associated with feed quality and quantity, thermoregulation and energy reserve. Ethiopia is characterized by distinct agro-eco-climatic zones and human ethnic farmer diversity of ancient origin, which have impacted on the variation of its livestock species. Here, we investigated autosomal genome-wide profiles of 11 local Ethiopian sheep populations using the Ovine 50K SNP BeadChip assay. Sheep from the Caribbean, Europe, Middle East, China, and western, northern and southern Africa were included to address globally, the genetic variation and history of Ethiopian sheep. Population structure and phylogenetic analysis separated the Ethiopian fat-tail sheep from their North African and Middle Eastern counterparts. It indicates two main genetic backgrounds and supports two genetic histories for the African fat-tailed sheep. Within Ethiopia, four genetic backgrounds are observed which occur at different proportions among fat-rump and long fat-tail sheep from western and southern Ethiopia while the short fat-tail sheep do not represent a monophyletic group. The fat-rump sheep share one genetic background with Sudanese thin-tail sheep. Selection signature analysis identified eight candidate genomic regions spanning genes influencing growth traits and fat deposition (*NPR2*, *HINT2*, *SPAG8*, *INSR*), development and formation of limbs, skeleton and tail (*ALX4*, *HOXB13* and *BMP4*), embryonic development of tendons, bones and cartilage (*EYA2*, *SULF2*), regulation of body temperature (*TRPM8*), weight and height variation (*DIS3L2*), control of lipogenesis and intracellular transport of long-chain fatty acids (*FABP3*), the presence/absence and morphology of horns (*RXFP2*), male reproduction (*SPATA24*) and response to heat stress (*DNAJC18*). Our findings indicate that Ethiopian indigenous fat-tail sheep represent an admixed but distinct genepool and is an important resource for understanding the genetic control of skeletal growth, fat metabolism and associated physiological processes.
Submitters email address: j.mwacharo@cgiar.org

W654: **International Sheep Genomics Consortium**

Genome-Wide Association with Ovine Footrot

Mehmet Ulas Cinar[1,2], Ryan D. Oliveira[1], Tracy Hadfield[3], Anne Lichtenwalner[4], Richard J. Brzozowski[4], C. Thomas Settlemire[5], Susan B. Schoenian[6], Charles Parker[7], Noelle Cockett[3] and **Stephen N. White**[1,8], (1)Washington State University, Pullman, WA, (2)Erciyes University, Kayseri, Turkey, (3)Utah State University, Logan, UT, (4)University of Maine, Orono, ME, (5)Bowdoin College, Brunswick, ME, (6)Western Maryland Research & Education Center, Keedysville, MD, (7)retired, Columbus, OH, (8)USDA, ARS, Animal Disease Research Unit, Pullman, WA

Ovine footrot is a bi-microbial infectious disease in which *Dichelobacter nodosus* and *Fusobacterium necrophorum* play major roles, and it is characterized by hoof separation from the underlying tissue associated with severe lameness. Economic losses due to footrot have been estimated at more than $20 million per year in the U.K, and are known to be much higher globally. While there are a variety of treatment methods, many are labor-intensive and expensive due to the treatment or culling of the diseased animal. Genetic markers to enable selective breeding for resistance to footrot would be a valuable tool to advance disease control. Footrot condition scores were collected from 251 U.S. sheep including Katahdin, Blackbelly, and European-influenced crossbred sheep. Of these, 200 were genotyped with the OvineHD array (>600,000 SNP) and 51 others were genotyped on the Ovine50K array (>50,000 SNP). The sheep genotyped with the Ovine50K array were imputed up to OvineHD density to create a uniform genotype dataset, and genome-wide association was performed using a model accounting for breed group. Genome-wide significance was observed for loci on ovine chromosomes 1 and 2, including a gene known to be expressed in skin but with little additional prior functional annotation. Furthermore, genome-wide suggestive results were observed on numerous other chromosomes. These data provide a basis for discovery of specific genetic variants underlying footrot susceptibility, and they will improve genetic and genomic selection to selectively breed healthy sheep.

W6**55: International Wheat Genome Sequencing Consortium (IWGSC)**

Studying Global Wheat Genetic Diversity of Germplasm Banks

Carolina Paola Sansaloni, CIMMYT- International Maize and Wheat Improvement Centre, Texcoco, Mexico

W6**56: International Wheat Genome Sequencing Consortium (IWGSC)**

No Magic Involved: Chromosome-Scale Sequence Assembly of Wheat Genomes with Open-Source Tools

Martin Mascher, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

The completion of a chromosome-scale reference sequence for bread wheat cultivar Chinese Spring marks a milestone for wheat genomics. A possible direction for future research in wheat genomics is the assembly of genome sequence for representatives of global wheat diversity. An important component in this endeavour will be a reproducible, open-source pipeline for sequence assembly, enabling fast and cost-effective assembly of a large number of genotypes. In this presentation, I will introduce a computational workflow to construct chromosome-scale sequence assemblies in Triticeae species based on a combination paired-end, mate-pair, Chromium 10X and chromosome conformation capture sequencing data. The pipeline was evaluated on the published raw data for the IWGSC Chinese Spring reference assembly and the assembly of wild emmer accession Zavitan. Moreover, first results on a reference genome sequence assembly for Ae. sharonensis, a diploid wild relative of wheat, will be presented. The implications for wheat pan-genomics and genetic analyses in wheat wild relatives will be discussed.

W6**57: International Wheat Genome Sequencing Consortium (IWGSC)**

Finishing Strategy for the IWGSC RefSeq Genome Sequence

Philippe Rigault[1,2], Sébastien Boisvert[1], Josquin Tibbits[3], Gabriel Keeble-Gagnère[3] and Rudi Appels[3], (1)GYDLE, Québec, QC, Canada, (2)Center for Organismal Studies (COS), University of Heidelberg, Heidelberg, Germany, (3)Agriculture Victoria, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia

The IWGSC RefSeq v1.0 assembly represents a major milestone in wheat research by providing the first ordered and annotated reference genome sequence for bread wheat (*Triticum aestivum*). The reference pseudomolecules represent 88% of the chromosome sequences (14.1 Gb out of an estimated 16 Gb genome), with 481 Mb of additional scaffolds unassigned to chromosomes, and approximately 1 Gb missing. The assembly sequence also contains over 130,000 gaps of unknown size and 261 Mb of uncalled bases in over 600,000 sized gaps.

We present a strategy aimed at finishing the wheat genome reference sequence in a way that maximizes the impact for the research community, by combining whole genome improvements (resolving gaps, unassigned and missing sequences) with targeted methods (e.g focusing on gene sequences and regions of interest), while allowing the tracking of changes across versions.

The bulk of this strategy can be achieved without any additional data acquisition, by using the data acquired by the IWGSC during the production of the IWGSC RefSeq v1.0 assembly and processing it with Gydle software to map the raw data to the assembly and then perform gap, repeat and connectivity resolution as well as targeted finishing with visualisation-assisted editing. Additional resolution methods target the ribosomal DNA and sattelite DNA clusters.

The value of using additional data sources will be demonstrated using chromosome 7A, where an independent BAC-based assembly and BioNano maps were used in combination with the IWGSC RefSeq data for finishing large targeted regions. Hybrid assembly resolution with BioNano and Illumina data using Gydle software helped achieve megabase-size gapless sequences and resolve tandem repeats that were collapsed in the IWGSC assembly. The average increase in sequence size after resolving tandem repeats was 8%, which corresponds to the fraction of missing sequence in the reference assembly, corroborating the hypothesis that most of the missing sequence in the IWGSC RefSeq v1.0 assembly lies in collapsed repeats.

W6**58: International Wheat Genome Sequencing Consortium (IWGSC)**

The Comparative Transcriptomic Analysis of Wheat, Barley and Wheat-Barley Chromosome-Arm 7HL Addition Line Reveals Major Changes Affecting the Gene Content of both the Alien Telosome and the Wheat Genome

Elodie Rey[1,2], Michael Abrouk[1,2], Gabriel Keeble-Gagnere[3], Miroslava Karafiátová[2], Jan Vrána[4], Sandrine Balzergue[5,6], Ludivine Soubidou-Taconnat[5], Veronique Brunaud[5], Marie_Laure Martin-Magniette[5,7], Takashi R. Endo[8], Jan Bartoš[4], Kellye Eversole[9], Rudi

Appels[10] and Jaroslav Doležel[2], (1)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, (2)Institute of Experimental Botany, Olomouc, Czech Republic, (3)AgriBio, Melbourne, Australia, (4)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (5)Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université Evry, Université Paris-Saclay, Paris Diderot, Sorbonne Paris-Cité, Orsay, France, (6)IRHS, Université d'Angers, INRA, AGROCAMPUS-Ouest, SFR4207 QUASAV, Université Bretagne Loire, Beaucouzé, France, (7)UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, Paris, France, (8)Department of Plant Life Science, Faculty of Agriculture, Ryukoku University, Shiga, Japan, (9)IWGSC, Bethesda, MD, (10)University of Melbourne, Melbourne, Australia
Bread wheat (*Triticum aestivum* L.) is one of the most important crops, providing staple food for 34 % of the human population. In order to satisfy the increasing demand, novel genetic diversity must be introduced into wheat cultivars to increase yield, quality and resilience to adverse environmental conditions. An attractive way is to introduce beneficial genes and alleles from related species, a process called alien introgression. However, poor knowledge on the interactions between the host and alien chromatin and their impact on gene regulation have hampered the use of wheat-alien introgression into agriculture.
In the present study, we used the recent reference genomes of wheat and barley to study the differential transcription between bread wheat, barley and the wheat-barley chromosome 7HL addition line. The study revealed a large proportion (~42%) of barley 7HL genes differentially transcribed in the addition line, while only 3% of wheat genes was affected. Moreover, the genes experiencing differential transcription in the addition line were randomly distributed across the genome, but a non-stochastic relation to their function. Our results identified a set of genes of potential relevance for interspecific incompatibility between wheat and its related species. We also revealed an important rearrangement of the wheat host genome in the addition line compared to the IWGSC RefSeq.
Our study represents a major step towards a better understanding of the interactions between wheat and alien genes in introgression lines and opens the way for further transcriptomic analysis in other alien introgression lines, including breeding plant material.

W659: **International Wheat Genome Sequencing Consortium (IWGSC)**
Designer Roots for Future Wheats
Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

W660: **International Wheat Genome Sequencing Consortium (IWGSC)**
Possibility of CRISPR-Based Precision Breeding in Wheat
Eduard Akhunov, Kansas State University, Department of Plant Pathology, Manhattan, KS

W661: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
ELIXIR Resources for Interoperability
Frederik Coppens, VIB, Gent, Belgium
Overview of ELIXIR infrastructure for Interoperability: bioconda, biocontainers, Galaxy, CWL,
Registries such as bio.tools, myExperiment, BioContainers

W662: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
Cyberinfrastructure for Landscape Genomics: Connecting Biological Databases, Metadata, and Intelligent Analytics
Jill L. Wegrzyn, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

W663: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
Soykb and Kbcommons Framework for Enabling Multiomics Research Applications
Shuai Zeng, University of Missouri, Columbia, MO
Advancement of next generation sequencing and high-throughput technologies has resulted in generation of multi-level of 'OMICS' data for many organisms. However, these data are often individually scattered across different repositories based on data type, making it difficult to integrate them. We have addressed this issue through our in- house developed Soybean Knowledge Base (SoyKB) framework, a comprehensive web-based resource. It acts as a centralized repository for soybean multi-omics data, and is equipped with an array of bioinformatics analytical and graphical visualization tools. It is available at http://soykb.org and has proven to be a great success with more than 500 registered users. Users working on other biological organisms including plants, animals and biomedical diseases have similar needs and the developed framework can be expanded to make the visualization and analysis tools function for other organisms, without having to reinvent the wheel. To achieve this we have developed KBCommons, a platform that automates the process of establishing the database and making the tools for other organisms available via a dedicated web resource. It provides information for six entities including genes/proteins, microRNAs/sRNAs, metabolites, SNP, traits as well as plant introduction or strains/populations. It also incorporates several multi-omics datasets including transcriptomics, proteomics, metabolomics, epigenomics, molecular breeding and other types. We have currently expanded KBCommons framework and tools to Zea mays , Arabidopsis, Mus musculus and Homo sapiens. We have integrated various genomics dataset for maize including RNAseq B73 mutants and Tassel meristem from our collaborators. It provides a suite of tools such as the gene/metabolite pathway viewer, Protein Bio-Viewer, heatmaps, scatter plots and hierarchical clustering. It also provides access to PGen, Pegasus analytics workflows developed for genomics variations analysis. It also has suite of tools for differential expression analysis of transcriptomics and other multi-omics datasets including venn diagrams, volcano plots, function enrichment and gene modules.

W664: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
Sustainable Business Model for Scientific Data Resources

Trilok Prithvi, Phoenix Bioinformatics, LLC, Frement, CA

W665: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
Workflows and Best Practices for High Throughput RNA-Seq Processing and Long Non-Coding RNA Identification across Eukaryotes
Andrew Nelson, School of Plant Sciences, University of Arizona, Tucson, AZ

W666: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
Data Federation in the Tripal Community and Beyond with Structured Cross-Site Searching
Margaret Staton[1], Bradford Condon[1], Jill L. Wegrzyn[2], Dorrie Main[3], Stephen P. Ficklin[4] and Abdullah Almsaeed[1], (1)University of Tennessee, Knoxville, Knoxville, TN, (2)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (3)Washington State University, Pullman, WA, (4)Dept of Horticulture, Washington State University, Pullman, WA
Community scientific databases serve important and necessary functions by being tailored for their specific users, however, the diversity of underlying storage and software systems presents a major challenge to data federation. Many sites solve this problem by offering application programming interfaces (APIs) as web services, which have numerous advantages (common structure, consistent behavior) as well as disadvantages, including the time and effort required to port standardized APIs to each site/database with a different underlying structure. A complementary solution is to use a powerful open source search engine such as Elasticsearch. Within the Tripal framework, we have implemented an Elasticsearch module that enables cross-site searching for users. A Tripal site installs and indexes all of their content with Elasticsearch, then exposes that index online, enabling any other site to to gain read only access to the index and present search results to users. For example, a user searching by keyword in Hardwood Genomics will now also see relevant search results from TreeGenes and the Citrus Genome Database. In comparison to web services, this is a quick way to begin integrating data between two different databases with little programming, and it allows users to interact with different data stores via one interface. While the primary disadvantage to this solution can be a lack of sophisticated data structure, there are a number of ways to maintain and take advantage of structured page text via Elasticsearch or in combination with web services. Examples within Tripal are now available, and applications to other software systems are feasible. Future work is focused on using ontologies and standardized metadata to structure search results and enable sophisticated data integration. As progress in this area is largely dependent on collaboration between data resource groups, all interested scientists are encouraged to join the Data Sharing working group within the the AgBioData consortium, a group of agricultural biological databases and associated resources working toward data federation standards.

W667: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
Recycling Data to Enable Investigation of Alternative Splicing across the Plant Kingdom
Brad Barbazuk, University of Florida, Gainesville, FL
The emergence of technologies that facilitate the high-throughput and inexpensive acquisition of genome sequence has resulted in an explosion of data that is available through public data repositories such as the NCBI SRA. Access to these data enables investigation of important biological questions without requiring additional sequence data to be generated. My lab is specifically interested in examining alternative splicing (AS) across the plant kingdom. Alternative splicing plays important roles in many plant biological processes, but its conservation across the plant kingdom is not known. We describe a methodology to identify AS events and identify conserved AS events across large phylogenetic distances using large publicly available RNA-Seq datasets. We have characterized and compared AS across 15 individual species, several maize lines and 10 rice relatives, and we are expanding this to include 50 plant species. I will discuss the project strategy, current limitation in using public data, current results and the implications of the research. I will also address the use of public data and publicly available genome analysis platforms to introduce undergraduates to high-throughput genomics.

W668: **Interoperability and Federation Across Bioinformatic Platforms and Resources**
TBD
David LeBauer, University of Arizona, Tucson, AZ

W669: **IRIC: Rice Informatics for the Global Community**
The Power of Machine Learning-Based Genomic Prediction in Hybrid Breeding
Sebastian J. Schultheiss, Computomics GmbH, Tuebingen, Germany

W670: **IRIC: Rice Informatics for the Global Community**
Unravelling the Diversity of Native Rice in Vietnam for Future Climatic Scenarios
Janet Higgins[1], Bruno A. Santos[2], Tran Dang Khanh[3], Khuat Huu Trung[3], Tran Duy Duong[3], Nguyen Thi Phuong Doai[3], Nguyen Truong Khoa[3], Dang Thi Thanh Ha[3], Le Huy Ham[3], Mario Caccamo[2] and **Jose De Vega**[1], (1)Earlham Institute, Norwich, United Kingdom, (2)NIAB, Cambridge, United Kingdom, (3)Institute of Agricultural Genetics, Hanoi, Viet Nam
Vietnam is an important country for rice production, both for export and for providing a staple food for circa 100 million people. Vietnam's national seedbank holds a collection of over 9,000 samples which represent the rich diversity of rice germplasm grown in Vietnam. This includes native and local landraces and varieties adapted to growing in the adverse growing conditions in the low-lying rice deltas. These regions are particularly vulnerable to the effects of climate change, meaning that there is an urgent need to understand the local diversity and use this knowledge to accelerate the breeding of new climate-resilient rice varieties.
In order to identify unexplored diversity, we have sequenced the whole genome of 616 rice accession native from Vietnam, which represent the rich diversity of rice germplasm grown in the country. To date, we have characterised the population structure, linking our results to these described by the international community. The results evidenced a clear division among the var. *indica* elite lines (particularly found in and

around the Mekong delta region) and landraces, which classed by climate/latitude. Within the var. *indica* landraces, we have observed two new well-defined subpopulations of around the Red River delta region. In Vietnam, the var. *japonica* subpopulations primarily distinguished between upland and lowland environments. This population structure has allowed to identifying evidences of genomic signatures of recent selectionin specific subpopulations: domestication in elite accessions, coolness in northern accessions, drought in upland accessions, and salinity in lowland coastal accessions. We are also carrying out genome-wide phenotype-genotype association analysis on a variety of morphological traits, including grain characteristics, heading date, and inflorescence and leaf characteristics. This data set offers opportunities to extend the analysis to more complex traits in the future, particularly related to tolerance to pests and abiotic stress.

W671: **IRIC: Rice Informatics for the Global Community**

The Rise and Fall of African Rice Cultivation History Revealed by Extensive Genome-Resequencing and Population Genomics Modeling Approaches

Philippe Cubry[1], Christine Tranchant-Dubreuil[1], Anne Celine Thuillet[1], Cécile Monat[1], Marie-Noelle Ndjiondjop[2], Karine Labadie[3], Corinne Cruaud[3], Stefan Engelen[3], Nora Scarcelli[1], Bénédicte Rhone[4], Concetta Burgarella[5], Christian Dupuy[6], Pierre Larmande[1], Patrick Wincker[3], Laurence Albar[1], Hélène Pidon[1], Stéphane Jouannic[1], Hélène Adam[1], Olivier Francois[7], François Sabot[1] and Yves P. Vigouroux[1], (1)IRD - UMR DIADE, Montpellier, France, (2)Africa Rice, Cotonou, Benin, (3)CEA - Genoscope, Evry, France, (4)IRD - UMR DIADE, Lyon, France, (5)CIRAD - UMR AGAP, Montpellier, France, (6)Institut des mondes africains (IMAF), Paris, France, (7)Grenoble INP TIMC-IMAG, Faculty of Medicine, Grenoble, France

Of interest for agriculture and early agricultural society history, crop domestication has been deeply investigated using archeological and ethnological approaches. The recent development of genetic statistical methods and the availability of full genome sequences offer new opportunities to understand this domestication process. We combined investigation of past population effective size history and spatially explicit coalescent-based simulations analyses to investigate the spatial origin and domestication history of African rice, using 163 cultivated and 83 wild relatives resequenced genomes. We shown that wild rice populations in West Africa experienced a long and steady period of population decline beginning more than 10,000 years ago, a feature common with the domesticated species. This decrease was certainly caused by external ecological factors rather than directly human induced. This result suggests that drying of the Sahara depleted wild resource and might have favored domestication of African rice. The wild species recovered 3,200 years ago, corresponding to the end of the Sahara drying, but cultivated population only spread In West Africa 2,000 years ago. We also detected a very recent decrease of population size in African rice cultivation, in the last 500 years. This decrease is the likely consequence of a documented introduction of Asian rice crops in West Africa, replacing African rice cultivation. Using a spatial model, we give a probability map for the origin of its cultivation, with the highest probability found in the region of the Inner Niger Delta. This inference is in perfect agreement with the oldest archeological remains of African rice and with the extant of cultivated genetic diversity. Investigating the genes under selection during domestication allowed us to evidence convergent evolution during domestication process of both asian and african rice on key genes and pathways implicated in plant architecture or seed dehiscence. Lastly, we took advantages of this large genomic dataset to perform Genome Wide Association Studies on panicle architecture, flowering time and resistance to Rice Yellow Mottle Virus. Altogether, our study refined the knowledge on domestication history of African rice by mobilizing advanced statistical and population genetics tools to address where, when and how the domestication of rice in Africa had occurred.

W672: **IRIC: Rice Informatics for the Global Community**

Rice Genome-Scale Network Integration Reveals Transcriptional Regulators of Grass Cell Wall Synthesis

Kangmei Zhao[1], Fan Lin[2], Sandra Romero-Gamboa[3], Prasenjit Saha[2], Hyung-Jung Goh[4], Gynheung An[5], Ki-Hong Jung[5], Samuel P. Hazen[6] and Laura Bartley[2], (1)Carnegie Institution for Science, Palo Alto, CA, (2)University of Oklahoma, Norman, OK, (3)University of Massachusetts, Amherst, MA, (4)Kyung Hee University, Yongin, South Korea, (5)Kyung Hee University, Yongin, Korea, Republic of (South), (6)Biology Department, University of Massachusetts, Amherst, MA

Grasses have evolved distinct cell wall composition and patterning relative to dicotyledonous plants. To better interrogate grass-specific processes using network analysis, we developed a genome-scale integrated network from publicly available rice networks, ROAD, PlaNet and RiceNet v2. Our heuristic strategy was to use a generalized linear model (GLM) to recalibrate the edge scores between genes within the two coexpression networks, ROAD and PlaNet to the scoring system of a functional network RiceNet v2. The resulting Rice Combined inverse Ranked Network (RCRN) covers 93% of annotated rice (*Oryza sativa*) genes, is high quality, and includes most grass-specific cell wall genes, such as mixed-linkage glucan synthases and hydroxycinnamoyl acyltransferases. Comparing the RCRN and an equivalent *Arabidopsis* network suggests that grass orthologs of most genetically verified eudicot cell wall regulators also control this process in grasses, but some vary significantly in network connectivity between these divergent species. Protoplast-based transient assays validated 10 out of 15 predicted transcription factors for association with cell wall biosynthesis in rice, including 6 of 11 that were previously unstudied. Reverse genetics, yeast-one-hybrid, and protoplast-based assays reveal that OsMYB61a, a co-ortholog of functionally characterized cell wall-associated transcription factor in Arabidopsis, activates a grass-specific acyltransferase promoter, which confirms network predictions and supports grass-specific cell wall synthesis genes being incorporated into conserved regulatory circuits. This study provides insight into the evolution of cell wall regulation and highlights the quality of the RCRN for examining rice biology.

W673: **IRIC: Rice Informatics for the Global Community**

Structural Variation in the 3000 Rice Genomes

Dmytro Chebotarov, International Rice Research Institute, Los Baños, Philippines

W674: **IWGSC – Wheat Genome Manual and Functional Annotation**

Online Real-Time Chromosome-Scale Visualisation and Analysis of Wheat Genomes with Pretzel

Josquin Tibbits, Agriculture Victoria, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia

W675: **IWGSC – Wheat Genome Manual and Functional Annotation**
Phylogenomic Studies to Annotate Amplified Gene Families
Daniel Lang, Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany

W676: **IWGSC – Wheat Genome Manual and Functional Annotation**
Gene Networks to Predict Gene Function
Philippa Borrill[1], Ricardo Humberto Ramirez Gonzalez[2], Andrea Braeutigam[3] and Cristobal Uauy[2], (1)University of Birmingham, Birmingham, United Kingdom, (2)John Innes Centre, Norwich, United Kingdom, (3)Bielefeld University, Bielefeld, Germany
The publication of the completely annotated RefSeqv1.0 wheat genome sequence has revealed full length gene models for over 100,000 wheat genes. However, the challenge remains to identify the functions of these genes and how to use this knowledge for crop improvement. Through combining RNA-Seq data with gene network modelling approaches it is now possible to generate hypotheses about gene function.
In this study we have leveraged publicly available RNA-Seq data (529 samples from 28 studies) and new studies (321 samples) to explore homoeologue-specific global gene expression in hexaploid wheat using the RefSeqv1.0 genome assembly. We classified samples according to their tissues, developmental stages, cultivars and environmental conditions. Using these 850 samples we have developed separate co-expression networks across a range of tissues (grain, leaf, spike and root) and stress conditions (abiotic and biotic), alongside a global network which infers transcription factor targets.
We will present examples of using these gene networks to identify potential flowering regulators, stress resistance genes and genes which regulate root development. These networks now enable the formulation of hypotheses about gene function in wheat itself, which extends the information that can be gathered from studying orthologues in model plants such as Arabidopsis and rice. This network-enabled approach will contribute to understanding the function of wheat genes and how interactions between genes control developmental and stress responses. These networks are available to the wheat community through http://knetminer.rothamsted.ac.uk/ to help facilitate the identification of genes regulating agronomically relevant traits.

W677: **IWGSC – Wheat Genome Manual and Functional Annotation**
Assigning GO Terms to Wheat Gene Models
Dennis Psaroudakis[1], Wimalanathan (Gokul) Kokulapalan[2] and **Carolyn J. Lawrence-Dill**[1], (1)Iowa State University, Ames, IA, (2)Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA
We constructed a high-coverage and reproducible functional annotation dataset for wheat based on Gene Ontology (GO) term assignments, covering 94% of the 107,891 high-confidence gene models in IWGSC's RefSeq 1.1 genome with 9.5 annotations per gene model on average. To derive this annotation set, we used the GOMAP pipeline, which includes sequence similarity and protein domain presence methods as well as mixed-method pipelines that were developed for the Critical Assessment of Function Annotation (CAFA) challenge (https://biofunctionprediction.org/cafa/). Whereas maize annotations derived from the pipeline were quality checked (based on hand-curated functional annotations; https://doi.org/10.1002/pld3.52), no such gold-standard exists for wheat, so these predictions lack quality assessment. Ideas for how best to assemble a gold-standard dataset to enable quality assessments for these predictions will be discussed.
The GOMAP pipeline is available at https://github.com/Dill-PICL/GOMAP-singularity.
Wheat GOMAP annotations are available at https://dill-picl.org/projects/maize-gamer/maize-gamer-datasets/.

W678: **IWGSC – Wheat Genome Manual and Functional Annotation**
Map-Based Cloning of Agronomically Important Genes
Vijay K. Tiwari, University of Maryland, College Park, MD

W679: **IWGSC – Wheat Genome Manual and Functional Annotation**
Genomic and functional analysis of glutenin and gliadin genes in Chinese varieties
Daowen Wang, Institute of Genetics and Developmental Biology, CAS, Beijing, China and Kunpu Zhang, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
[1]The State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
[2]The State Key Laboratory of Wheat and Maize Crop Science, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
Glutenin and gliadin proteins are key determinants of wheat dough functionality and end-use traits. However, the multiplicity and extensive genetic variation of these proteins have made it very difficult to study their individual and combined effects on dough and end-use properties. To address this knowledge gap, we have combined mutagenetic, functional genomic and rheological approaches to investigate the spectrum of glutenin and gliadin proteins expressed in two Chinese winter wheat cultivars (Xiaoyan 54 and Xiaoyan 81) and their single and interactive effects on dough and breadmaking performance. In high-molecular-weight glutenin subunits (HMW-GSs), we developed knockout mutants lacking one or more subunits or one or more of the three homoeologous *Glu-1* loci. These mutants have been successfully used to demonstrate functional differences among HMW-GSs and in between different *Glu-1* loci. In low-molecular-weight glutenin subunits (LMW-GSs), we identified the LMW-GSs accumulated in both cultivars, and found a positive correlation between the number of actively expressed LMW-GS genes and the potency of the contribution of LMW-GSs to gluten and dough strength. In gliadins, combined transcriptomic and proteomic analysis showed that Xiaoyan 81 grains accumulated at least 38 gliadin proteins including 21 alpha, 11 gamma, 1 delta, and 5 omega-gliadins. Lastly, we studied the structure and function of *Gli-D2* locus located on 6D chromosome, and found that the null allele of *Gli-D2* is useful for improving the breadmaking and health-related (e.g., lysine content) traits of common wheat. Our work shows that the complexity of glutenin and gliadin genes can be tackled by combining genetic and genomic approaches. The implications of our research on further studies of glutenin and gliadin genes and their functions in end-use quality control will be discussed.

W680: **IWGSC – Wheat Genome Manual and Functional Annotation**
The PPR Gene Family in Wheat
Ian Small, The University of Western Australia, Crawley, WA, Australia

W681: **IWGSC – Wheat Genome Manual and Functional Annotation**
Wheat Prolamin Gene Regions in Chinese Spring using the Sequences Generated by PacBio Long Reads and Validated by Bionano Maps
Yong Q. Gu, USDA ARS, Western Regional Research Center, Albany, CA
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 prolamin 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 prolamin 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 prolamin gene loci from the wheat A, B, and D genomes were annotated manually to identify a complete set of wheat prolamin genes including both intact genes and pseudogenes from a single wheat cultivar cv Chinese Spring, and compared with the orthologous regions from different species (rice, *Brachypodium*, and sorghum). Our results indicated that rapid evolutionary dynamics are present only in the wheat genomes. The high frequency of sequence rearrangements including deletion, duplication and translocation events have resulted in considerable synteny erosion in the prolamin genomic regions. We propose that the HMW-glutenin genes originate from a tandem duplication of an ancestral globulin gene, while other prolamin genes (LMW-glutenin, alpha-, gamma-, and omega-gliadins) are likely derived from gene translocations followed by multiple rounds of gene duplications. The expression of individual prolamin genes were analyzed using transcriptome data. We found that the A genome contributes the least to prolamin expression in cv Chinese Spring because of its smaller number of expressed intact genes and their low expression levels, while the B and D genome contribute at higher and approximately equal levels. Our study also provided insights into the evolution of CD epitopes and identified that a single indel event in the hexaploid wheat D genome likely resulted in the generation of the highly toxic 33-mer CD epitope. The knowledge gained from this study can facilitate the breeding of wheat varieties with improved end-use traits and reduced immunogenic potential.

W682: **IWGSC – Wheat Genome Manual and Functional Annotation**
Visualizing the IWGSC Refseq v1.0 Wheat Assembly in Ensembl Plants
Guy Naamati, EMBL-EBI, Hinxton, United Kingdom
The Ensembl Plants project offers an integrative collection of interfaces for accessing and comparing genome-scale data for over 50 plant species, including the recently published IWGSC RefSeq V1.0 assembly.
This presentation will provide an overview for visualizing and analysing the IWGSC assembly in the Ensembl Plants platform with specific reference to all the variation data that has been aligned to the assembly (including the TILLING population) and also comparative genomics for the gene annotation set. In addition this presentation will describe how to view and analyse the different genomic components in the form of a polyploidy view and comparison of homoeologues across the A, B and D components.
Finally the presentation will provide a guide for the community to use the IWGSC assembly as a reference for different types of metadata which we can be stored in Ensembl Plants and visualized.

W683: **IWGSC – Wheat Genome Manual and Functional Annotation**
The Wheat Genome Visualized in Persephone
Max Troukhan, Persephone Software, LLC, Agoura Hills, CA
The large size of the wheat genome presents a challenge for developers of tools that analyze and visualize genomic data on a large scale.
A multi-genome browser, called Persephone, has been built with this problem in focus. The browser is designed to show several maps side by side, allowing the users to smoothly zoom in/out and quickly switch the view from the whole genome overview to visualizing individual nucleotides. The wheat chromosomes, some exceeding 800 Mbp in size, can be aligned with genetic maps by connecting common markers. This comparison is helpful in confirming or correcting marker positions or in translating QTL information from genetic maps to the sequence. Aligning sequence maps, such as homoeologous A and B wheat chromosomes, and connecting homologous genes helps reveal large scale sequence rearrangements.
A desktop version of Persephone has been used daily in large corporations and has been proven efficient in handling millions of chromosome/scaffold maps with different types of data, including gene models, markers, QTLs, SNPs, RNA-seq, gene expression, BLAST hits, etc. Application of proprietary algorithms to data compression allows Persephone to load information at speeds of more than one million data points per second. The high data visualization rate facilitates navigation across multiple genomes.
The recently-introduced web version of Persephone is open for the scientific community at http://web.persephonesoft.com. The loaded wheat genome information can be used to showcase the capabilities of Persephone in dealing with big data.

W684: **IWGSC – Wheat Genome Manual and Functional Annotation**
Visualization, Tools, and Resources for Wheat at GrainGenes
Taner Z. Sen, USDA-ARS / GrainGenes, Albany, CA
GrainGenes (https://graingenes.org; https://wheat.pw.usda.gov) is the centralized, curated USDA-ARS database for wheat, barley, oat, and rye, ensuring long-term data sustainability for small grains researchers. GrainGenes hosts a JBrowse genome browser instance for the IWGSC

RefSeq v1.0 assembly, which is populated with multiple diversity tracks generated by the Akhunov (Kansas State Univ.) and Dubcovsky (UC Davis) Labs. GrainGenes contributes to the JBrowse genome browser development, and is developing a BLAST plug-in that can perform sequence similarity searches from the genome browser tracks. This plug-in will be available for implementation by other JBrowse users. GrainGenes is an active participant of the Wheat Information System (WheatIS) under the Wheat Initiative: for facilitated data discovery, GrainGenes indexed QTL, germplasm, and genes from the Wheat Gene Catalogue into WheatIS.

W6**85: IWGSC – Wheat Genome Manual and Functional Annotation**
Wheat Genome Resources in the Wheat@URGI
Thomas Letellier, URGI, INRA, Université Paris-Saclay, Versailles, France

W6**86: IWGSC – Wheat Genome Manual and Functional Annotation**
Open Discussion and Community Input on a Mechanism for the Curation and Validation of Manual and Functional Annotation of IWGSC Refseq v1.0
Rudi Appels, University of Melbourne, Melbourne, Australia, Kellye Eversole, IWGSC, Bethesda, MD, Etienne Paux, GDEC, INRA, UCA, Clermont-Ferrand, France and Ute Baumann, The University of Adelaide, Urrbrae, South Australia, Australia
Phase II of the IWGSC's activities includes developing a pipeline for integrating into annual releases manual and functional annotation of IWGSC RefSeq v1.0 that may be developed through IWGSC projects or by individuals within the wheat community. This will be a discussion section of the workshop where we will welcome input from the community regarding contributions that would improve IWGSC RefSeq. We also aim to draw on experience from other, perhaps more advanced, reference genome sequence projects, including processes used for regular updates of genome coordinates to EnsemblPlant as details of gene families in the assembly are completed.

W6**87: JBrowse, a Next Generation Genome Browser**
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 (easier than GBrowse!) 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.

W6**88: Legumes**
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 that could help with this has become feasible. Nevertheless, significant challenges remain due to the sheer volume of such genotype data. Comparisons between genotyping projects are additionally complicated by the lack of common markers among data sets, differences in accession names, SNPs called from different reference genomes, and by inconsistent data formats. This talk will describe the recently developed Soy SNP Registry, a database for soybean genotyping data. The Registry currently contains data from published U.S., Brazilian, Canadian, Chinese, and Korean soybean and results of comparisons between the genotyping datasets will be discussed.

W6**89: Legumes**
Domesticating the Legume-Rhizobium Symbiosis: From Wild Systems to Agricultural Applications
Douglas R Cook[1], Alex Greenlon[2], Laura Perilla-Henao[3] and Betsy Alford[1], (1)University of California-Davis, Davis, CA, (2)UC Berkeley, Berkeley, CA, (3)UC Davis, Davis, CA
Legume species are key components of both natural and agricultural ecosystems. Their importance derives in large part from their capacity for symbiotic nitrogen fixation with soil bacteria, enabling them to return vital nitrogen to the soil environment and to create seed and forage of high protein content. Two decades of molecular and genomic studies in model systems have revealed the presence of exquisite genetic pathways that initiate symbiosis, but despite these advances we have essentially no understanding of genes that regulate symbiotic performance in the natural or agricultural environment. Here we aim to understand the evolution of symbiotic performance in the wild progenitors of chickpea and the ways in which human selection has reshaped this potential during domestication. Sequencing the genomes of >1,000 chickpea symbionts from a systematic global survey of wild and cultivated systems reveals a domestication-driven network of horizontal gene transfer that has expanded the species capable of fixing nitrogen with chickpea from three to greater than fifteen. In parallel, we are characterizing the nodule-associated microbiome to identify non-rhizobia that provide services to the symbiotic organ. By analyzing a matrix of bacterial genotypes x host genotype combinations we are using genetics to infer impacts of domestication and identify genes that enhance the host's capacity for nitrogen fixation. Current data indicates that cultivated species have a broader range of effective symbiotic partner species, but with lower average benefit from symbiosis, consistent with a selection trade-off during domestication. We will present strategies to leverage such information to improve the effectiveness of nitrogen fixation in legume agricultural systems.

W690: Legumes

Integrating the Alflafa Gene Expression Atlas into the Alfalfa Breeders Toolbox

Maria J. Monteros[1], Nadim Tayeh[2], A. Brice Cazenave[3], Silvas J. Prince[1], M. Rokebul Anower[1], Christy M. Motes[1], Tim D. Hernandez[1] and Chunlin He[1], (1)Noble Research Institute, Ardmore, OK, (2)INRA, Dijon, France, (3)Virginia Tech University, Suffolk, VA

Increased demands for agricultural products developed using environmentally-friendly practices are driving forces to develop agricultural systems that are more productive, adaptable to erratic or extreme weather patterns and require fewer on-farm inputs. Current factors limiting crop production on a global scale include abiotic and biotic stresses. Integration of new technologies in genomics, transcriptomics and phenomics facilitates the development of efficient solutions to these agricultural challenges. Alfalfa (*Medicago sativa* L.) is a perennial forage legume with global agricultural value. Our approach includes evaluating genetic variation for stress adaptation and performance in alfalfa, understanding the underlying mechanisms associated with these traits and identifying functional sequence variants that can be used during selection to increase the frequency of favorable alleles. Field-based high-throughput phenotyping strategies that utilize drones and sensors mounted on pheno-mobiles were used to identify variation for key traits in alfalfa germplasm evaluated in the field. Additionally, specific root traits or phenes that enhance the capacity for efficient nutrient and water uptake have been associated with increased productivity under abiotic stress conditions. The alfalfa genome sequence, the gene expression atlas generated from plants grown under different stress conditions, and genome-wide single nucleotide polymorphisms (SNPs) associated with phenotypic traits were integrated into the Alfalfa Breeder's Toolbox (www.alfalfatoolbox.org). The use of this resource can facilitate implementation of approaches to track the accumulation of favorable alleles during selection and breeding. Further, the Toolbox will enable the integration of genomics and phenomics datasets to accelerate genetic gains in alfalfa breeding programs and effectively address current and future challenges to crop productivity.

W691: Legumes

PacBio Resequencing of *Medicago truncatula* A17 Genome and Discovery of Symbiotic Co-Regulated Genomic Islands

Yann Pecrix[1], S. Evan Staton[2], Erika Sallet[1], Christine Lelandais-Brière[3], Sandra Moreau[1], Sébastien Carrère[1], Thomas Blein[3], Marie-Françoise Jardinaud[1], David Latrasse[3], Helene Berges[4], Andreas Niebel[1], Julia Buitink[5], Moussa Benhamed[3], Martin Crespi[6], Jérôme Gouzy[1] and Pascal Gamas[1], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS/Université de Toulouse, Castanet-Tolosan, France, (2)Department of Botany/University of British Columbia, Vancouver, BC, Canada, (3)Institut des Sciences des Plantes - Paris-Saclay (IPS2), CNRS/INRA/Universities of Paris Diderot and Sorbonne Paris Cité, Gif sur Yvette, France, (4)French Plant Genomic Center (CNRGV), INRA, Castanet-Tolosan, France, (5)Institut de Recherche en Horticulture et Semences, IRHS, Agrocampus-Ouest, INRA, Université d'Angers, SFR 4207 QuaSaV, Beaucouzé, France, (6)Institute of Plant Sciences Paris Saclay, GIF SUR YVETTE, France

Deciphering the functional architecture of eukaryotic genomes is facilitated by recent breakthroughs in sequencing technologies. Thus, PacBio sequencing allowed us to generate a substantially improved genome assembly of *Medicago truncatula*, a legume model species notably for endosymbiosis studies. Using high-depth (> 100x) PacBio sequencing, as well as previous (Young et al., Nature 2011) and new (BioNano technology) optical maps, a highly contiguous reference genome of 430 Mb (termed Mt5.0) was generated in only 64 sequence contigs (including 3.59 Mb in 32 unanchored contigs), thanks to a meta-assembly protocol based on a combination of several assemblers. This enabled a thorough analysis of transposable elements and repeats, as well as the identification of new players involved in rhizobium-induced nodule development, with in particular 1,037 strongly upregulated long non-coding RNAs (lncRNAs). We also discovered that a substantial proportion (~35%) of the genes upregulated in nodules or expressed in the nodule differentiation zone co-localize in more than two hundreds physical gene clusters, here termed symbiotic islands. We found these islands to be highly enriched in lncRNAs and to display differentially both DNA methylation and histone marks. Epigenetic regulations and lncRNAs are therefore attractive candidate elements, in addition to gene duplications, for the functional organization of symbiotic genes in *M. truncatula*.

Pecrix et al. 2018 Nat Plants. doi: 10.1038/s41477-018-0286-7

Integrative web portal, including a M. truncatula genome browser:
https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/

W692: Legumes

Two Genomes Gone Global: Insights from Sequencing White Clover and its Progenitors

Andrew G. Griffiths[1], Roger Moraga[1], Stig Uggerhøj Andersen[2], Vikas Gupta[2], Marni Tausen[2], Matthew A. Campbell[3], Rachael Ashby[4], Istvan Nagy[5], Anar Khan[6], Craig Anderson[1], Benjamin K. Franzmayr[1], Kerry Hancock[1], Alicia Scott[1], Nick Ellison[1], Murray P. Cox[7], Torben Asp[8], Thomas Mailund[9] and Mikkel Heide Schierup[2], (1)AgResearch, Palmerston North, New Zealand, (2)Aarhus University, Aarhus, Denmark, (3)Department of Ecology and Evolutionary Biology, Santa Cruz, CA, (4)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (5)Aarhus University, Slagelse, Denmark, (6)AgResearch Invermay, Mosgiel, New Zealand, (7)Massey University, Palmerston North, New Zealand, (8)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, (9)Bioinformatics Research Centre, Aarhus, Denmark

Merging genomes via allopolyploidisation generates adaptive potential through increased genetic diversity and access to 'genomic toolboxes' from the contributing genomes. White clover (*Trifolium repens* L.) is an allotetraploid (2n=4x=32) forage legume found throughout temperate grasslands, and is derived from two diploid progenitors: *T. occidentale* and *T. pallescens*, each confined to markedly different coastal and montane niches, respectively.

Genome and transcriptome sequencing and subsequent assembly of this species complex has provided a wealth of data to gain insight into the genesis and evolution of white clover. The sequence data have confirmed the progenitors and, in contrast to many allopolylpoids, shown that the progenitor subgenomes within white clover have largely retained their integrity and gene expression activity. Furthermore, we show that this hybridization event occurred during the depths of the last glaciation at a time when the European progenitor ranges (coastal and montane) likely overlapped.

White clover, therefore, represents a clear example of allopolyploidy-facilitated niche expansion, where the two progenitor genomes expanded from disparate and highly specialized European habitats to a ubiquitous global presence. Perhaps underpinning this evolutionary success, we found high polymorphism levels in white clover, demonstrating diversity carry-over from its progenitors. Furthermore, we have also found evidence of tissue-specific expression switching between subgenome copies of genes involved in flavonoid biosynthesis, a key pathway involved in adaptive traits such as plant/microbial interactions. This genomic resource not only provides insight into this species, but also a deeper understanding of subgenome interactions that may be exploited in breeding programmes.

W693: Legumes

Phylogenetic and Phylogenomic Approaches to Decipher the Evolution of the Nitrogen Fixing Root Nodule Symbiosis

Jean Florent Paul Keller, Laboratoire de Recherche en Sciences Végétales, UMR5546 CNRS-UPS Pôle de Biotechnologies Végétales, Castanet-Tolosan, France
Keller J. [1], Vernié T. [1], San Clemente H. [1], Girou C. [1], Delaux PM. [1]
[1]*Laboratoire de Recherche en Sciences Végétales, UMR5546 CNRS-UPS Pôle de Biotechnologies Végétales*
24 Chemin de Borde Rouge, BP42617, 31326 Castanet-Tolosan

The nitrogen fixing root nodule symbiosis between plants and bacteria has a major importance in both natural ecosystems and agriculture. This symbiosis is found in four orders of plants (Fabales, Fagales, Cucurbitales and Rosales, forming the NFN clade) in which nodulating species are spread. Thus, one of the main question regarding the evolutionary dynamics of this symbiosis is to determine whether it appeared or has been lost multiple times during the evolution of the NFN clade. Recently, via a phylogenomics approach of nodulating and non-nodulating species in the NFN clade we found signatures of multiple and independent losses in the *NODULE INCEPTION* (*NIN*) gene that support the hypothesis of multiple losses of the NFN symbiosis across the NFN clade (Griesmann et al., 2018). However, given that essential gene such as *NIN* are also present in non-symbiotic families outside the NFN clade, question remains about the molecular changes that provided the basis for the NFN symbiosis to evolve in that particular clade. The aim of our work is to identify the genetic elements that could explain the evolution of symbiosis in the NFN clade. Our approach is based on phylogenomics and phylogenetics study of nodulating and non-nodulating species to detect specific elements related to the nitrogen fixing symbiosis. We are developing a new genome database (more than 100 genomes) dedicated to symbiosis evolution investigation. This database also contains redefined othology relationships between species with different inflation parameters, taking into account the evolutionary dynamics of genes and gene families.

Griesmann et al., *Science,* https://doi.org/10.1126/science.aat1743.

W694: Linkage and Deletion Mapping

A Novel Software to Detect Homozygous Deletions

Jiangqi Wen, Liang Sun, Yinbing Ge, Andrew Bancroft and Xiaofei Cheng, Noble Research Institute, Ardmore, OK

Deletion mutagenesis has been widely used for forward and reverse genetics studies in functional genomics. Traditionally, the costly, time-consuming map-based cloning is used to locate causal deletions in deletion mutants. In recent years, comparative genomic hybridization (CGH) has been used to speed up and scale up the lesion identification in deletion mutants. However, limitations of low accuracy and sensitivity for small deletions in the CGH approach are apparent. With the next generation sequencing (NGS) becoming affordable for most users, NGS-based bioinformatics tools are more appealing. Although several deletion callers are available, these tools are not very efficient in detecting small deletions. Population-scale deletion callers that can identify both small and large deletions are rare. We were motivated to create a population-scale deletion detection tool, called FNBtools, to identify homozygous causal deletions in mutant populations by using next generation sequencing data. FNBtools is a tool to call deletions from NGS data at a population-scale and to achieve high accuracy at different levels of coverage. In addition, FNBtools can detect both small and large deletions with the ability to identify unique deletions in a mutant pool by filtering deletions that exist in a wild-type or control pool. FNBtools outperforms four existing popular deletion callers in detecting small deletions at different coverage levels. To better study identified deletions, FNBtools is also able to visualize all identified deletions in a genome-wide scope by using Circos. We applied FNBtools to analyze a salt-tolerant mutant in Medicago truncatula and identified the unique deletion locus that is tightly linked with this trait. The causal deletion in the mutant was confirmed by PCR amplification, sequencing and genetic linkage analyses. FNBtools can be used for homozygous deletion identification in any species with reference genome sequences. FNBtools is publicly available at https://github.com/noble-research-institute/fnbtools.

W695: Linkage and Deletion Mapping

Creating Linkage Maps using Traveling Salesperson Problem Solvers with TSPmap

J. Grey Monroe[1], Zachariah Allen[1], Paul Tanger[1], Jack L. Mullen[1], John Lovell[2], Brook T. Moyers[1], Darrell Whitley[1] and John McKay[1], (1)Colorado State University, Fort Collins, CO, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL

Advances in nucleic acid sequencing technologies have led to a considerable increase in the number of markers available to generate genetic linkage maps. However, genetic map construction from these large marker datasets can be computationally prohibitive and error prone. We recently developed *TSPmap*, a method which implements both approximate and exact Traveling Salesperson Problem solvers to generate linkage maps. We found that for simulated datasets with large numbers of genomic markers (e.g. 10000) and in multiple population types generated from inbred parents, *TSPmap* can rapidly produce high quality linkage maps with low sensitivity to missing and erroneous genotyping data. We also find that *TSPmap* can create reliable linkage maps from real genotype data in several species. *TSPmap* is open source and freely available as an R package. Here we present the background, implementation, and highlight results of using this approach for linkage map construction to facilitate its use by others to create high quality maps using a large number of genomic markers.

W696: Linkage and Deletion Mapping

Linkage Mapping with Low Coverage High-Throughput Sequencing Data in Outcrossed Populations

Timothy P. Bilton[1,2], Matthew R. Schofield[1], Michael A. Black[3], David Chagné[3,4], Phillip L. Wilcox[1] and Ken G. Dodds[2], (1)University of Otago, Department of Mathematics and Statistics, Dunedin, New Zealand, (2)AgResearch, Invermay Agricultural

Centre, Mosgiel, New Zealand, (3)University of Otago, Department of Biochemistry, Dunedin, New Zealand, (4)The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand
High-throughput sequencing methods (e.g., genotyping-by-sequencing, restriction site-associated DNA sequencing, exome capture sequencing) provide cost- and time-efficient methods for performing SNP discovery and genotyping and are increasingly being utilized in a wide range of species. These sequencing methods allow for substantially more markers than previous technologies, providing opportunities for building high-density genetic linkage maps. However, constructing genetic maps using sequencing data is complicated by the presence of sequencing errors and genotyping errors resulting from missing parental alleles due to low sequencing depth. Linkage maps are highly sensitive to these types of errors and even low error rates can lead to substantial map inflation.
We present a new software package GUSMap for performing linkage mapping in F1 biparental outcrossed (full-sib) populations using sequencing data. The novelty of GUSMap is that it accounts for the errors associated with sequencing data using a probabilistic approach, which enables construction of linkage maps with low-coverage data without *ad-hoc* data correction. We discuss the workflow in GUSMap for performing linkage mapping and highlight some of its functionality. We also apply GUSMap to a sequencing dataset consisting of a F1 population of mānuka plants. Results show that GUSMap is able to construct high-density linkage maps without large map inflation. GUSMap is publicly available at https://github.com/tpbilton/GUSMap.

W697: **Linkage and Deletion Mapping**
Bioinformatics of Ultra-High Density Mapping: How to Avoid Map Length Inflation Despite Genotyping Errors and Missing Data
Abraham Korol, The Institute of Evolution, University of Haifa, Haifa, Israel
Any organism of economical or ecological importance has (or will soon) become a target for genetic mapping, which is a prerequisite for most genomic practical applications. Moreover, in the last decade genetic mapping emerged as an efficient tool for structural genomics (including genome sequence assembly), studies of meiosis and recombination, evolution of genome organization, comparative genomics, etc. The rapidly deceasing genotyping cost warrants broad adoption of high-density mapping in many fields of biology. However, quite unexpectedly, it became obvious that the genetic lengths of a considerable part of the published high-density linkage maps for many species exceed the expectations (sometimes several-fold) based on the previous genetics and cytogenetics studies. Quite often, these discrepancies are simply ignored, or authors opt to perform either “map rescaling” or to choose “justified best rescaling parameters” to get more consistent estimates of map length. For some of the broadly used bioinformatics tools, the aforementioned map length inflation grows with increasing number of markers. Here we describe the application of an analytical system (MultiPoint-uld) for ultra-high density map construction that shows just an opposite trend: the map quality increases with increase in the size of the genotyping dataset. It considers cosegregating markers as more trustable candidates to be included into the primary skeletal map, followed by several cycles of resampling-based assessment of local map stability, removal of inconsistent markers and skeletal map saturation by adding carefully selected singleton markers. The efficiency of the system was tested by intensive simulations. Re-analysis of the publicly available data underlying the published inflated maps, allowed us to considerably reduce the map lengths without losing map coverage and collinearity with the published genome sequence.

W698: **Linkage and Deletion Mapping**
How to Anchor Genomes with Linkage Maps
Pasi M Rastas, University of Helsinki, Institute of Biotechnology, Helsinki, Finland
Typically (de novo) genome assemblies contain errors and are fragmented, making it difficult to know how the assembled contigs are physically located with respect to each other. Linkage mapping is one of the best tools to detect errors and anchor assembled contigs within chromosomes. Even a mapping cross of ten individuals can be used, but with more individuals and markers, more contigs can be corrected and anchored. Linkage maps with more markers than recombinations have multiple markers at most map positions. This will anchor contigs more reliable by locally pinpointing each recombination. However, most linkage mapping tools are not well suited for very large number of markers nor individuals.
Recent state-of-the-art linkage mapping software Lep-MAP3 is optimal for large data possible originating from low coverage whole genome sequencing. Here we discuss how Lep-MAP3 maps can be used to anchor genomes and show some real examples of the obtained (almost) completely anchored genomes. We also show that anchoring is possible for polyploid species where genome assembly is very difficult.

W699: **Linkage and Deletion Mapping**
IBD Estimation and QTL Mapping in Polyploids
Peter M Bourke, Wageningen University & Research, Wageningen, Netherlands

W700: **Maize**
Estimating the "Unknown Unknowns": Incorporating Environmental Data *via* Physiological Modeling and Genomic Prediction
Addie Thompson, Michigan State University, East Lansing, MI

W701: **Maize**
European Flint Genomes Complementing the Maize Pan- and Core-Genome
Georg Haberer[1], **Manuel Spannagl**[1], Nadia Kamal[1], Iris Fischer[1], Alevtina Ruban[2], Nina Opitz[3], Michael Seidel[1], Caroline Marcon[3], Heidrun Gundlach[1], Andreas Houben[2], Thomas Presterl[4], Milena Ouzunova[4], Frank Hochholdinger[3], Klaus F.X. Mayer[1], Eva Bauer[5] and Chris-Carolin Schön[5], (1)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (2)Leibniz Institute of Plant Genetics and Crop Plant

Research (IPK), Gatersleben, Germany, (3)INRES Crop Functional Genomics, University of Bonn, Bonn, Germany, (4)KWS SAAT AG, Einbeck, Germany, (5)Technical University Munich, Freising, Germany
Maize (*Zea mays*) is one of the most important and widely grown crops worldwide. It exhibits an enormous genetic diversity and can be grouped into a variety of major germplasm groups used in maize breeding including flint, dent, sweet and pop corn. All currently available reference genome sequences are focusing on US dent types (e.g. B73). Here, we report the analysis of reference genomes for four European flint lines: three elite lines (F7, EP1 and DK105) that are important founders of European breeding programs, and a doubled haploid (DH) line derived from Petkuser, a European landrace. All lines were assembled to pseudo-chromosomes with scaffold N50 ranging from 6 to 10 Mb using the DeNovoMagic™ pipeline. The high consistency between the physical and genetic map derived from a cross of PH207 (dent) and EP1 (flint) demonstrated the excellent quality of the assemblies. Comparative studies and whole-genome alignments of two dent (B73, PH207) and our four flint references revealed line- and germplasm-specific as well as core genomic regions shared by all six lines. The core genome clusters around low recombinogenic, (peri-)centromeric regions in the maize genome. Detailed inspection of the gene content suggests that the genic presence/absence variation (PAV) in maize may have been overestimated by recent studies due to e.g. missing or low quality gene models. Thus, genic PAVs seem to have a minor contribution to germplasm differences while repeats account for most of the observed variation and breakpoints of whole genome alignments. In support of this, *in situ* hybridizations revealed a high plasticity of knob regions within our lines. In contrast to genic variation, reprogramming of the regulatory code seems to discriminate flint and dent germplasms to a greater extent. We identified several hundred differentially expressed genes in a panel of 12 genotypes of each flint and dent, including well characterized genes, such as genes encoding for components of the light harvesting complex or involved in starch biosynthesis or degradation. Detailed genomic and transcriptomic analysis of economically important storage proteins greatly illustrates the benefits of multiple reference genomes.
Funding acknowledgements: Bavarian State Ministry of the Environment and Consumer Protection (BayKlimaFit; http://www.bayklimafit.de/) and the German Federal Ministry of Education and Research (BMBF; MAZE; http://www.europeanmaize.net/)

W702: **Maize**
Division Plane Orientation in Plant Cells
Carolyn Rasmussen, University of California, Riverside, Riverside, CA

W703: **Maize**
Biostatistical Tools in Breeding: Predicting Genotypes from Satellites to Genes
Lucia Gutierrez, University of Wisconsin - Madison, Madison, WI

W704: **Maize**
Breeding Maize with Optimized Root Systems
Philip N. Benfey, Duke University and Howard Hughes Medical Institute, Durham, NC
Underground Exploration
How roots explore their soil environment determines their ability to acquire nutrients and water. Very few genes controlling root growth in soil are known, primarily because of the difficulty of observing the underground environment. From mutagenized populations of rice, we have identified a gene that controls 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 find biopores in soil horizons. We have also identified novel signaling compounds that modulate root system architecture in dicots and monocots. Finally, at Hi Fidelity Genetics we have developed an inexpensive device, the RootTracker, that can monitor maize root growth in soil over time.

W705: **Mammalian Cloning. Process, Progress and Pitfalls, Practicality, Preservation and Potential of Cloning and other Advanced Genetic and Reproductive Technologies**
What has Changed Since Dolly, the Sheep, was Cloned? Introduction to Mammalian Cloning Workshop.
Heather Koshinsky, STIER, El Cerrito, CA

W706: **Mammalian Cloning. Process, Progress and Pitfalls, Practicality, Preservation and Potential of Cloning and other Advanced Genetic and Reproductive Technologies**
What Is Cloning? History, Terminology, Groundwork.
Irina Polejaeva, Utah State University, ADVS, Logan, UT

W707: **Mammalian Cloning. Process, Progress and Pitfalls, Practicality, Preservation and Potential of Cloning and other Advanced Genetic and Reproductive Technologies**
Broken Conceptus-Maternal Communication and the Consequences for Cloned Animals
Fernando Biase, Auburn University, College of Agriculture, Auburn, AL
Interactions between the embryo and endometrium at implantation are critical for the establishment and progression of pregnancy. These reciprocal actions involve the exchange of paracrine signals that govern implantation and placentation. The initiation of this signaling is triggered by key factors produced by the conceptus, which are translated by the endometrial cells into actions that will condition the trajectory of embryo development, as well as resulting progeny phenotype. In mammalian species, including human, rodents and ruminants, the delicate balance in embryonic-maternal communication is affected by the embryos' source (i.e.: natural mating, artificial insemination (AI), in vitro fertilization or somatic cell nuclear transfer) and by the sensor-driven properties of the endometrium defined by intrinsic maternal factors (i.e.: maternal metabolism, ageing) and environmental perturbations (i.e.: pathogens, nutrition). We have recently identified that at gestation day 18 in cattle, nearly all (98%) of the genes expressed in the AI derived the conceptus' extraembryonic tissue (EET) have a strong quantitative association with genes expressed in the endometrium's caruncular (CAR, 85% genes) tissues. Similarly, all 9,548 of the genes expressed in EET have a

quantitative association with genes expressed in the endometrium's intercaruncular (ICAR, 89% genes) tissues. We noted that in AI derived pregnancies genes that are highly co-expressed between EET and CAR form a unique blueprint that distinguishes individual pregnancies. By comparison, pregnancies initiated by the transfer of cloned blastocysts show a high degree of dysregulated gene expression. At gestation day 18, nearly 50% of the genes have altered expression profiles in cloned conceptuses compared to those produced by AI. At this stage, the endometrium already detects this aberrant expression and a few dozen genes have their expression derailed by harboring a cloned conceptus. By gestation day 34, a cattle conceptus has two main EETs, namely: chorion and allantois. Impressively, the chorion is very similar in clones compared to AI counterparts. On the other hand, the allantois has over 200 genes with dysregulated expression in cloned conceptuses, including genes responsible for vascularization, which is likely a major etiology of hydrallantois. Focusing on endometrial differences at gestation day 34, over 36% of the genes expressed in ICAR tissues dysregulated expression. Comparative co-expression analysis demonstrated that the dysregulated expression in pregnancies initiated by a cloned conceptus has massive impact on the interaction with the endometrium at the level of gene expression. The blueprint formed by highly co-expressed genes between AI-derived conceptus and endometrium is not established in cloned pregnancies; and the lack of interaction at the gene regulatory level is likely to have major impact on placenta malformation and consequently, reduced survivability of clones.

W708: Mammalian Cloning. Process, Progress and Pitfalls, Practicality, Preservation and Potential of Cloning and other Advanced Genetic and Reproductive Technologies

Can Advanced Genetic and Reproductive Technologies Really Contribute to Conservation Species and their Genetic Diversity?

Oliver Ryder, San Diego Zoo Institute for Conservation Research, Escondido, CA

Somatic cell nuclear transfer (SCNT) cloning has been discussed as offering potential for restoring genetic variation that has been lost in small populations. As retention of existing gene pools (standing variation), including potential for adaptation and reducing homozygosity due to identity-by-descent (inbreeding) are major conservation goals, progress in development and application of SCNT is desirable as an addition to the toolkit of methods to conserve biodiversity. However, the promise of SCNT has yet to be fulfilled as a functional management option, with notable exceptions. Assisted reproductive technologies (ART), such as artificial insemination, in vitro fertilization, embryo transfer, and germ-line chimeras also offer promise and have been put into limited use in endangered species conservation efforts. Nonetheless, to address the urgency of loss of biodiversity at the species level and at the intraspecific level through diminution of genome-wide genetic variation, strategies for maintaining and restoring genetic variation need to be expanded.

First and foremost, the role of viable cells for genetic rescue with the aim of maintaining or restoring genetic variation for current and future efforts to ensure population sustainability should be emphasized. Gene editing may very well play a significant role, but this will best be done in cells of the same species; this will be possible if tissue cultures have been established and cryobanked in facilities like San Diego Zoo Global's Frozen Zoo® and elsewhere across nations, regions, and continents.

Fibroblast cells have been reprogrammed to become induced pluripotent stem cells (iPSC) for a diverse but limited number of vertebrate taxa. The potential for iPSC to contribute to biodiversity characterization and conservation efforts is enormous, proven in principle in experiments with the laboratory mouse, but otherwise still speculative, although conservation-focused initiatives are underway. One such initiative is the northern white rhinoceros genetic rescue effort. The rationale for the project was developed at an international workshop in Vienna in 2015 and described in a publication (Saragusty, et al., Zoo Biology 35:289; 2016). Notable progress has been made in the interim period.

The northern white rhinoceros initiative and other efforts to use advanced genetic and reproductive technologies may alter existing paradigms for species conservation and preserve a wider range of options for future interventions to save species from extinction, if viable cell cultures have been saved.

W709: Mammalian Cloning. Process, Progress and Pitfalls, Practicality, Preservation and Potential of Cloning and other Advanced Genetic and Reproductive Technologies

Does Cloning Face Competition from Other Cell Based Methods

Bruce Whitelaw, The Roslin Institute, Midlothian, United Kingdom

Genome editing technology is set to transform the livestock breeding industry. We already see genome edited animals that address pressing animal welfare concerns, such as Polled cattle produced by Recombinetics, and spectacular progress in producing disease resistant animals, illustrated by the PRRSV resistant pigs produced by Genus. With the tried and tested methods of direct zygote injection being revitalised by genome editing technology and the wide spread use of SCNT as delivery routes much progress has been achieved, while in poultry it is through the PGC route. Although these methods remain the three mainstays by which to produce genome edited livestock, other cell-based platforms are emerging focussing on SSC and ESC. Simply put, we now have a plethora of delivery routes for producing genome edited livestock.

W710: Meiotic Recombination

Towards Harnessing Meiosis in Barley

Stefan Heckman, IPK Gatersleben, Gatersleben, Germany

W711: Meiotic Recombination

Recombination-Mediated Generation of Novel Synaptonemal Complex Alleles Promotes Meiotic Adaptation to Whole Genome Duplication in *Arabidopsis lyrata*

James David Higgins, University of Leicester, Leicester, United Kingdom

Paul J. Seear, Catherine L. Gregory, Darren Heavens, Roswitha Schmickl, Levi Yant and James D. Higgins

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 autotretaploid *Arabidopsis arenosa,* thus promoting meiotic stability and we are now investigating whether parallel evolution is taking place in the closely related *Arabidopsis lyrata*. Using a combination of next

generation and Sanger sequencing we have determined that there is extensive gene flow of synaptonemal complex (SC) genes between *A. lyrata* and *A. arenosa* autotetraploids. In addition, recombination events such as gene conversion have generated novel chimeric SC alleles (*ASY1* and *ZYP1*) and unequal crossing over (*ASY3*) that have been selected to promote meiotic stability by reducing numbers of crossovers and altering position. Interestingly, the majority of the modifications affect putative phosphorylation sites in the proteins and may be regulated by a small number of kinases active during meiosis such as ATM, CDKA and CKII . We have phenotyped individual plants for meiotic stability collected from the *A. arenosa* and *A. lyrata* hybrid zone in central Europe and are currently in the process of genotyping these plants to determine the quantitative effect of the individual alleles.

W712: **Meiotic Recombination**
Genetic Architecture of Genome-Wide Recombination Rate Variation in Wheat Revealed by Nested Association Mapping
Katherine Jordan, Kansas State University, Manhattan, KS

W713: **Meiotic Recombination**
Using Machine Learning to Dissect Recombination Landscape
Minghui Wang, Bioinformatics Facility, Cornell University, Ithaca, NY

W714: **Meiotic Recombination**
Elevated Temperature Increases Meiotic Crossover Frequency *via* the Interfering Pathway in *Arabidopsis thaliana*
Jiyue Huang, UNC-CH, Chapel Hill, NC

W715: **Meiotic Recombination**
Detecting *de novo* Homoeologous Recombination in *Brassica napus*
Erin Higgins[1], Wayne Clarke[1], Elaine C. Howell[2], Susan Armstrong[3] and Isobel Parkin[1], (1)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (2)University of Birmingham, Birmingham, United Kingdom, (3)University of Birmingham, West Midlands, United Kingdom
Brassica napus is a recent allotetraploid, formed from the hybridisation of the A genome of *Brassica rapa* and the C genome of *Brassica oleracea* approximately ten thousand years ago. Sequencing of a reference *B. napus* genome revealed several homoeologous recombination events between the A and C genomes that have become fixed in modern cultivars due to their association with important adaptations that were instrumental in establishing oilseed rape as a major worldwide crop. Molecular markers and whole genome resequencing have shown that these homoeologous recombination events continue to occur but the rate, distribution and size of events has been difficult to measure. Testcross F_1 populations from 11 spring-type *B. napus* lines were analysed using a Brassica 60K Illumina Infinium SNP array and reciprocal gain and loss of alleles between A and C genome in the F_1 individuals was used to measure the rate of homoeologous recombination in the parental lines. There was significant variation in the level of recombination between testcross families and events were biased to sub-telomeric regions. Events varied in size from less than 100 kb to entire chromosomes and aneuploids were observed in almost 5% of gametes. The high level of *de novo* homoeologous recombination indicates that it is a continuing source of variation in established *B. napus* lines and further understanding and potential manipulation of this phenomenon offers a novel mechanism for increasing genetic variation in this important oilseed crop.

W716: **NCBI Genome Resources**
Submission of Genomes to GenBank
Karen Clark, National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD
Because of improvements to the submission and processing steps, the process of submitting eukaryotic genomes to GenBank has become easier, especially for simple genome assemblies. For example, batch submissions of fasta sequences are accepted, and some simple contamination can be removed by the GenBank staff. Improvements have also been implemented for more complicated genome assemblies, eg annotated eukaryotes or diploid assemblies. This talk will discuss some of the recent improvements in the Submission Portal and details of some of the validations that are performed before a genome is accepted.

W717: **NCBI Genome Resources**
GEO Submissions and Usage
Steve Wilhite, NIH/NLM/NCBI, Bethesda, MD
The NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) stores comprehensive gene expression and epigenomic data sets that are generated using both microarray-based and sequence-based technologies (e.g., RNA-seq, ChIP-seq, single-cell-seq, methyl-seq). Data sets are submitted to GEO primarily by researchers who are publishing their results in journals that require original data to be made freely available for review and analysis. This talk will provide an overview of GEO database content, a step-by-step guide on how to submit your data to GEO, and a brief overview of tools to help users search and analyze GEO data.

W718: **NCBI Genome Resources**
From Annotation to Visualization: Exploring Genes and Genomes with NCBI Tools
Eric Cox, Francoise Thibaud-Nissen, Terence D. Murphy and The Eukaryotic Genome Annotation Team, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD
NCBI serves the agricultural genomics community by providing annotation of public genome assemblies, and the resulting products of annotation, including RefSeq genomic, transcript and protein sequences. This talk will provide an overview of our annotation pipeline and some of the ways our users can explore the resulting data, including a new and improved search interface that delivers more relevant results, as well as NCBI's genome browser, the Genome Data Viewer, and the Gene database.The NCBI Eukaryotic Genome Annotation Pipeline

(https://www.ncbi.nlm.nih.gov/genome/annotation_euk/) has been used to annotate nearly 500 organisms, including animals of agricultural importance. The pipeline provides content for various NCBI resources including RefSeq sequence databases, Gene, BLAST databases and the Genome Data Viewer. NCBI's Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) is an advanced genome browser that allows users to visualize genome annotations provided by NCBI's annotation pipelines, submitters, and external groups such as Ensembl in an interactive context. Genome Data Viewer provides access to a diverse set of tracks, including BLAST alignments, expression, comparative genomics, and more. In addition, users can access data from track hubs or upload their own tracks in a variety of formats. NCBI's Gene database (https://www.ncbi.nlm.nih.gov/gene) centralizes gene-related information into individual records. Many different types of gene-specific data are connected to the record including sequence accessions, nomenclature, genomic location and organization, publications, gene products and their attributes, expression, interactions, pathways, homology, and useful links to databases both internal and external to NCBI. By providing a better understanding of the genome annotation process, the resulting data, and ways to access this data, this talk should help users to use NCBI resources more effectively in pursuit of their animal genomics research goals.
This research was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

W719: **NCBI Genome Resources**
Programmatic Access to Genomic Data: E-Utilities and FTP
Vamsi K. Kodali, Francoise Thibaud-Nissen, Terence D. Murphy and The Eukaryotic Genome Annotation Team, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD
NCBI is home to a large amount of biological sequence and annotation data, most of which is accessible for public consumption. While the web-based frontends provide useful access to this data in a piecemeal manner, some users may want to leverage big data tools to analyze a large number of these at once. I will present a brief introduction to the data that is available for download from the NCBI FTP site and how users can use E-utilities, specifically the EDirect tools on the unix command line, to query Entrez, fetch data and extract specific information from them in bulk.

W720: **NCBI Genome Resources**
NCBI Resources for Phyletically-Defined Next Generation Analysis in and out of the Cloud (a.k.a. Cool New Stuff!)
Ben Busby, NIH/NCBI, Bethesda, MD
For years, you've enjoyed taxonomic slicing and dicing of your favorite NCBI BLAST databases on the web. If you've been a bit disappointed that this functionality doesn't extend to the lightning fast workflows you've built on the command line, we've heard your call and the taxonomically aware BLASTDBv5 is here! It's also on the cloud with a lot of other BLAST functionality! Need to define orthologs? Map deleterious variation across species? Find specific isoforms in your species, genus, family, clade or order? Want to do it in the cloud? We'll show you how.
Back at the ranch -- and across the United States -- we've also been running hackathons at a breakneck pace (there's a virus hunting hackathon right before PAG)! You've probably heard of this by now, and if you want to get involved, check out https://biohackathons.github.io! What you might not know is that the NCBI-hackathons program ran a pilot Visiting Bioinformaticians Program over the last couple years. One of the visiting bioinformaticians, Jose Die, built GeneHummus, an automated pipeline to study plant gene families based on protein domain organization. https://github.com/NCBI-Hackathons/GeneHummus. Check out the Digital Resources talk about it this on Wednesday -- spoiler alert: you can also get expression of these families using R. Interested in graph genomes? Check out and contribute to Chaochih Liu's progress on https://github.com/NCBI-Hackathons/HummusGraph, because haplotypes aren't just for human genome researchers!
Moving forward, we will be generating contigs for many of the datasets in the sequence read archive, and generating count summaries for many of our RNA-seq datasets as well as making even more resources available on the cloud!

W721: **New Approaches for Developing Disease Resistance in Cereals**
Mining Vavilov Wheat Landraces for Seedling Stripe Rust Resistance
Raghvendra Sharma, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia; Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia
Wheat is one of the most important staple food and agricultural crop worldwide. Along with meeting the needs of the growing world population, the wheat production in major growing regions is challenged by the emergence of widely virulent fungal pathogens. These include aggressive races of stripe rust pathogen (*Puccinia striiformis* f. sp. *tritici*) that have appeared in wheat fields in Australia, India, China, Europe, Africa, the Middle East and New Zealand. Routinely, stripe rust disease is effectively managed using genetic resistance. However, due to the lack of genetic diversity, many current wheat varieties are susceptible or vulnerable to the recently evolved stripe rust races. Therefore, new sources of resistance from unexplored materials, such as landrace and wild wheat collections are required urgently. Hence, the present study aims to identify and characterize novel genes for resistance in landraces collected by famous Russian botanist Nikolai Ivanovich Vavilov in the early 1900s. In initial screening of 300 accessions with prevalent Australian *Pst* races, we identified six highly resistant landrace accessions. Segregating mapping populations were developed to identify the resistance genes and linked molecular markers that will permit transfer of resistance in elite breeding materials. Furthermore, we aim to isolate the genes by recently developed mutagenesis and resistance gene enrichment and sequencing (MutRenSeq) approach. Along with the use of linked or gene specific markers for rapid gene selection, we will be using speed breeding facilities to accelerate the development of resistant wheat varieties.

W722: **New Approaches for Developing Disease Resistance in Cereals**
Understanding and Exploiting Disease Resistance in Wild Wheats
Brande B.H. Wulff, John Innes Centre, Norwich, United Kingdom

W723: **New Approaches for Developing Disease Resistance in Cereals**
Using AgrenSeq to Identify Wheat Blast Resistance Genes

Paul Nicholson, John Innes Centre, Department of Crop Genetics, Norwich, United Kingdom

W7**24: New Approaches for Developing Disease Resistance in Cereals**

Genomic Prediction for Leaf Rust Resistance in Hybrid Wheat with High Resolution Exome Sequencing Data

Yusheng Zhao, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Genome-wide prediction promises to boost selection gain for important agronomic traits. It is well known that relatedness between training and test populations strongly affects the prediction accuracy of single-cross hybrids, which makes it challenging to predict new hybrids derived from novel parental lines. In our study, we draw advantage of exome capture sequencing to establish powerful genome-wide prediction models for leaf rust resistance in a population 1,574 wheat hybrids and their 133 parental lines. We extended genomic best linear unbiased prediction considering additive and dominance effects (AD-GBLUP) towards a model allowing for sub-population-specific marker effects (GSAD-BLUP: general and sub-population-specific AD-GBLUP). Using the empirical hybrid wheat data, we showed that applying GSAD-BLUP increased the prediction ability by 27% for hybrids derived from independent parental lines when compared with AD-BLUP modeling one additive and dominance effect for all sub-populations. In addition, we also emphasized the potential of GSAD-BLUP to improve genome-wide hybrid prediction for scenarios of genetically diverse parental populations. The experimental findings were further substantiated with computer simulations. Because of the advantages of the GSAD-BLUP model in dealing with hybrids from different parental populations, it may also be a promising approach to boost the prediction ability for hybrids bred using genetically diverse heterotic groups.

W7**25: New Approaches for Developing Disease Resistance in Cereals**

Genetic Dissection of Disease Resistance to Goss' Wilt in Maize

Sanzhen Liu, Kansas State University, Manhattan, KS

Goss's wilt (GW) of maize, caused by the Gram-positive bacterium *Clavibacter michiganensis* subsp. *nebraskensis*, has spread in recent years throughout the Great Plains, posing a threat to production. The genetic basis of plant resistance is unknown. Here, a simple method for quantifying disease symptoms was developed and used to select cohorts of highly resistant and highly susceptible lines known as extreme phenotypes (XP). Genome-wide association (GWAS) and copy number variation analyses (CNV) of bulked XP sequences, along with standard GWAS and quantitative trait loci analyses of three bi-parental populations, revealed GW disease-associated genomic loci. Specifically, the *rp1* locus was associated with GW resistance based on XP-CNV. Multiple *Rp1* accessions with distinct *rp1* haplotypes in an otherwise susceptible accession exhibited hypersensitive responses to Cmn. The results will facilitate breeding strategies to control this emerging disease, and indicate that GW is an excellent system for genetically dissecting the interaction between plants and Gram-positive bacteria.

W7**26: New breeding technologies: Prospects and regulatory hurdles**

Current Regulation of NPBTs; A Global Map

Jochen Menz, Julius Kuehn-Institut, Quedlinburg, Germany

W7**27: New breeding technologies: Prospects and regulatory hurdles**

Current Applications of Genome Editing in Agriculture: A Systematic Map

Dominik Johannes Christian Modrzejewski, Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Institute for Biosafety in Plant Biotechnology, Quedlinburg, Germany

Plant breeding is a development process and new breeding methods have evolved continuously over time. Within few decades, genome editing techniques such as Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated proteins (CRISPR/Cas), Transcription Activator-Like Effector Nucleases (TALENs), Zinc-Finger Nucleases (ZFN), Meganucleases (MN) and Oligonucleotide-Directed Mutagenesis (ODM) have been developed enabling a precise modification of DNA sequences in many plant species.

In order to provide a comprehensive overview about the fast growing available evidence for the application of genome editing in plants a systematic map was conducted. A systematic map is based on a broad review question aiming to identify, collect and evaluate the available academic and grey literature in a systematic and transparent manner. The detailed determination and documentation of the data collection allows a consistent updating and supplementing of the existing literature.

The results of the systematic map cover the period between 1996 and May 2018 and show more than 1300 applications in model plants as well as in cultivated plants. The corresponding authors/ leading institutions came by a considerable margin from China (612) followed by the USA (487), Japan (92) and Germany (88). The major part of applications are related to rice (475), followed by Arabidopsis (214), tobacco (107), tomato (87) and maize (77). Altogether, in the period until May 2018 applications in 46 different model and cultivated plants were documented. In most of the studies (92%) mutations comparable to spontaneous mutations or undirected mutagenesis were induced by genome editing. Besides many basic research studies, 98 applications with regards to plants and general traits were allocated as market-oriented developments, including improved growth characteristics and yield improvement, improved food and feed quality, increased tolerance to abiotic and biotic stress and herbicide tolerance.

W7**28: New breeding technologies: Prospects and regulatory hurdles**

Regulation of NPBTs in the EU, with a Focus on the CJEU Ruling

Ulrich Ehlers, Federal Office of Consumer Protection and Food Safety, Berlin, Germany

In Europe, genome edited organisms are not explicitly regulated, genetically modified organisms (GMO), however, are. Of utmost importance in European gene technology legislation is the definition of a GMO in Article 2(2) of the Directive 2001/18/EC, according to which a GMO *"means an organism, with the exception of human beings, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination."*

That is the reason why the ruling of the Court of Justice of the European Union (CJEU) on the interpretation of said definition with regard to organisms created by targeted mutagenesis, i.e. genome editing, has been highly anticipated – especially as the correct interpretation of the GMO definition has been the subject of many legal debates.

The CJEU on July 25th 2018 in essence ruled that all organisms obtained by means of techniques/methods of mutagenesis have to be considered GMO and that organisms created by methods of genome editing methods do not fall under the existing exemption clause for organisms created by conventional methods of (untargeted) mutagenesis. Therefore in Europe currently all genome edited organisms – even if the editing is limited to a point mutation in the genome (SDN-1) – are governed by the same rules that apply to transgenic organisms, i.e. products of conventional genetic engineering.
The ruling has been widely criticized by many scientists. At least in Germany, a huge part of the large mainstream media also displayed criticism towards the CJEU ruling.
The implementation of the CJEU ruling is currently subject of intense discussions among the member states and the European institutions. There are many open questions. For example, it is questionable if the principle that no adventitious presence of unauthorized GMO in seeds is tolerated can be enforced, since mutations obtained by genome editing are principally undistinguishable from naturally occurring mutations. The implications for international law like the Cartagena Protocol or world trade rules are also unclear as for now.
In theory, there is a number of different legislative options to solve the problems created by the CJEU ruling. However, basically all of these require an amendment of the central Directive 2001/18/EC.
On November 13th 2018, the Scientific Advisory Mechanism (SAM), a high level group that provides independent scientific advice to the European Commission, issued a report in which it recommended to revise the existing GMO Directive to reflect current knowledge and scientific evidence, in particular on gene editing and established techniques of genetic modification.
Even if there was a necessary political majority among the EU Member States for such an amendment, it would probably be a rather lengthy procedure: The last amendment of Directive 2001/18/EC („opt-out“ Directive (EU) No. 2015/412) in the normal legislative procedure took about five years. Therefore, immediate changes seem unlikely, so that for now genome editing in Europe without any differentiation is being treated like conventional genetic engineering.

W729: New breeding technologies: Prospects and regulatory hurdles
Current Situation of Genome Editing Regulation in Japan
Masashi Tachikawa, Nagoya University, Nagoya, Japan

W730: New breeding technologies: Prospects and regulatory hurdles
Calyxt NPBTs Presentation
Chloe Pavely, Calyxt, INC., Roseville, MN

W731: Next Generation Gene Editing in Plants and Animals
Gene Editing: Historical Perspective
Dan Voytas, University of Minnesota, Saint Paul, MN

W732: Next Generation Gene Editing in Plants and Animals
Gene Editing: Tools and Technology Overview
Steve Jacobsen, University of California at Los Angeles/Howard Hughes Medical Institute, Los Angeles, CA

W733: Next Generation Gene Editing in Plants and Animals
Gene Editing: Plants Overview
Zach Lippman, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
With the recent widespread deployment of genome editing technologies in plants, we have entered an era of great excitement and enormous opportunity for rapidly advancing fundamental discoveries in plant biology and translating those discoveries to agriculture. Based on knowledge gained from exploiting CRISPR-Cas9 to reveal and harness mechanisms of flowering and flower production in tomato, I will present what is already possible and will likely become feasible in the coming decade of plant genetics and crop improvement. This includes our work on integrating discoveries in plant stem cell biology with agriculture to customize and optimize major productivity traits, and an approach we have developed to customize and fine-tune quantitative trait variation. Beyond these immediate applications, I will discuss a broader view on how efficiency in breeding could be improved through future developments in this technology, which could help pave the way to breaking yield barriers.

W734: Next Generation Gene Editing in Plants and Animals
Gene Editing: Animal Overview
A. Mark Cigan, Genus, DeForest, WI

W735: Next Generation Gene Editing in Plants and Animals
Panel Discussion Moderator
Alison Van Eenennaam, University of California, Davis, Davis, CA

W736: Next Generation Genome Annotation and Analysis
Comparative Annotation Toolkit (CAT) - Simultaneous Clade and Personal Genome Annotation
Ian T. Fiddes[1], Joel Armstrong[2], Mark Diekhans[2], Stefanie Nachtweide[3], Zev N. Kronenberg[4], Jason Underwood[5], Thomas Keane[6], Evan Eichler[7], David Haussler[2], Mario Stanke[3] and Benedict Paten[2], (1)10x Genomics, Pleasanton, CA, (2)Genomics Institute, University of California Santa Cruz and Howard Hughes Medical Institute, Santa Cruz, CA, (3)University of Greifswald,

Greifswald, Germany, (4)Phase Genomics, Seattle, WA, (5)Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, (6)EMBL EBI, Hinxton, United Kingdom, (7)University of Washington, Seattle, WA
The recent introductions of low-cost, long-read, and read-cloud sequencing technologies coupled with intense efforts to develop efficient algorithms have made affordable, high-quality de novo sequence assembly a realistic proposition. The result is an explosion of new, ultracontiguous genome assemblies. To compare these genomes, we need robust methods for genome annotation. We describe the fully open source Comparative Annotation Toolkit (CAT), which provides a flexible way to simultaneously annotate entire clades and identify orthology relationships. We show that CAT can be used to improve annotations on the rat genome, annotate the great apes, annotate a diverse set of mammals, and annotate personal, diploid human genomes. We demonstrate the resulting discovery of novel genes, isoforms, and structural variants—even in genomes as well studied as rat and the great apes—and how these annotations improve cross-species RNA expression experiments and enable analysis of single-cell RNA expression analysis on non-model organisms.

W737: Next Generation Genome Annotation and Analysis
Streamlining Submission of Genomic Assemblies and Annotations to NCBI with Genome Annotation Generator
Scott Geib, USDA-ARS, Hilo, HI
One of the most overlooked, yet critical, components of a whole genome sequencing (WGS) project is the submission and curation of the data to a genomic repository, most commonly the National Center for Biotechnology Information (NCBI). While large genome centers or genome groups have developed software tools for post-annotation assembly filtering, annotation, and conversion into the NCBI's annotation table format, these tools typically require back-end setup and connection to an Structured Query Language (SQL) database and/or some knowledge of programming (Perl, Python) to implement. With WGS becoming commonplace, genome sequencing projects are moving away from the genome centers and into the ecology or biology lab, where fewer resources are present to support the process of genome assembly curation. To fill this gap, we developed software to assess, filter, and transfer annotation and convert a draft genome assembly and annotation set into the NCBI annotation table (*.tbl*) format, facilitating submission to the NCBI Genome Assembly database. This software has no dependencies, is compatible across platforms, and utilizes a simple command to perform a variety of simple and complex post-analysis, pre-NCBI submission WGS project tasks. A workflow of data submission for whole genome and whole transcriptome datasets will be discussed using this tool, along with other platforms that are present for whole genome data submission.

W738: Next Generation Genome Annotation and Analysis
EMBLmyGFF3
Jacques Dainat, Uppsala University, UPPSALA, Sweden

W739: Next Generation Genome Annotation and Analysis
Pigeonomics: Using Whole Genome Data to Uncover the Genetic Bases of Morphological Variation in Domestic Pigeons
Elena F. Boer[1], Eric T. Domyan[2], Zev N. Kronenberg[3], Carson Holt[4], Mark Yandell[4] and Michael D. Shapiro[1], (1)School of Biological Sciences, University of Utah, Salt Lake City, UT, (2)Utah Valley University, Orem, UT, (3)Phase Genomics, Seattle, WA, (4)Department of Human Genetics, University of Utah, Salt Lake City, UT
Deciphering the genetic mechanisms of morphological variation remains a critical challenge in evolutionary and developmental biology. The domestic pigeon (*Columba livia*) offers a unique balance of extraordinary phenotypic variation and experimental accessibility, making it an outstanding system to study the genetic and developmental underpinnings of morphological variation. Artificial selection by pigeon fanciers has resulted in more than 350 breeds that collectively display tremendous variation in a variety of complex traits, including body size, beak morphology, feather color and patterning, and ornamental feathering. To identify genomic variants associated with phenotypic variation, we have generated a powerful "pigeonomics" toolkit, which includes a high-quality reference genome and ~150 resequenced whole genomes representing more than 50 diverse pigeon breeds. Our toolkit captures genomic variation at more than 200 million positions in the pigeon genome and has been used to successfully identify genomic regions associated with a variety of simple and complex morphological traits, including feather color and patterning, ornamental feathering, and beak size and shape. In combination with classical genetic and embryological approaches, our pigeonomics toolkit provides a powerful resource to identify precise genetic and developmental changes that underlie morphological variation.

W740: Next Generation Genome Annotation and Analysis
CLfinder-OrthNet: Creating Comparative Genomics Framework for Closely-Related Genomes using Co-Linearity Networks
Dong-Ha Oh, Louisiana State University, Baton Rouge, LA
This is a tentative abstract, I will update it before Dec 1st
The CLfinder-OrthNet pipeline (1) detects co-linearity among multiple closely-related genomes, (2) finds orthologous gene groups, and (3) encodes the evolutionary history of each ortholog group into an Ortholog Network (OrthNet). OrthNets connect orthologs with edges representing either the presence or absence of co-linearity between them. Each OrthNet encodes in its network topology the evolutionary history of an orthologous locus, including different modes of gene duplication, deletion, transposition, and combinations of them, occurred in a lineage or multiple lineages. Orthologous gene groups with the same evolutionary history can be retrieved by searching OrthNets with a network topology query.
As a proof-of-concept, we applied CLfinder-OrthNet to characterize gene transposition-duplication (*tr-d*) events among six Brassicaceae genomes, including those of *Arabidopsis thaliana* and two extremophytes, *Eutrema salsugineum* and *Schrenkiella parvula*. We identified subsets of lineage-specific *tr-d* events with signatures of selective retention and sub-functionalization in all six genomes. These included lineage-specific *tr-d* of genes that may be critical for the local adaptation of extremophytes, such as orthologs of *SALT TOLERANCE 32* and *ZINC TRANSPORTER 3*.
CLfinder-OrthNet offers a flexible toolset for systematic comparative studies of closely-related genomes. Beside the detection of all orthologs showing the same evolutionary history, the application includes but not limited to: (1) improving orthology inference assisted by co-linearity, (2)

identification of gene duplication or transposition events co-occurring with certain phenotypic traits among closely-related genomes, and (3) detection of truncated, split, and chimeric gene models based on co-linearity.
CLfinder-OrthNet is available at https://github.com/ohdongha/CL_finder

W741: **Non-coding RNA**

Long Non-Coding RNA Annotation in Chickens using Novel Iso-Seq Methods

Richard Kuo[1], Katarzyna Miedzinska[2], Mike McGrew[3], Adam Balic[3], Jacqueline Smith[4], Alan L. Archibald[5] and David W. Burt[6], (1)Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, (2)The Roslin Institute, Edinburgh, United Kingdom, (3)The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, (4)Roslin Institute, Edinburgh, United Kingdom, (5)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (6)The University of Queensland, St Lucia, Australia

Long non-coding RNA (lncRNA) are one of the least understood biotypes in transcriptomics. A combination of their lack of sequence conservation, lower expression levels, and mysterious composition motif have made them incredibly elusive. Due to the lack of external verification methods, full length transcript sequencing has become a useful approach for lncRNA identification. We used Iso-Seq sequencing in conjunction with novel cDNA library preparation methods to identify lncRNA within the chicken genome. Our cDNA library customization increased the efficiency of sequencing thus allowing us to create the most extensive annotation of lncRNA of any animal species. These results suggest that many of the prior assumptions on lncRNA characteristics may have been incorrect and thus led to biased approaches in lncRNA research.

W742: **Non-coding RNA**

An Atlas of Chicken Long Non-Coding RNAs gathering Multiple Sources and Expression across more than Twenty Tissues

Frédéric Jehl, INRA - Agrocampus Ouest, UMR 1348 PEGASE, Rennes, France

Long non-coding RNAs (lncRNAs) appear as key regulators of gene expression through numerous mechanisms. Yet, particularly in domesticated species like chicken, their annotation is incomplete and their expression in different tissues poorly characterized. Here, we used data from 4 public databases (Ensembl, NONCODE, NCBI, ALDB), a recent lncRNA catalogue, and genes models from 364 RNA-seq samples from our lab to generate an extensive annotation of the chicken genome aggregating most of the currently known lncRNAs, plus robust models of newly discovered lncRNA from our dataset. Our extended lncRNAs annotation comprises 30084 lncRNA loci versus 4640 in the Ensembl reference. These lncRNAs follow the same trends as those in the reference for their genomic localization, structure and expression. To further characterize these lncRNAs, we studied their expression in over 20 chicken tissues with minimum 6 biological replicates. This allowed us to explore expression variability of both lncRNAs and protein-coding genes (mRNAs) across tissues, and across replicates in each tissue. Across these chicken tissues, we found interesting differences in the tissue-specificity patterns between lncRNAs and mRNAs. We characterized the expressions of both types of genes in relation with (*i*) their tissue-specificity and/or (*ii*) their genomic localization relative to each other (convergent, divergent, antisense, distant or proximal) in order to highlight some lncRNA–mRNA couples with a role in the functions or the maintenance of a given organ in chicken. Finally, we started a conservation study of interesting lncRNA–mRNA couples across species. Project funded by ANR-13-ADAP and H2020-Feed-a-Gene projects.

W743: **Non-Seed Plants**

Progress in Streptophyte Algae Genomics and Genetics

Tomoaki Nishiyama[1], Hiroaki Kamada[2], Masahiro Kasahara[3], Hidetoshi Sakayama[4], Yuki Tsuchikane[5] and Hiroyuki Sekimoto[5], (1)Kanazawa University, Kanazawa, Japan, (2)Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan, (3)University of Tokyo, Kashiwa, Japan, (4)Kobe University, Kobe, Japan, (5)Japan Women's University, Bunkyo-ku, Japan

Chara braunii is a close relative of land plants and the draft genome was compared with other algae and land plants (Nishiyama et al. 2018). The species *C. braunii* is distributed world-side and two ecotypes of *C. braunii,* shallow-water and deep-water types*,* are recognized. Variants were identified between *C. braunii* strain S276 (shallow-water) and strain S277 (deep-water) and PCR-RFLP markers were designed. Genetic cross was established based on the segregation of upon the variants. A genetic map was constructed based on low depth sequencing of 96 progeny obtained by outcrossing S276×S277, which covered 1.6 Gb of the scaffolds in 14 linkage groups. Further collection of *C. braunii* in East Asia and South East Asia is on going to reveal genetic basis for the adaptation to diverse environments.
The earliest branch of streptophytes includes *Chlorokybus* and *Mesostigma*. *Chlorokybus* sp. NIES-160 was sequenced with PacBio to reach 84 Mb in 222 contigs with N50 contig length of 857 kb. *Mesostigma viride* NIES-475 was also sequenced but the nuclear genome requires some more effort while the richettial symbiont (Yang 2016) genome was complete.
The closest living relatives of land plants are Zygnematales, including unicellular desmids. *Closterium peracerosum-strigosum-littorale* complex plus (NIES-67) and minus (NIES-68) mating type strains were sequenced to obtain 360 and 340 Mb assemblies, with N50 contig length of 350 and 276 kb, respectively. Genetic transformation and CRISPR/Cas9 based gene targeting were established using native promoters including the U6 promoter, enlightening development of genetics in new organisms utilizing the genome sequence.

W744: **Non-Seed Plants**

***De novo* Genome Assembly and Annotation of *Sanionia uncinata* (Amblystegiaceae: Hypnales), a Pleurocarpous Moss Dominant in Antarctica**

Hyoungseok Lee[1], Hyun Park[1,2], Mi Young Byun[1], Mira Park[1], Jungeun Lee[1] and Ralf Reski[3], (1)Korea Polar Research Institute, Incheon, South Korea, (2)University of Science & Technology, Incheon, South Korea, (3)Plant Biotechnology, University of Freiburg, Freiburg, Germany

Mosses in Antarctica grow mostly in coastal areas and are expected to have developed various unique physiological/molecular mechanisms to survive in extreme environments. *Sanionia uncinata* (Amblystegiaceae: Hypnales) is a dominant moss species in the maritime Antarctic and considered as a good target to investigate genes associated with abiotic stress tolerance of mosses. It has several distinct characteristics when

compared to *Physcomitrella patens*, the first model moss species. First, *S. uncinata* is a pleurocarpous moss. Second, it belongs to the order Hypnales which contains the largest number of moss species. Third, it is an alpine species that lives in cold regions unlike *P. patens* mostly found in temperate regions. Here, we report the draft genome sequence of an Antarctic *S. uncinata*, obtained using third-generation PacBio sequencing technology. About 1 million reads were attained from four Sequel sequencing runs and merged together into a single dataset of 21 Gb. The *de novo* assembly produced 673 contigs with an N50 contig length of 2.18 Mb, and a total of 28,651 coding genes were inferred. Our dataset can be useful as a comparative genome for evolution and speciation studies for bryophytes, as well as for the analysis of molecular adaptation of plants to harsh environment.

W745: Non-Seed Plants
Sexually Antagonistic Selection in an Ancient Interaction between Moss and Microarthropods
Leslie M. Kollar, University of Florida, Gainesville, FL
A central goal in evolution is to understand the mechanisms that maintain genetic variation for fitness. Across much of the tree of life, males and females are differentiated in many non-reproductive traits, presumably because selection favors different trait optima in each sex. Thus, an allele that increases fitness in one sex can be deleterious in the opposite sex, causing a form of genetic conflict, sexual antagonism. The role of genetic conflict in maintaining variation for fitness depends upon the degree to which males and females respond similarly to an allelic substitution (i.e., the cross-sex correlation) and the difference in optimum phenotypes between the sexes, both poorly understood quantities. Here, I estimated the cross-sex correlation for many life history traits in the moss, Ceratodon purpureus using a common greenhouse experiment with 48 haploid-sibling families, each comprising at least three male and three female offspring. One life history trait was the volatile organic compound (VOC) profiles. Analogous to flowering plant-pollinator mutualisms, female C. purpureus gametophytes emit abundant VOCs to attract sperm-dispersing microarthropods, which significantly increase fertilization rates in moss. Male mosses produce fewer VOCs than female mosses, suggesting that VOC production may be costly. Next I plan to conduct competitive mating experiments in controlled mesocosms to identify traits linked with female and male reproductive success.

W746: Non-Seed Plants
Three *de novo* Draft Genomes from P*hyscomitrella*'s Close Relatives
Nico van Gessel[1], Daniel Lang[2], Peter Szovenyi[3] and Ralf Reski[1,4], (1)Plant Biotechnology, University of Freiburg, Freiburg, Germany, (2)Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany, (3)Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland, (4)BIOSS Centre for Biological Signalling Studies, Freiburg, Germany
For decades the Funariaceae have been targets for various biological studies covering the fields of classical genetics, evo-devo and molecular biology and with *Physcomitrella patens*, the family comprises the first established model species representing the early-diverging land plant lineage of bryophytes. While the whole family is characterized by profound differences in morphological complexity of the sporophytic generation, *P. patens* and *Funaria hygrometrica* constitute the respective end-points of a broad spectrum. Despite substantial differences in genome sizes and chromosome counts, comparative analyses of their genomes have indicated their chromosome history to be largely shared, including two rounds of whole genome duplications. However, mosses of the *Physcomitrium* genus display sporophytic morphologies intermediate between the extreme forms and some have been suspected to have undergone an additional genome duplication due to autopolyploidization or a hybrid origin, respectively. In order to gain further insights into the individual and shared genome evolution, we have assembled and annotated three novel draft genomes of this genus, namely *Physcomitrium eurystomum*, *Physcomitrium pyriforme* and *Physcomitrium sphaericum*. We will present first analyses on genome and gene family evolution with a focus on the extent of genome duplications.

W747: Non-Seed Plants
Pattern of Gene Expression during Shift in Ploidy in *Funaria hygrometrica*
Nasim Rahmatpour, University of Connecticut, Storrs, CT
The life cycle of all land plants consists of an alternation between a multicellular haploid gametophyte and a multicellular diploid sporophyte, linked by sexual reproduction and meiosis. These two generations differ in their ploidy and function: the gametophyte develops sex organs and gametes and the sporophyte produces spores via meiosis. The dramatic morphological, developmental and physiological differences between the two generations seems correlated with the shifts in ploidy between generations. However, ploidy alone does not determine function and this is evident from the occurrence of the polyploid series within plants. Furthermore, apospory and apogamy, two phenomena in which sporophyte and gametophyte can readily convert to each other without shift in ploidy, indicate that ploidy is not a determent factor in alternation of generations. Apogamy occurs through sporophyte development without fertilization, apospory is the ability of sporophyte to develop homoploid gametophyte. Apospory and apogamy are common well-known phenomena in bryophytes and ferns. Apogamy also occurs in angiosperms known as apomixis in which seeds are developed directly from maternal tissue without fertilization through inhibition of meiosis either by apospory or diplospory. In bryophytes, apogamy is more common in nature and both apospory and apogamy can be induced easily in laboratory conditions. For instance, apospory requires harvesting sporophyte at early developmental stage, cutting immature, spear-shaped sporophytes and place it on media. Here, we are studying the effect of ploidy on gene expression by contrasting transcriptomes of generations (haploid gametophyte, sporophyte and aposporous diploid gametophyte) that differ in function and or ploidy and examine how gene expression changes along with shift in ploidy and function. If ploidy shift is not necessary during alternation of generation, how would it effect on gene expression? Whether it has influence on gene expression or not and, if its effect interfering with normal moss development are the kind of questions we are seeking for an answer.

W748: Non-Seed Plants
The *Isoetes* Genome and the Evolution of Aquatic CAM Photosynthesis

Fay-Wei Li[1,2], Li-Yaung Kuo[1], Shane Poplawski[3], Ta-Chun Lin[4], Yao-Moan Huang[5], Tai-Chung Wu[4], Chun-Neng Wang[4], Wen-Yuan Kao[4] and Todd Michael[3], (1)Boyce Thompson Institute, Ithaca, NY, (2)Plant Biology Section, Cornell University, Ithaca, NY, (3)J. Craig Venter Institute, La Jolla, CA, (4)National Taiwan University, Taipei, Taiwan, (5)Taiwan Forestry Research Institute, Taipei, Taiwan
Since the publication of the first plant genome *Arabidopsis thaliana* in 2000, reference genomes for all major green lineages have been published, except for a few orphan clades such as *Isoetes* (Isoetales). *Isoetes* is one of the three lycophyte lineages, together with Selaginellales (*Selaginella*) and Lycopodiales (e.g. *Lycopodium*). *Isoetes* harbors an array of unique morphological and developmental characteristics, making it an important species to understand major transitions during land plant evolution. In addition, despite having an aquatic lifestyle, some *Isoetes* species can carry out crassulacean acid metabolism (CAM) photosynthesis, which is usually associated with xeric adaptation. Here I will present the genome of *Isoetes taiwanensis* sequenced using the Oxford Nanopore technology. We will also report on the ongoing research using *I. taiwanensis* as a model to study the gene regulatory network of aquatic CAM photosynthesis.

W7**49: Oats, Wild and Cultivated**
Genome Assembly and Phylogenomics in Diploid *Avena* Species
Peter J. Maughan[1], Rachel N. Walstead[2], Rebekah Lee[1], Jessica Schlueter[2], Shelvasha Burkes[2], Tim Langdon[3], Eric Jackson[4] and Rick N. Jellen[1], (1)Brigham Young University, Provo, UT, (2)University of North Carolina at Charlotte, Charlotte, NC, (3)IBERS, Aberystwyth University, Aberystwyth, United Kingdom, (4)25:2 solutions, Rockford, MN
Common oat (*Avena*) has held a significant place within the global crop community for centuries; although its cultivation has decreased over the past century, its nutritional benefits have recently garnered increased interest for human consumption. Here we present chromosome scale sequence assemblies and annotations for the A- and C-genome diploids, (*Avena atlantica* Baum and Fedak and *Avena eriantha* Durieu, respectively). The *A. atlantica* and *A. eriantha* assemblies are composed of a total of 2,195 and 2,652 scaffolds, spanning 3.69 and 3.78 Gb with an N50 of 513 and 588 Mb, respectively, and an L50 of 4 for both assemblies. Analysis of the assemblies classified much of the genomes as repetitive sequence (~83%) with LTR retrotransposons (Copia- and Gypsy-like elements) making up the majority of the classified elements. Annotation of the genome, using sequenced transcriptome identified ~50,000 gene models in both species. Shared ancestry between barley (*H. vulgare* L.) and the two diploid *Avena* genomes can be seen in the significant level of synteny observed between the seven *Avena* and *Hordeum* chromosomes.

W7**50: Oats, Wild and Cultivated**
The Repetitive DNA Landscape in Avena (Poaceae): Chromosome and Genome Evolution defined by Major Repeat Classes in Whole-Genome Sequence Reads
Qing Liu[1], Xiaoyu Li[2], Xiangying Zhou[2], Mingzhi Li[3], Fengjiao Zhang[4], Trude Schwarzacher[5] and Pat Heslop-Harrison[6], (1)South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, Guangdong, China, (2)China Botanical Garden, Guangzhou, China, (3)Genepioneer Biotechnologies Co. Ltd., Nanjing, China, (4)Chinese Academy of Sciences, Nanjing, China, (5)University of Leicester, Leicester, United Kingdom, (6)Department of Genetics, University of Leicester, Leicester, United Kingdom
Common oat, *Avena sativa* (2*n*=6*x*=42), includes A, C and D genomes from diploid *Avena* (2*n*=2*x*=14) progenitors. Here, we aimed to define the nature, abundance and organization of all the repetitive DNA sequences in a range of diploid and the hexaploid species. We put the results into the context of species or genome evolution and their chromosomal distribution. Using raw reads from genomic survey sequencing, we show that some 200 repeated DNA motifs make up 70% of the *Avena* genome, with less than 20 families making up 20% of the total. Retroelements represent the major component, with Ty3/Gypsy elements representing more than 40% of all the DNA, nearly three times more abundant than Ty1/Copia elements. DNA transposons are about 5% of the total, while tandemly repeated, satellite DNA sequences fit into 55 families and represent about 2% of the genome. The *Avena* species are monophyletic, but both bioinformatic comparisons of repeats in the different genomes, and *in situ* hybridization to metaphase chromosomes from the hexaploid species, shows that some repeat families are specific to individual genomes, or the A and D genomes together. Notably, there are terminal regions of many chromosomes showing different repeat families from the rest of the chromosome, suggesting presence of translocations between the genomes. The relatively small number of repeat families shows there are evolutionary constraints on their nature and amplification, with mechanism leading to homogenization, while repeat characterization is useful in providing genome markers and to assist with future assemblies of this large genome (c. 4,100 Mb in the diploid). The frequency of inter-genomic translocations suggests optimum strategies to exploit genetic variation from diploid oats for improvement of the hexaploid may differ from those used widely in bread wheat.

W7**51: Oats, Wild and Cultivated**
Transcriptomics and Metabolomics to Identify Drivers of Seed Composition in Oat
Haixiao Hu[1], Trevor H. Yeats[1], Maryn Carlson[1], Mark E. Sorrells[1], Melanie Caffe-Treml[2], Kevin P. Smith[3], Lucia Gutierrez[4], Michael A. Gore[5] and **Jean-Luc Jannink**[6], (1)Cornell University, Ithaca, NY, (2)South Dakota State University, Brookings, SD, (3)University of Minnesota, St. Paul, MN, (4)University of Wisconsin - Madison, Madison, WI, (5)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (6)USDA-ARS / Cornell University, Ithaca, NY
In plant breeding it is often possible to make large full-sib families and therefore to impose strong within-family selection. Prediction of performance within families, enabling selection prior to extensive phenotyping, is challenging because linkage disequilibrium relationships among loci within a family may be different from those within the general population that is the source of the training population. The general premise of the research is that improved identification of causal loci within a diversity panel can improve prediction within an elite panel. To test this hypothesis, we are collecting metabolomic and transcriptomic data on a diversity panel. We will analyze these data to identify drivers of oat seed metabolite composition. We will use transcriptomic data to improve causal locus identification. We present data from a pilot study used to optimize methods to collect transcriptomic data, showing how we plan to use that data for locus prioritization as applied to oat oil content.

W7**52: Oats, Wild and Cultivated**

Efficient NGS Strategies and Data Integration for Crop Genomics

Brian Boyle[1], Davoud Torkamaneh[1,2], Yann Dorant[1], Wubishet Bekele[3], Asuka Itaya[3], Nicholas A. Tinker[3], Louis Bernatchez[1] and Francois Belzile[4], (1)Université Laval, Québec, QC, Canada, (2)University of Guelph, Guelph, ON, Canada, (3)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (4)Laval University, Quebec, QC, Canada

Next-generation sequencing (NGS) has revolutionized plant and animal research by providing powerful genotyping methods which offer a wide range of applications from genome-wide analysis to routine screening with a high level of accuracy and reproducibility. Furthermore, they provide a straightforward workflow to identify, validate, and screen genetic variants in a short time at low cost. NGS-based genotyping methods include whole-genome re-sequencing, SNP arrays, and reduced representation sequencing, which are widely applied in crops. The main challenges facing breeders and geneticists today is how to choose an appropriate genotyping method and how to integrate genotyping data sets obtained from various sources. In this short talk, we discuss the advantages and challenges of several NGS methods for genome-wide genetic marker development and genotyping in crop plants. We also discuss how imputation methods can be used to both fill in missing data in genotypic data sets and to integrate data sets obtained using different genotyping tools.

W7**53: Oats, Wild and Cultivated**

Genomic Selection Strategy in Oat

Ian G. McNish[1], Jeffrey L. Neyhart[2] and Kevin P. Smith[2], (1)University of Minnesota, St Paul, MN, (2)University of Minnesota, St. Paul, MN

Genomic selection is a routine process in the breeding pipelines of maize and soybean programs, but the application of genomic selection to oat breeding is relatively new. Genomic selection is a process where the performance of a set of lines is predicted using genotypic information and the known performance of a related set of lines. This process is efficient because the cost to genotype a line is less than the cost to phenotype a line. Genomic selection can be used in a breeding program to increase the rate of genetic gain, decrease costs, and increase morale. In this presentation, I will explain the principles of genomic selection, present some results from the oat breeding program at the University of Minnesota, and present some simulation results that show how the oat breeding community can can collaborate to improve genomic selection.

W7**54: Organellar Genetics**

Interpreting Patterns of Sequence Conservation in Plastid Genomes to Identify Novel Functional Elements

Ian Small, The University of Western Australia, Crawley, WA, Australia

W7**55: Organellar Genetics**

Genetic and Molecular Basis of Chromoplast Division

Tianhu Sun, Cornell University, Ithaca, NY and **Li Li**, USDA-ARS/Cornell University, Ithaca, NY

Chromoplast is the colored plastid that synthesizes and stores large amount of carotenoids. Chromoplast number and size define its sink strength for carotenoid accumulation in plants. However, nothing is known about the mechanisms underlying chromoplast duplication. Plant mutants or natural variants defect in chromoplast replication provide excellent genetic materials to reveal the basis of chromoplast division. A number of mutants and variants with carotenoid accumulation in specific organs are conferred by an *Orange* (*Or*) gene mutation. *Or* initiates chromoplast biogenesis for carotenoid accumulation, but restrains chromoplast division with only one or two large chromoplasts per affected cell. Our genetic and molecular studies suggest that chromoplast division employs the same chloroplast binary fission machinery. *Or* exerts no influence on the expression of plastid division genes. OR affects chromoplast duplication by its ability to specifically interact with the Membrane Occupation and Recognition Nexus domain of plastid division factor ARC3. Such a specific interaction hinders the association of ARC3 with PARC6, resulting in restricted chromoplast division. Our findings demonstrate OR as a chromoplast division regulator and provide novel mechanistic insights into chromoplast replication in plants.

W7**56: Organellar Genetics**

Development of a Novel Mini-Synplastome System to Genetically Engineer Chloroplasts in Plants

Alessandro Occhialini, University of Tennessee, Knoxville, TN and Alexander C. Pfotenhauer; Agnieszka Piatek; C. Neal Stewart, Jr; and Scott C. Lenaghan

Over the last 3 decades the traditional approach of chloroplast genome (plastome) engineering in higher plants has been based on homologous recombination of transgenes organized in simple synthetic operons. This classical method is extremely laborious requiring multiple cycles of tissue culture to reach homoplasmy, and in case of installing complex metabolic pathways, multiple transformation/selection/regeneration steps are necessary. Consequently, the efficient installation of multigene pathways into higher plant chloroplasts will require novel engineering approaches.

Here we describe a new synthetic platform for advanced engineering of chloroplast genome in higher plants. We demonstrated the successful installation of extra-plastomic DNA “Mini-Synplastome” able to autonomously replicate in chloroplasts of *Solanum tuberosum* (potato) transplastomic plants. In order to demonstrate engineering with extra-plastomic DNA, vectors equipped with sequences homologous to the inverted repeat region (IR) and containing a chloroplast origin of replication (*ori-A*) were developed. To reduce vector/plastome recombination and increase the probability of episomal propagation, these vectors were designed using heterologous sequences from a different plant species (*Nicotiana tabacum*). The transformation of leaf tissue with these IR constructs generated transplastomic plants with an episomally replicating plasmids that persisted over multiple cycles of tissue culture and regeneration of the second generation of plants from transplastomic leaf tissue (~15 months).

This new engineering approach would represent an alternative and valuable method, not only for transgene integration into higher plant plastids, but would also provide a substantial advancement in terms of speed and flexibility, facilitating the installation of complex multigene pathways organized in several synthetic constructs. Moreover, this “Mini-Synplastome” platform can be used like a classical cloning vector to integrate transgenes in synthetic operons by traditional cloning, and their ability to episomally replicate in chloroplasts allows similar flexibility of using bacterial cells.

W757: **Organellar Genetics**
Engineered RNA-Binding Protein for Transgene Activation in Non-Green Plastids
Qiguo Yu, Rutgers University, Piscataway, NJ

W758: **Organellar Genetics**
Rare Maternal and Bi-Parental Transmission of the Cucumber Mitochondrial DNA Reveals Sorting of Polymorphisms across Generations
Michael J. Havey, USDA-ARS and University of Wisconsin, Madison, WI
The mitochondrial (mt) DNA of cucumber is paternally transmitted, and it is unclear if relatively rare maternal or biparental transmission occurs. We used a mt mutant (MSC16) of cucumber to screen for maternally or biparentally transmitted mt DNA. Wild-type progenies from crosses of MSC16 as the male were selected and genotyped for indel polymorphisms in the mt DNA, and rare maternal and biparental transmission of mt polymorphisms were detected. We then used normal and droplet-digital (dd) PCR to study transmission of polymorphic mt markers across generations. Different combinations of maternal and paternal mt polymorphisms were revealed in progenies across generations, indicating that maternal regions were transmitted to progenies and can predominate. These results indicate that rare maternal or biparental transmission of mt DNA occurs in cucumber, and sorting of these polymorphisms contributes to mt DNA diversity over generations.

W759: **Ornamentals**
High-Quality Chromosome-Level Rose Genome Sequence provides insights on Domestication and Major Traits
Jérémy Just, École Normale Supérieure de Lyon / CNRS, Lyon, France and On the behalf of the The Rose Genome Sequencing Initiative(1)
Roses are among the most commonly cultivated ornamental plants worldwide since antiquity, and have high cultural and economic importance for the ornamental plant market and in the perfume industry. The genus *Rosa* contains approximately 200 species, more than half of which are polyploid. Owing to natural autoincompatibility and artificial interspecific hybridisations, all roses have highly heterozygous genomes that are challenging to assemble despite their relatively small size (560 Mb). During the past years, we generated a number of biotechnology and molecular tools that allowed discovering the molecular mechanisms controlling flower formation and scent biosynthesis[2-6]. However, to date, attempts to assemble rose genomes with short reads have led to highly fragmented assemblies composed of thousands of scaffolds. To overcome these bottlenecks and obtain a reference genome, we produced a homozygous genome from *Rosa chinensis,* known to have extensively participated in breeding and the creation of modern roses, through an original *in vitro* culture protocol[7], and we sequenced this homozygous genome with long-read sequencing technology.
We used PacBio Single Molecule Real-Time sequencing at a depth of coverage of 80× and an original meta-assembly approach to obtain a very high-quality genome assembly[1]. The final assembly was composed of 82 contigs for an N50 of 24 Mb. Using a genetic map and Hi-C chromosomal-contact-map data, we successfully built the seven pseudomolecules of *R. chinensis*, containing 97.7% of the assembly. Resequencing of the genome of 14 major genotypes that contributed to rose domestication, along with genome diversity analyses, highlighted the mosaic origin of the genome of modern rose hybrids that combines European species traits and Chinese species traits[1]. Expert gene annotations along with gene expression data permitted the reconstruction of gene regulatory pathways associated with major rose traits, and allowed to describe epigenetic variation landscapes along the rose genome[1]. These analyses permitted in particular to identify the molecular mechanism by which double flower are formed[8], which had been a research topic for decades. Comparative genomics investigation gave insight into rose recent history, and paleohistory within the Rosaceae family. Together, these resources provide a solid foundation for understanding the mechanisms governing rose traits and their diversity and will accelerate knowledge discovery in roses, *Rosaceae* and ornamentals.
Raymond et al. (2018). The *Rosa* genome provides new insights into the domestication of modern roses.
Nature Genetics **50**:772–777. DOI: 10.1038/s41588-018-0110-3.
Bendahmane et al. (2013). *J Exp Bot* **64**(4):847-857.
Dubois et al. (2012). *BMC Genomics* **13**:638.
Dubois et al. (2011). *PLoS ONE* **6**:e28455.
Dubois et al. (2010). *PLoS ONE* **5**:e9288.
Magnard et al. (2015). *Science* **349**:81-83.
Vergne et al. (2018). *Protocol Exchange*. DOI: 10.1038/protex.2018.067.
François et al. (2018). *Scientific Reports* **8**:12912. DOI: 10.1038/s41598-018-30918-4.

W760: **Ornamentals**
Comparative Genomics Identifies Patterns of Selection in Roses (Rosaceae)
Jinyong Hu, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

W761: **Ornamentals**
Genetic Variation in Historic Populations of *Chrysanthemum Arcticum* Using GBS (DArTseqLD)
Neil O Anderson, University of Minnesota, Department of Horticultural Science, Saint Paul, MN
Chrysanthemum arcticum L. (=*Leucanthemum arcticum* L.) and its two subspecies, *C.a.* subsp. *arcticum* L. and *C.a.* subsp. *polare* Hultén, is the sole chrysanthemum species native to the New World, with its centers' of origin and diversity in the State of Alaska. The species is of interest since it is salt tolerant with a groundcover habit and may possess other ornamental traits of value to chrysanthemum breeders. It is of interest to determine genetic structure within and among populations for targeting future collections of extant genotypes. Recent explorations have shown the species is becoming endangered or threatened which limits obtaining extant specimens for genetic and breeding analyses. Determination of the populations with the highest levels of genetic diversity would enable more targeted collection of extant samples. We collected 362 historic specimens (dating as far back as 1865) using destructive leaf sampling at 11 herbaria across N. America for DNA extraction; three extant

specimens were obtained from USDA GRIN for fresh leaf comparisons. DNA quality/quantity and bioanalyzer assays were performed to determine the levels (fragment sizes) of DNA degradation. A pilot study using random sampling of 96 genotypes across its distributional range were analyzed using GBS (DArTSeqLD). Considerable degradation occurred with historic samples, most likely due to severe conditions post-sampling in the wild before they were submitted to herbaria for preservation. This minimized the amount of usable GBS data. Methods of circumventing issues connected with using DNA from historic samples of Arctic daisy will be presented.

W762: Ornamentals
Linkage and QTL Mapping in Autohexaploid Chrysanthemum

Geert van Geest[1,2], Peter M. Bourke[1], Aike Post[2] and Paul Arens[1], (1)Wageningen University & Research, Plant Breeding, Wageningen, Netherlands, (2)Deliflor Chrysanten B.V., Maasdijk, Netherlands

Genetic linkage maps are indispensable for DNA-informed breeding. However, for hexaploids like chrysanthemum (2n = 6x = 54) linkage map construction is complex, requiring specialised methods and software. We genotyped a large bi-parental population of 406 individuals with a set of 183,000 SNP markers on an Axiom array. This data showed that chrysanthemum has polysomic inheritance, which means that pairing of homologous chromosomes occurs at random at meiosis. For linkage analysis we used the R package named polymapR, which can be used for all steps required for linkage mapping in polyploids, like marker filtering, clustering, ordering and map integration. By applying the software, we integrated 54 homologous chromosomes of the maternal parent and 53 homologous chromosomes of the paternal parent into a linkage map with the expected 9 chromosomes. Because the map was phased and integrated, we could estimate the inheritance of each homologous chromosome in each member of the progeny. Based on this information we applied an interval QTL mapping approach in which the combinations of each of the six possible alleles at each locus are taken into account. This yielded different QTL models for flowering and postharvest traits. The developed methods and results provide a major step forward for DNA-informed breeding in chrysanthemum. We are currently investigating the possibilities of sequence-based genotyping methods. A promising method is the sequencing of fragmented DNA selected with RNA baits (bait-seq). Sequence based genotyping enables the reconstruction of local haplotypes that can be used as multi-allelic markers. Such markers can be very helpful to improve linkage map integration and better estimate QTL models.

W763: Ornamentals
Transcriptome Analysis and Identification of Single Nucleotide Polymorphisms for Powdery Mildew Resistance in Gerbera Daisy

Krishna Bhattarai and Zhanao Deng, University of Florida, Wimauma, FL

Powdery mildew (PM), caused by *Podosphaera xanthii*, is one of the most challenging diseases to gerbera daisy production. It is difficult to control with chemical and biological control methods. Development of PM resistance in gerbera can provide an economic, effective and environmentally-safe control of the disease. Although the loci controlling the PM resistance in resistant gerbera are known, the genes underlying are yet to be identified and tagged with molecular markers. The leaf transcriptomes of resistant and susceptible gerbera lines were sequenced using the Illumina HiSeq2000 system. A transcriptome assembly was developed using Trinity, TransAbyss, Velvet and SoapDenovo, which consists of 145,348 contigs. The sequence reads from the resistant and susceptible gerbera lines were mapped to the assembly using Tophat2 and quantified using RSEM. Blast2Go was used to functionally annotate the transcriptome and identify differentially expressed genes. Analysis of the quantified reads revealed that 5,603 transcripts were differentially expressed: 2,057 up-regulated and 3,546 down-regulated. Based on keyword search in the functionally annotated transcriptome, 265 genes were involved in “disease resistance”, among which 16 and 20 were up- and down-regulated, respectively. Sequence variant calling using GATK, SAMtools and Freebayes revealed 580,141, 399,594 and 1,331,235 SNPs in the resistant and susceptible gerbera lines, respectively. Based on GATK variant calling, 1,222 SNPs were present in the up- or down-regulated “disease resistance” genes. These SNPs are a valuable resource for developing molecular markers that can be used in gerbera breeding for PM resistance.

W764: Ornamentals
Molecular Regulation of Petal Size and Nectar Production in Field Pennycress (*Thlaspi arvense*)

Jason Thomas, University of Minnesota Twin Cities, St. Paul, MN

W765: Perennial Grasses
Genomics Approaches to Gene Regulatory Networks for Lignin Biosynthesis in Grass

Xiaolan Rao, University of North Texas, Denton, TX

W766: Perennial Grasses
Analysis of Leaf Characteristics in the Halophyte Seashore Paspalum

Katrien M. Devos[1], Joanne Lee[2], Hua Chen[2], Melanie Rug[2], Paul Raymer[3] and John J. Spiekerman[1], (1)University of Georgia, Athens, GA, (2)The Australian National University, Canberra, Australia, (3)University of Georgia, Griffin, GA

Soil salinity caused by irrigation and urbanization is a growing problem worldwide and results in significant crop yield losses. Breeding for salt tolerant crops has met with only limited success, likely because the trait is genetically complex. Halophytism, the ability to complete a life cycle under saline conditions, occurs in less than 1% of angiosperms, but has originated multiple times independently. Some taxonomic groups may therefore contain many of the pathway elements needed to confer salt tolerance. We have focused our attention on seashore paspalum, *Paspalum vaginatum* Sw., a halophytic turfgrass belonging to the grass subfamily Panicoideae. In addition to conducting population structure analyses and genetic mapping, we are also investigating whether the leaf surface topology plays a role in the high salt tolerance levels of seashore paspalum. Scanning electron microscopy (SEM) of the leaf epidermis revealed the presence of ridges of large papillae that overarch stomata. We demonstrated with CoroNa Green staining and cryo-SEM combined with energy-dispersive X-ray spectroscopy (EDS) that the epidermal cells and papillae contain sodium (Na) and likely play a role in Na sequestration. Adaxial leaf surfaces of seashore paspalum are also highly

hydrophobic, and we hypothesize that the hydrophobic barrier reduces both salt uptake and water loss. Our results suggest that leaf topology in seashore paspalum may be an important contributor to salinity tolerance.

W767: Perennial Grasses
Comparative Sequence and Synteny Analysis across 13 Complete *de novo* Grass Genomes
John T. Lovell and Jeremy Schmutz, HudsonAlpha Institute for Biotechnology, Huntsville, AL
The grass family is characterized by high levels of genetic diversity but relatively conserved chromosome-scale synteny. Here we model the chromosomal organization of syntenic sequence blocks across 13 de novo grass genomes and track the evolution of a duplication that predated most grass diversification. We then explore the processes of coding sequence loss and gain within a set of gene families and discuss the distribution of presence-absence variation across the phylogenetic history of grasses.

W768: Perennial Grasses
Understanding the Epigenomic Landscape and Transcriptional Architecture of Switchgrass Imposed with Single and Combinations of Drought and Heat Stress
Venu (Kal) Kalavacharla[1], Rita Hayford[2], Vasudevan Ayyappan[1], Shaojun Xie[3], Desalegn D. Serba[4], Malay C. Saha[5], Jyothi Thimmapuram[3] and Cathy Wu[2], (1)Delaware State University, Dover, DE, (2)University of Delaware, Newark, DE, (3)Purdue University, West Lafayette, IN, (4)Kansas State University, Hays, KS, (5)Noble Research Institute, Ardmore, OK
Plants in their natural environment are affected by a combination of simultaneous stresses. High temperatures and water deficiency severely affect crop growth and yield. Despite the documented damage of heat and drought on crops, their interaction and the crop's physiological response to these combined stresses remain unclear. Hence, there is a need to understand the molecular and physiological mechanisms of crop tolerance under heat and drought stresses. Recent developments in high-throughput sequencing technologies have improved our understanding of the molecular mechanisms in heat and drought tolerance. Switchgrass is an economically important bioenergy crop, grown in a wide geographical region. This is due to its relatively high biomass accumulation even when grown in marginal lands, its ability to assimilate nitrogen, and its adaptability to climatic variability. It is necessary to develop tools to facilitate the identification, characterization, and regulation of important genes that contribute to heat and drought adaptation. Analyses of the epigenetic landscape and combining this with RNA profiling in response to heat and drought have not been carried out in switchgrass. In this study, we profiled the genome-wide methylated-histone mark, H3K4me3, using ChIP-seq analysis coupled with RNA-seq to determine differential histone modifications and expression of genes related to heat, drought, and the combination of both stresses. Our study provides the first attempt at integrated transcriptome and epigenomic profiling of switchgrass under drought and combination of drought and heat stress. Methylated peaks (H3K4me3) were enriched downstream of the transcription start sites of protein-coding genes in both drought and heat treated samples.. The transcriptome analysis revealed unique expression patterns of genes and transcription factors associated with photosynthetic machinery, carbon fixation and phenylpropanoid pathways. Additionally, we have developed reference methylomes, transcriptomes, and epigenomes in upland and lowland switchgrass ecotypes. Understanding the chromatin status and transcriptional responses brought about by these combined stresses are necessary to unravel unique genes and signaling pathways to improve abiotic stress tolerance and biomass production in switchgrass. The present study would also be useful in understanding the cross talk between the multiple stresses, which will help to unravel multifunctional genes to combat abiotic stresses in context of climate change.

W769: Perennial Grasses
Optimizing Selective Breeding in Switchgrass (*Panicum virgatum*) combining Genetics and Genomics
Serge Edme, USDA-ARS Wheat, Sorghum, and Forage Research Unit, Lincoln, NE

W770: Perennial Grasses
Transgenic Pollen Containment in *Brachypodium sylvaticum* and *Panicum virgatum*
Jonathan Willis, USDA-ARS, Albany, CA

W771: Plant and Animal Paleogenomics
From Dinosaurs to Birds: Genomic Evolution from a Chromosomal Perspective
Rebecca O'Connor, University of Kent, Canterbury, United Kingdom
Rebecca E. O'Connor[1], Michael N. Romanov[1], Lucas G. Kiazim[1], Paul M. Barrett[2], Marta Farré[3], Joana Damas[3], Malcolm Ferguson-Smith[4], Nicole Valenzuela[5], Denis M. Larkin[3*], Darren K. Griffin[1*].
[1] School of Biosciences, University of Kent, Canterbury, UK
[2] Department of Earth Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK
[3] Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, UK
[4] Department of Veterinary Medicine, Cambridge University, Cambridge, UK
[5] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, USA
Dinosaurs have been ever present in popular culture and the creative arts since the earliest fossil discoveries. While organismal studies focus primarily on their morphology, relationships, likely behaviour and ecology, there have been few academic studies that have made extensive extrapolations about the nature of non-avian dinosaur genome structure. Here, we apply bioinformatic and molecular cytogenetic approaches to determine the genomic structure of the diapsid common ancestor. We then infer the events that likely occurred along this lineage to theropod dinosaurs and modern birds. Our results suggest that most elements of a typical 'avian-like' karyotype (40 chromosome pairs, including 30 microchromosomes) were in place before the divergence of turtles from birds ~255 mya. This genome organisation therefore predates the emergence of early dinosaurs and pterosaurs and the evolution of flight. Remaining largely unchanged interchromosomally through the dinosaur–theropod route, intrachromosomal changes nonetheless reveal evolutionary breakpoint regions enriched for genes with ontology terms related to chromatin organisation and transcription. This genomic structure therefore appears highly stable yet contributes to a large degree of phenotypic

diversity, as well as underpinning adaptive responses to major environmental disruptions via intrachromosomal repatterning. We propose therefore that the overall genome organization and evolution of dinosaur chromosomes (inclusive of the avian radiation) had deeper origins than previously appreciated and was a major contributing factor to the morphology, physiology, ecology, evolutionary change, and ultimately survival, of this fascinating group of animals.

W772: Plant and Animal Paleogenomics
Evolutionary Dynamics of Grass Genomes
Arnaud Bellec, French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France
Modern grass genomes exhibit a large range of genome structures. They are the result of both whole genome duplication (i.e. polyploidization) and chromosomal rearrangement events, such as chromosomes fusion, inversion and translocation. These alterations of ancestral genome structures, by modifying gene dosage and regulation process, can affect genes at both expression and sequence level leading to deletion, neo-functionalization and sub-functionalization. In this study, we have focused on chromosomal inversions of modern grass species, as a major driver of synteny decay. The inversions have been accurately detected by comparing gene order of modern species with that of the reconstructed Ancestor Grass Karyotype of 12 ancestral chromosomes. Seven grass species have been investigated representing the *Pooideae*, *Panicoideae* and *Ehrhartoideae* subfamilies that have diverged 50 million years ago. The chromosomal inversions have been addressed in the light of grasses evolutionary history.

W773: Plant and Animal Paleogenomics
Reconstructing the Large-Scale Evolution of Genomes and Gene Functions
Paul D Thomas, University of Southern California, Los Angeles, CA
Over time, the gene content of genomes evolves by processes of gene duplication, horizontal transfer, de novo gene origination, and gene loss. My group has reconstructed the evolutionary history of over 1 million genes in 15,000 gene families, covering all domains of life. From these gene family trees, we have inferred the gene content of common ancestral genomes, and the history of duplication, transfer, origination and loss along each branch of the species tree. With the Gene Ontology Consortium, we have also constructed models of function evolution through thousands of these families. I will give an overview of our reconstruction methods, and findings from this reconstruction, such as the deep history of the human genome.

W774: Plant and Animal Paleogenomics
Re-Analysizing Plant Genome Sequences Unfolds their Recursively Folded Structure
Xiyin Wang, NCST and UGA, Athens, GA
Polyploidization has recursively affected the evolution of plant genomes, and after each event, whole genome or all chromosomes duplicated or triplicated, and then thousands of gene lost, genome repatterned, and chromosomes repacked often returning to small numbers. These repetitive operations make the plant genomes rather complex. With collaborating efforts by colleagues, we developed methods to infer collinear genes between different genomes or within each of them, and by considering the divergence between collinear genes, we tried to link them to different evolutionary events, esp. each of ancient polyploidization events. This helps us to find the evolutionary history of genomes and chromosomes, and functional innovations of genes and families. Here, we show our findings in selected land plants.

W775: Plant and Animal Paleogenomics
High-Resolution Comparative Analysis of Great Ape Genomes
Zev Kronenberg, Phase Genomics Inc, Washington, DC
Genetic studies of human evolution require high-quality contiguous ape genome assemblies that are not guided by the human reference. Here, we couple long-read sequence assembly, full-length cDNA sequencing, and a multi-platform scaffolding approach to produce *ab initio*chimpanzee and orangutan genome assemblies where most genes are complete, gaps are closed, and novel gene models are identified. Using two long-read *de novo* human genome assemblies and a previous gorilla genome assembly, we characterize lineage-specific and shared great ape genetic variation ranging from single base-pair to megabase-sized variants. We identify 17,789 fixed structural variants more than doubling the number of genic and putative regulatory changes that emerged in humans since divergence from nonhuman apes. Interestingly, fixed human-specific deletions are enriched near genes that are downregulated in human compared to chimpanzee cerebral organoids, particularly in cells analogous to radial glial neural progenitors.

W776: Plant and Animal Paleogenomics
Hybrid Conflict, Biased Gene Losses and Developmental Innovation: The Continuing Impact of Ancient Polyploidies on Genome Structure and Function
Gavin C. Conant[1,2], Marianne L. Emery[3], M. Madeline Willis[3], Yue Hao[4], Kerrie W. Barry[5], Khouanchy Oakgrove[6], Yi Peng[6], Jeremy Schmutz[7], Eric Lyons[8], J. Chris Pires[9] and Patrick P. Edger[10], (1)Department of Biological Sciences/North Carolina State University, Raleigh, NC, (2)Division of Animal Sciences, University of Missouri, Columbia, MO, (3)University of Missouri, Columbia, MO, (4)North Carolina State University, Raleigh, NC, (5)DOE Joint Genome Institute, Walnut Creek, CA, (6)Department of Energy Joint Genome Institute, Walnut Creek, CA, (7)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (8)The University of Arizona, Tucson, AZ, (9)Division of Biological Sciences, University of Missouri, Columbia, MO, (10)Department of Horticulture, Michigan State University, East Lansing, MI
Once source of the innovative potential of polyploidy may be in the polyploid organisms' origins in the merging and mixing of genomes from two different species (e.g., allopolyploidy). Using POInT (the PolyploidOrthology Inference Tool), we model the resolution of four allopolyploidy events that occurred in the ancestors of the grasses, the thale cress *Arabidopsis thaliana*, the teleost fishes and bakers' yeast. This model detects robust evidence for the existence of *biased fractionation*in all lineages, whereby genes from one of the two parental subgenomes were more likely to be lost than those from the other subgenome. The pattern of biased fractionation after the *Arabidopsis*and grass allopolyploid

events was surprisingly constant in time. In strong contrast, the yeast allopolyploid event shows evidence of biased fractionation only immediately after the event, with balanced gene losses more recently. In the yeasts, *Arabidopsis*and its relatives and in the fishes, similar groups of genes were also rapidly returned to single-copy after the polyploidy, suggesting that selection may favor the removal of specific duplicates in the initial phases of WGD. In the case of the fishes, the surviving duplicates from the polyploidy are also less likely to function in early embryo development than are singletons. Instead, the pattern of which tissues these duplicates are expressed in and their functions lend support to recent suggestions the fish genome duplication was the source of a morphological innovation in the structure of the teleost retina.In sum, we suggest that, after allopolyploidy, there are functional conflicts between interacting genes encoded in different subgenomes that are ultimately resolved through preferential duplicate loss, with surviving duplicates then becoming available for co-option into novel functions.

W777: Plant Chromosome Biology
TBA

Wojtek Pawlowski, Cornell University - Section of Plant Biology, School of Integrative Plant Science, Ithaca, NY

W778: Plant Chromosome Biology
TBA

Isabelle M. Henry, University of California, Davis, Davis, CA

W779: Plant Chromosome Biology
Ribosomal RNA Gene Loci in Hexaploid Wheat Resolved By Chromosome Sorting and Optical Mapping

Hana Simkova[1], Zuzana Tulpova[1], Helena Toegelova[1], Veronika Kapustova[1], Jan Vrána[1], Tingting Zhu[2], Ming-Cheng Luo[2], Martin Mascher[3], István Molnár[4], Eva Hřibová[4] and Jaroslav Doležel[1], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Department of Plant Sciences, University of California, Davis, Davis, CA, (3)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany, (4)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic

Ribosomal RNA gene loci pose an indispensable part of both prokaryotic and eukaryotic genomes. They are mostly organized as long head-to-tail tandem arrays spanning several hundred kilobases to multiple megabases, which precludes their complete assembling from NGS data and impedes characterization of particular loci.

In our study, we identified and analysed 26S-5.8S-18S rRNA multigene loci in bread wheat genome by Bionano genome (BNG) mapping, a technology that visualizes short sequence motives along DNA molecules of 150 kb to 1 Mb. The rDNA arrays can be recognised in wheat BNG maps as a regular label pattern with ~9-kb unit. The maps have been produced from separated rDNA-bearing chromosome arms discriminated by flow cytometry. The BNG map data enabled precise positioning of the rDNA arrays in the sequence assembly and also quantification of rDNA units in particular chromosome arms. Locus-specific sequences of rDNA units were reconstructed from chromosome-arm-specific raw data using RepeatExplorer pipeline. This information combined with transcriptomics data (RNA-seq, Iso-Seq) enabled a tissue-specific analysis of expression pattern for particular major and minor rRNA loci.

W780: Plant Chromosome Biology
Pervasive Numerical and Structural Chromosomal Variation in a Newly Formed Segmental Allotetraploid Rice

Ying Wu, Northeast Normal University, Changchun, China

Polyploidy or whole genome duplication (WGD) has occurred ubiquitously across the tree of angiosperms and is known as a pervasive driving force in both plant and animal evolution. Two major types of polyploidy are recognized, that is, autopolyploidy (doubling of a single species genome) and allopolyploidy (doubling of a hybrid genome formed by two or more species). Between these two types of polyploidy is a continuum which can be collectively termed segmental allopolyploidy. A fascinating finding with respect to genome evolution is that numeric chromosome variation, i.e., aneuploidy, appears to generally associate with nascent polyploidization. In the case of allopolyploidy and segmental allopolyploidy wherein homoeologous chromosomes exist, structural chromosome variation in the form of homoeologous exchanges (HEs) is recognized. Both numerical and structural chromosomal variations in an allopolyploid or a segmental allopolyploid background may lead to rapid phenotypic diversity and thus enhanced adaptability at the population level. Here, we used whole-genome re-sequencing and molecular cytogenetic tools (FISH) to investigate the occurrence and its trend of aneuploidy and HEs in novel segmental allotetraploid rice system. The system was constructed by colchicine-mediated genome doubling of reciprocal F1 hybrids between the two subspecies (*japonica* and*indica*) of Asian cultivated *Oryza sativa*. The resequencing of 312 randomly selected tetraploid individuals revealed that HE event is rampant in the rice segmental allotetraploids. HE events converted *ca.*30-60% heterozygous genomic regions into homogenized state favoring either of the parental subgenomes. Besides the structural chromosomal variations, we also identified *ca.*40% aneuploids harboring 55 distinct aneuploid karyotypes in the population under normal conditions. Interestingly, the proportion of aneuploids in the survived population shifted to *ca.*65% under acute salt stress, suggesting aneuploidy in general is more tolerant to the stress than euploidy in the tetraploid rice background. We currently assess the effect of HEs on salt and other abiotic stresses, and the possible relationship between HEs and aneuploidy. Regardless, these rampant HEs and aneuploidization events working together have caused dramatic phenotypic diversity at the population level under field conditions.

W781: Plant Chromosome Biology
Back Spliced RNAs from Centromeric Retrotransposon Promote Centromere Chromatin Loops and Affect CENH3 Localization

Yalin Liu, **Handong Su**, Jing Zhang, Yang Liu, Chao Feng and Fangpu Han, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

In most eukaryotes, centromeric DNA contains hierarchical arrays of highly repetitive sequences. The processes of higher order chromatin organization on centromeric DNA are confused, and the interactions between centromere chromatin architecture and centromeric specific histone (CENH3) localization are unclear. Centromeric RNA can affect CENH3 deposition, but the interaction between centromere chromatin and RNA

is unknown. The question of why specific repeats are enriched in the centromere region during centromere evolution is not yet answered. Here, we found three types of circular RNAs from maize centromeric retrotransposon CRM1 with the same back-splicing site. These circular RNAs bind in the centromere by R-loops with centromeric CRM1 elements, promoting chromatin loop formation in the centromere regions. Knocking down the CRM1 circular RNAs by RNAi mainly decreased the RNA level of free circular RNAs, and the level of circular RNAs with R-loop formation was increased. Chromatin loops in the CRM1 regions were increased and CENH3 localizations in the centromere were decreased in the RNAi plants. The higher level of R-loops and chromatin loops in the CRM1 regions may be resistant to CENH3 localization. The back-splicing process of retrotransposon is conserved in numerous crops. Our work reveals the centromere chromatin organization by centromeric retrotransposon CRM1 via R-loops and chromatin loops, and improves the understanding of regulation of CENH3 localizations based on the chromatin structure. Centromeric DNA can not only provide sequences for CENH3 loading but also regulate centromere chromatin structure to regulate CENH3 maintenance, which is important in artificial chromosome design. These results unveil potential mechanisms for repeat sequences in centromere activity and centromere evolution.

W782: Plant Chromosome Biology
Small RNAs in Rice Gametes
Jonathan I Gent, University of Georgia, Athens, GA

Egg and sperm egg nuclei are dramatically different from each other in terms of chromatin composition, and phenomena such as delayed activation of the paternal genome in the zygote and imprinted gene expression in the endosperm indicate that the differences can have post-fertilization consequences. Small RNA in gametes likely direct chromatin modifications in gametes and post-fertilization as well. We are investigating small RNAs and associated DNA methylation patterns in gametes of rice (Oryza sativa). We have found that 24nt siRNAs are lost from euchromatin-heterochromatin boundaries in both egg and sperm and 24nt siRNAs are gained in normally heterochromatic regions in sperm specifically. Sperm and egg also have distinct miRNA profiles though both share miR159 as the dominant miRNA. Redistributions of 24nt siRNAs in gametes do not correspond to changes in DNA methylation, which is largely similar between both gametes and vegetative cells. In the long term, we aim to connect small RNAs to cell-type-specific chromatin modifications and gene regulation, pre- and post-fertilization.

W783: Plant Cytogenetics
Genetic Control of Homoeologous Recombination in Wheat
Bikram S. Gill, Kansas State University, Manhattan, KS

In wheat, a large number of genes that promote or suppress homologous and homoeologous pairing have been mapped to chromosomes using aneuploids. Among these, the pairing homoeologous gene *Ph1* has been widely studied and deployed to manipulate homoeologous recombination. However, the frequency of homoeologous recombination is very low and recombination events are localized to the distal ends of chromosomes. We have identified a potent gene, *Hpp-5M^g*, where the carrier chromosome 5M^g of *Aegilops geniculata* escapes control of *Ph1*-imposed suppression of homoeologous recombination. Chromosome *5M^g* and its wheat homoeologous chromosome 5D freely recombined in the presence of *Ph1*. The *Hpp-5M^g* gene, in the absence of *Ph1*, led to a vast genome-wide increase in homoeologous recombination in euchromatic as well as heterochromatic and proximal chromosomal regions where even homologous recombination is suppressed. These results, and the insights they provide into the process of meiosis and the use of this system for wheat crop improvement, will be discussed.

W784: Plant Cytogenetics
Interplay between Axial Elements and Meiotic Recombination in Maize
Chung-Ju Rachel Wang, IPMB/Academia Sinica, Taipei, Taiwan

Double strand breaks (DSBs) are generated at the beginning of meiosis by the evolutionarily conserved SPO11 complex that targets preferentially open chromatin. However, how DSBs are formed with respect to chromosome axes in a timely controlled manner remain unclear. Here, we analyzed the maize *spo11-1* mutant and show that it strongly impairs DSB and bivalent formations. Notably, cytological characterization in the *spo11-1* mutant revealed abnormally twisted axial elements that persisted until pachytene. In contrast, examinations of precisely staged wild-type meiocytes uncovered a transient remodeling of axial elements, changing from a curly to linear morphology during leptotene to zygotene transition, which is coincident with DSB formation. Using a SPO11-1 antibody, approximately 300 foci were detected from leptotene to pachytene in wild-type meiocytes. Interestingly, when examining distances between SPO11-1 foci to chromosome axes by super-resolution microscopy, predominant loading of SPO11-1 onto axial elements occurs concordantly with alteration of axial elements during leptotene. Taken together, our results suggest a dynamic localization of SPO11-1 during early meiosis that is correlated with a remodeling of the axial element conformation.

W785: Plant Cytogenetics
REC8, Chromatin and Transcription Orchestrate Arabidopsis Meiotic Recombination
Ian R Henderson, University of Cambridge, Cambridge, United Kingdom

Meiotic chromosomes undergo DNA double-strand breaks (DSBs) that can be repaired using a homolog to produce reciprocal crossovers. REC8-cohesin organizes the recombining chromosome axis and is required to prevent fragmentation. Axis polycomplexes form in Arabidopsis *rec8* mutants, which recruit recombination foci with altered stoichiometry, leading to catastrophic non-homologous recombination. REC8 ChIP shows strong enrichment in centromeric heterochromatin, correlating with suppression of recombination, despite loading of the DSB machinery on the axis in these regions. Loss of the heterochromatic marks H3K9me2 and non-CG DNA methylation in *kyp suvh5 suvh6*causes REC8 remodeling and gain of meiotic recombination in repeated sequences, although centromeric cohesion is maintained. In the chromosome arms, REC8 is enriched within gene bodies, exons and GC-rich sequences, and anti-correlates with transcription, with highest REC8 occupancy in facultatively silent genes. Therefore, as REC8 organizes meiotic chromosome architecture and interhomolog recombination, it is shaped by multiple chromatin states and transcription.

W786: Plant Cytogenetics

Insights into Arabidopsis Gene Regulation and Wheat Nuclear Organisation using Single Molecule RNA FISH

Susan Duncan, Earlham Institute, Norwich, United Kingdom and Anthony Hall

Single molecule RNA FISH (smFISH) enables microscopic analyses of RNA at the cellular level. It uses multiple fluorescent probes that collectively bind and label target transcripts. Since adapting this method for use in plants we have designed probe sets to target exon sequences to generate fluorescent spots that each represent single mRNA. This approach has provided quantitative mRNA-per-cell data as well as sub-cellular localisation information for genes involved in vernalization, root development and boron uptake. By specifically labelling intronic RNA we have also investigated regulation of sense and anti-sense transcription at individual loci. Currently, we are labelling intron RNA from genes that undergo co-transcriptional splicing to investigate hexaploid wheat chromosome positioning. This labelling approach will enable us to generate 3D maps that reveal relative positions of sub-genomes within the nucleus. Together, these results demonstrate that smFISH is a versatile method that can be used to investigate various aspects of plant gene regulation and nuclear organisation.

W7**87: Plant Cytogenetics**

Knocking Down Wheat Meiotic Genes through Virus-Induced Gene-Silencing (VIGS)

Stuart Desjardins[1], Kostya Kanyuka[2] and James D. Higgins[1], (1)University of Leicester, Leicester, United Kingdom, (2)Rothamsted Research, Harpenden, United Kingdom

Virus-induced gene-silencing (VIGS) is a reverse genetic technology that can be used to assess gene function in model and crop species. The technique exploits a natural viral defence response in plants, whereby invading viral RNAs are detected, and this triggers a targeted post-transcriptional gene silencing (PTGS) mechanism. In VIGS, a small fragment of a plant gene of interest is inserted into a viral vector to generate a recombinant virus that, upon infection, leads to silencing of both the foreign viral RNA and the endogenous target gene; creating knockdown lines.

We have applied *barley stripe mosaic virus* (BSMV)-mediated VIGS to the study of recombination in hexaploid wheat, and have optimised the targeting of meiotic genes at peak expression. Following VIGS, successful down-regulation of gene expression can be assessed by qRT-PCR, and meiotic phenotypes analysed cytologically. An initial proof-of-concept study was conducted by silencing the recombinase *TaDMC1*, which resulted in a severe meiotic phenotype with the majority of homologous chromosomes remaining unpaired, visible as univalents, and we are currently expanding this approach to incorporate additional targets. VIGS is proving to be a rapid and cost-effective way of assessing the function of meiotic genes in wheat, enabling us to efficiently screen a broad range of candidate genes, so that further analyses using TILLING or CRISPR/Cas may be employed on promising targets.

W7**89: Plant Disease Resistance**

Towards Stacks of Resistance Genes in Lettuce

Anne Giesbers, UC Davis, Davis, CA

W7**90: Plant Disease Resistance**

A Genetic Toolbox for Managing Blast and Sheath Blight Diseases of Rice in the USA

Yulin Jia, USDA-ARS, Stuttgart, AR, Melissa Jia, DB NRRC, Stuttgart, AR, Wenqi Li, Jiangsu Academy of Agricultural Sciences, Nanjing, 210014, China., Nanjing, China and Don Groth, Rice Research Station, Louisiana State University, Rayne, LA

In the USA, rice blast disease caused by the fungus *Magnaporthe oryzae*, and sheath blight disease caused by the fungus *Rhizoctonia solani* are two major hurdles for stable rice production. *M. oryzae* is a highly adaptable fungus that can easily overcome newly acquired resistance, and *R. solani* is a necrotrophic pathogen where complete genetic resistance has not been identified, and fungicide resistant sheath blight isolates have been reported in commercial rice fields. Presently, rice plants with resistance (*R*) genes (major and minor) are being commonly deployed along with fungicides to reduce damage by these two fungal diseases. A total of 46 rice genetic stocks were developed with effective major and minor blast resistance genes from progenies involved in crosses of rice varieties, Katy, Zhe733, Cybonnet, Jasmine 85, and Saber. Their resistant reactions to differential blast races/isolates were correlated with the presences of DNA markers from resistant donors with artificial inoculation under greenhouse conditions. Resistance to blast of these genetic stocks was evaluated under upland and flooded conditions, without artificial pathogen inoculation, in Louisiana, USA in replicated field plots with controls. All susceptible checks were susceptible. All of them were resistant under upland and flooded field conditions except for one genetic stock which was susceptible under upland conditions but was resistant under flooded conditions. These findings suggest that both major and minor *R* genes identified under controlled greenhouse conditions can be effective in the field. For sheath blight, the major resistant QTL *qShB9-2* contributing 25 % of phenotypic variation was mapped within a 100 kilobase region with a BC_3F_4 mapping population with phenotypic data from pathogen inoculated fields. This result paved the road to develop more accurate genetic markers to monitor the incorporation of qShB9-2 and eventually lead to its cloning and functional validation. Together, we demonstrated that effective resistance genes can be identified under controlled conditions, and their resistant phenotypes are stable under field conditions. A strategy for rice crop protection and elucidating genetic mechanisms of disease resistance with this toolbox will be presented.

W7**91: Plant Disease Resistance**

Genetic Architecture of Fire Blight Susceptibility in Apple (*Malus x domestica*)

Jugpreet Singh[1], Awais Khan[1], Katchen Julliany Pereira Silva[1], Elsa Desnoues[1], Alvin Randall[1], Jie Arro[2], C. Thomas Chao[2] and Gan-Yuan Zhong[2], (1)Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, NY, (2)USDA-ARS, Geneva, NY

Fire blight, caused by gram negative bacteria *Erwinia amylovora*, leads to significant losses in apple production worldwide. Utilization of host genetic resistance can contribute to limit its impact. However, susceptibility to infection depends on genetic background of cultivars, virulence levels of bacterial strains, and environmental conditions. We conducted a multi-population analysis to dissect genomic regions related to fire blight susceptibility in apple (*Malus* x *domestica*). Two bi-parental populations (GMAL-4591 and GMAL-4592) and one multi-parental populations with F_1 individuals randomly selected from 7 segregating populations were evaluated for leaf, shoot and fire blight necrosis traits. Genotyping-by-sequencing (GBS) was used to obtain single nucleotide polymorphism (SNP) markers for genetic mapping and association analysis. Leaf, shoot and necrosis traits showed a range of variation in three segregating populations. Genetic analysis revealed several genomic

regions related to necrosis and percent lesion length in leaf (PLLL) and shoot (PLLS). Very few genomic regions controlling PLLL and PLLS were common between the three mapping populations. PLLL exhibited marker associations on chromosomes 2, 4, and 7 in the GMAL-4591, chromosome 7 and 9 in the GMAL-4592 populations, and across seven different chromosomes in the multi-parental population. These results suggest that genotypic background can significantly influence fire blight susceptibility of different apple cultivars. Some of the identified genomic regions also coincide with previously reported loci for fire blight resistance in domesticated apples. These genomic regions will lead to selection and fine mapping of gene candidates related to fire blight in apple.

W792: **Plant Disease Resistance**

Pathotype-Specificity of Angular Leaf Spot in Common Bean - Genome Wide Association Studies and Implications for Resistance Breeding in Latin America and Africa

Michelle Maria Nay, ETH Zürich, Zürich, Switzerland

W793: **Plant Disease Resistance**

From QTL Mapping to Gene Cloning: Phytophthora Crown and Root Rot Resistance, FaRPc2, in a Complex Allo-Octoploid Strawberry

Seonghee Lee, University of Florida, Wimauma, FL

W794: **Plant Disease Resistance**

Cloning of Wheat Leaf Rust Resistance Gene *Lr42*

Sanzhen Liu, Kansas State University, Manhattan, KS

Aegilops tauschii, the D-genome diploid donor of wheat, is a rich reservoir of resistance genetic elements for biotic and abiotic stresses. The *Lr42* locus from *Ae. tauschii* confers leaf rust resistance at both seedling and adult plant stages and remains effective against all leaf rust races reported to date. *Lr42* has been transferred from *Ae. tauschii* into bread wheat and extensively used in the CIMMYT wheat breeding program. We applied a bulked RNA-Seq based mapping strategy, BSR-Seq, to locate *Lr42* in the reference genome. The markers developed from BSR-Seq were then used for fine mapping, resulting in a few candidate genes. One candidate gene was proved the *Lr42* gene through ectopic overexpression, which encodes an NLR protein. Identification of *Lr42* expands the repertoire of cloned leaf rust resistant genes and our strategy demonstrated the value of BSR-Seq for cloning of disease resistance genes in wheat.

W795: **Plant Disease Resistance**

Transgenics Expressing Citrus-Derived Anti-*CLas* Proteins offer HLB Resistance

Ed Stover, U. S. Horticultural Research Laboratory, USDA-ARS, Fort Pierce, FL and **Goutam Gupta**, NMC Inc., Los Alamos, NM

The decade-long collaboration between USDA-ARS and LANL/NMC started with the concept that citrus immunity can be enhanced to facilitate clearance of *Candidatus* Liberibacter asiaticus (CLas) and prevention of Huanglongbing (HLB). The strategy is to construct transgenic citrus rootstocks and scions expressing anti-bacterial proteins that clear CLas and provide protection against HLB. Three generations of anti-CLas proteins have been designed. Before the availability of the completed citrus genome(s), tobacco Thionin, with anti-bacterial effect on gram-negative bacteria such as CLas, was chosen as the 1st generation anti-CLas protein. This tobacco Thionin showed 70% sequence homology with the published citrus Thionin(s). The tobacco Thionin was modified to increase activity and lower toxicity. Citrus transgenics expressing modified tobacco Thionin showed CLas clearance by *in vitro* and greenhouse *in planta* studies. We, therefore, designed, the 2nd and 3rd generation of anti-CLas proteins to further improve CLas clearance and HLB resistance. The 2nd generation citrus tansgenics expressing a chimera of citrus Thionin (modified to increase activity and lower toxicity) and a citrus LPS-Binding Peptide (LBP) that is added to target the CLas membrane. Preliminary studies showed that the 2nd generation citrus transgenics are more efficient in clearing CLas than the 1st generation citrus transgenics expressing modified tobacco Thionin. Finally, we constructed the 3rd generation citrus transgenics expressing chimeras of modified citrus Thionin and citrus Bacterial Permeability Increasing/Lipid-binding protein (BPI/LBP) or citrus Subtilisin. BPI/LBP was chosen to recognize the CLas membrane and Subtilisin was selected to cleave the CLas outer-membrane protein. Both the 2nd and 3rd generation transgenics are ready for greenhouse efficacy studies. Therefore, we believe that the combined studies on the three generations of transgenic citrus will demonstrate the efficacy of citrus-derived anti-CLas proteins in CLas clearance and HLB resistance.

W796: **Plant Epigenetics and Epigenomics**

Reaction Mechanisms of Pol IV, RDR2 and DCL3 in Heterochromatic siRNA Biosynthesis

Jasleen Singh[1], Vibhor Mishra[1], Feng Wang[1], Hsiao Yun Huang[1] and **Craig Pikaard**[2], (1)Indiana University, Bloomington, IN, (2)Howard Hughes Medical Institute, Indiana University, Bloomington, IN

In plants and mammals,transcriptional silencing of retrotransposons involves DNA methylation guided by small RNAs. Using purified and recombinant enzymes in vitro, have determined the reaction mechanisms that produce the siRNAs that guide *de novo* cytosine methylation in Arabidopsis. The process begins with DNA transcription by NUCLEAR RNA POLYMERASE IV (Pol IV), whose atypical termination mechanism, induced by DNA secondary structure, activates the physically associated enzyme, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). RDR2 converts Pol IV transcripts into double-stranded (ds) RNAs and then adds an untemplated nucleotide to the 3' end of its transcripts, creating 3' overhangs. The Dicer endonuclease, DCL3 cuts the duplexes from either end, generating 24 nt siRNAs and 23 nt siRNAs enriched for the 3' nontemplated nucleotides added by RDR2. Recapitulation of siRNA biogenesis *in vitro*provides explanations for Pol IV-RDR2 co-dependence, the short size of Pol IV transcripts and the origins of untemplated nucleotides in siRNAs *in vivo*.

W797: **Plant Epigenetics and Epigenomics**

The Evolution and Functional Roles of 24-nt Reproductive PhasiRNAs in Plants

Blake Meyers, Donald Danforth Plant Science Center, St. Louis, MO
In plants, 21 or 22-nt miRNAs or siRNAs typically negatively regulate target genes through mRNA cleavage or translational inhibition. Heterochromatic or Pol IV are 24-nt and function to maintain heterochromatin and silence transposons. Phased "secondary" siRNAs (phasiRNAs) are generated from mRNAs targeted by a typically 22-nt "trigger" miRNA, and are produced as either 21- or 24-mers via distinct pathways. Our prior work in maize and rice demonstrated the temporal and spatial distribution of two sets of "reproductive phasiRNAs", which are extraordinarily enriched in the male germline of the grasses. These two sets are the 21-nt (pre-meiotic) and 24-nt (meiotic) siRNAs. Both classes are produced from long, non-coding RNAs, generated by hundreds to thousands of loci, depending on the species. These phased siRNAs show striking similarity to mammalian piRNAs in terms of their abundance, distribution, distinctive staging, and timing of accumulation, but they have independent evolutionary origins. The functions for these small RNAs in plants remain poorly characterized. In monocots, the 24-nt phasiRNA pathway, triggered by miR2275 and abundant during meiosis, requires a recently-diverged Dicer known as DCL5, an interesting evolutionary elaboration of this pathway. I will describe our recent work investigating the evolutionary origins and characterizing the functions and biogenesis of 24-nt plant reproductive phasiRNAs.

W798: Plant Epigenetics and Epigenomics
Systematic Discovery of Gene Regulatory Elements using Chromatin Structure in Plant Genomes
Robert Schmitz[1], William Ricci[2], Zefu Lu[2], Lexiang Ji[1], Alex Marand[2], Maria Katherine Mejia Guerra[3], Christina Ethridge[2], Mary Galli[4], Jaclyn Noshay[5], Nathan M. Springer[5], Andrea Gallavotti[6], Edward S. Buckler[7] and Xiaoyu Zhang[1], (1)University of Georgia, Athens, GA, (2)University Of Georgia, Athens, GA, (3)Cornell University, Ithaca, NY, (4)Rutgers University, Piscataway, NJ, (5)University of Minnesota, St. Paul, MN, (6)Waksman Institute, Rutgers University, Piscataway, NJ, (7)USDA-ARS-Cornell University, Ithaca, NY
Significant progress has been made in recent years in plant genome assembly and gene annotation. However, the systematic identification of plant cis-regulatory DNA elements remains a challenge, as methods that are highly effective in animals do not translate to plants. A comprehensive and well-curated data set of plant cis-regulatory DNA elements is instrumental to understanding transcriptional regulation during development and/or in response to external stimuli. In addition, cis-regulatory DNA elements are also hotspots for genetic variations underlying key agronomical traits. We have discovered a plant-specific chromatin signature that is indicative of cis-regulatory DNA elements. We are using this newly identified signature in combination with high-throughput validation assays to systematically identify, analyze and functionally validate cis-regulatory elements in important crop species.

W799: Plant Epigenetics and Epigenomics
Mechanisms of DNA Methylation in Arabidopsis
Ming Zhou, Ana Marie Palanca and **Julie A. Law**, Salk Institute for Biological Studies, La Jolla, CA
Cytosine DNA methylation plays crucial roles in gene regulation, transposon silencing, and diverse developmental processes. Although the generation of specific DNA methylation patterns is critical for these processes, how methylation is regulated at individual loci remains unclear. In *Arabidopsis*, DNA methylation is established via the RNA-directed DNA methylation (RdDM) pathway, wherein RNA POLYMERASE-IV (Pol-IV), initiates biogenesis of 24-nucleotide small interfering RNAs (24nt-siRNAs) that guide methylation at cognate genomic loci. Using a combined genetic and genomic approach, we show that four Pol-IV-associated factors, CLASSY (CLSY) 1-4, act individually as locus-specific regulators of RdDM and together control the production of essentially all 24nt-siRNAs, demonstrating they are the master regulators of Pol-IV function. Mechanistically, the CLSYs function in connection with H3K9 and CG methylation to facilitate Pol-IV chromatin association and show a striking division of labor, with specific CLSY pairs preferentially regulating loci in the chromosome arms versus pericentromeric heterochromatin. These findings reveal an unanticipated layer of complexity within the RdDM pathway that enables locus-specific control of DNA methylation patterns. Given the conservation between methylation systems in plants and mammals, analogous pathways likely operate in a broad range of organisms.

W800: Plant Interactions with Pests and Pathogens
Gene Editing with CRISPR-Cas to Improve Disease Resistance in Maize and Weed Control in Sorghum
Huirong Gao, Coteva Agriscience, Johnston, IA
Crop productivity dramatically improved over the last century through a combination of technologies, including plant breeding, improved agronomic practices, increased use of fertilizer, use of crop protection chemicals, and transgenic traits. Going forward, genome engineering will be added to this list. The CRISPR/Cas9 system is currently the technology of choice for genome editing in plants, due to its simplicity, efficiency and versatility. Like other crops, maize and sorghum productivity is reduced by disease and insects and weeds. We have used genome editing to improve maize resistance to northern leaf blight disease and to improve sorghum resistance to the parasitic weed Striga.

W801: Plant Interactions with Pests and Pathogens
A Developmental Genomics Analysis of Soybean Defense Processes
Vincent P. Klink, Mississippi State University, Mississippi State, MS
A developmental genomics analysis has been performed regarding gene expression occurring in root cells of *Glycine max* undergoing defense. The analysis has led to the identification of candidate defense genes. Functional experiments have been performed to determine if specific genes truly function in the defense process. Gene expression studies have been performed to ascertain whether cross communication is occurring between the identified defense genes. The work provides a picture regarding the defense process.

W802: Plant Interactions with Pests and Pathogens
Aphid-Host Interactions: It's about Host Resistance, Aphid Virulence and Pest Control
Anna-Maria Botha, Stellenbosch University, Stellenbosch, South Africa

Aphids are a group of approximately 4,700 species of phloem-feeding insects with mainly temperate distributions. Although best known as agricultural pests, they are also valuable systems to study host plant specializations, bacterial symbiosis and environmentally induced morphologies (polyphenism). *Diuraphis noxia* (Kurdjumov, Hemiptera: Aphididae), a specialist phloem feeder, is such economically important cereal aphid pest afflicting wheat and barley yield in dry-land production regions. Unlike in areas where *D. noxia* is endemic and can reproduce through facultative parthenogenesis, in areas were it is invasive, it reproduce only asexually. Despite the lack of sexual recombination, new *D. noxia* biotypes with varying levels of virulence continue to develop and overcome previously resistant host plants, posing multiple threats to global food security. Host plant resistance, the most environmentally sound measure of pest resistance is defined by the response to aphid feeding (i.e., antibiosis, antixenosis or tolerance), with most *Dn* genes following the gene-for-gene relationship. With the availability of the draft genome of *D. noxia* and confounding evidence of genomic plasticity, we set out to determine the extent of DNA methylation in the genome of *D. noxia* in order to determine if epigenetic regulation contributes to changes in virulence. To this end, the global levels of methylation as well as the methylation profiles of the different biotypes were investigated. The global and specific methylation results suggest an inversely proportional relationship between virulence levels and DNA methylation, and that methylation may be associated with increased virulence. This study, being the first of its kind for *D. noxia*, has provided the groundwork for future research into methylation of this insect, and adds to a growing body of knowledge on hemipterans.

W803: Plant Interactions with Pests and Pathogens
A Family of Novel Glycoside Hydrolase Genes Originated from Bacteria in Hessian Fly Genome

Xiaoyan Cheng[1], Sandra Garcés Carrera[2], R. Jeff Whitworth[2], Yoonseong Park[2], Stuart Jeffrey[3] and Ming Chen[1], (1)Kansas State University, Manhattan, KS, (2)kansas state university, manhattan, KS, (3)Department of Entomology, Purdue University, West Lafayette, IN

Levan and inulin, made up of hundreds or thousands of fructose, are important carbohydrate reserves in over 45000 plant species. However, we still have little knowledge on whether or how they are used in animals or insects if those plants were taken as food. Here, we cloned and isolated a novel family of glycoside hydrolase (GH) genes from Hessian fly (*Mayetiola destructor*), which encode levanases/inulinases (MDLs). The MDLs only show identities with bacterial levanase and inulinase, while probable MDL homologous in other insects show similarities with each other. The phylogenetic tree also shows MDLs family could be clustered into a branch, which is different from those homologous in plants and bacteria. It is suggested that the new genes are horizontally transferred from *Bacillus*, and followed by gene duplication and diversification in host Hessian fly. The new MDLs consist of a group of sixteen members that have never been found in animals before. qPCR results confirm gene family members are expressed throughout the whole life cycle of Hessian fly, including three-instar stages of larvae, pupae, female and male adults. Twelve family members show higher levels of expression in larvae than in pupae and adults. And the remaining four members are expressed most in female adults. Family members are also expressed in different tissues, such as salivary glands, midgut, fat bodies and malpighian tubules. Three gene members are expressed most in salivary glands, six genes in midgut, four genes in fatty bodies, and two genes in malpighian tubules. Only one gene was expressed with similar level in four organs. Enzymatic activity assays revealed that protein extracts from whole insects can digest both levan and inulin *in vitro*. Furthermore, nine of recombinant MDL proteins are also able to digest both two carbohydrate substrates. Therefore, we confirm that MDLs family in Hessian fly could assist the insect in making good use of host carbohydrates. In addition, the expression pattern of this family gives us a valuable clue that that fructose may function as another important energy molecule in Hessian fly compared with the conventional glucose.

W804: Plant Interactions with Pests and Pathogens
Functional Characterization of the Predicted G-Protein Coupled Receptor Genes in the Wheat Scab Fungus *Fusarium graminearum*

Jin-Rong Xu, Purdue University, West Lafayette, IN

Although the cAMP-PKA and MAP kinase pathways are well-characterized for their functions in the wheat scab fungus, upstream receptors responsible for recognizing flowering wheat heads have not been identified. In this study, we systematically characterized the 105 GPCR genes that may function upstream from these conserved signaling pathways. Whereas none of them was important for vegetative growth, one non-pheromone GPCR and 14 others were involved in sexual and asexual reproduction, respectively. Although the CFEM domain-containing GPCRs are dispensable for plant infection, mutants deleted of five GIP genes up-regulated during infection were reduced in virulence. The GIP1 mutant was defective in infection cushion formation, which was partially suppressed by exogenous cAMP. Deletion of GIP1 reduced PKA activities and GMK1 phosphorylation stimulated by floral organs. GIP2 and GIP3 were important for infectious growth after penetration. Interestingly, the GIP genes are in the same phylogenetic cluster belonging to a subfamily of 22 closely-related GPCRs with many of them upregulated significantly during plant infection. Only three members of this subfamily are conserved in other filamentous fungi, suggesting their species-specific expansion in this important pathogen. Overall, our data showed that the wheat scab fungus, a floral pathogen, has an expanded subfamily of infection-related GPCRs that may recognize plant signals for regulating various infection processes.

W805: Plant Interactions with Pests and Pathogens
Exploring the Role of a Barley UDP-Glucosyltransferase in *Fusarium* Head Blight Resistance

Gary J. Muehlbauer, University of Minnesota, St. Paul, MN

Fusarium graminearum is the primary causal pathogen of Fusarium head blight of wheat and barley. Trichothecene mycotoxins (e.g., deoxynivalenol and nivalenol) accumulate during *F. graminearum* infection and act as virulence factors and result in reduced grain quality. Using transcriptomics and yeast complementation, we isolated a barley UDP-glucosyltransferase (*HvUGT13248*) and showed that it conferred resistance in Arabidopsis to the toxic effects of deoxynivalenol (DON). Transgenic wheat overexpressing *HvUGT13248* exhibited both type II resistance (resistance to spread of disease symptoms in the spike) and field resistance. In addition, these transgenic plants exhibited resistance to both DON- and nivalenol (NIV)-producing *F. graminearum*. Biochemical analysis of the transgenic wheat showed that HvUGT13248 metabolizes DON and NIV to the less toxic DON-3-glucoside and NIV-3-glucoside, respectively. Transgenic barley overexpressing *HvUGT13248* exhibited root resistance when grown on DON-containing media. Taken together, our transgenic results indicate that early timing of gene expression results in trichothecence and FHB resistance. To further functionally characterize *HvUGT13248*, we used a barley TILLING

population and identified two nonsynonomous mutations located in the UDP-D-glucose binding site. Both of these mutations resulted in reduced tolerance to growth on DON, indicating that *HvUGT13248* is a major trichothecene resistance gene in barley. Resequencing of *HvUGT13248* from a large collection of barleys revealed only a few nonsynonomous changes that do not likely impact the function of the gene, indicating that *HvUGT13248* is highly conserved. Our results provide a path towards developing FHB and trichothecene resistant barley and wheat.

W806: **Plant long non-coding RNAs**
Space Filler to be Changed By Presenter
Mark A Beilstein, University of Arizona, Tucson, AZ

W807: **Non-coding RNA**
The Genome of *Cucurbita argyrosperma* (silver-seed gourd) Reveals Faster Rates of Protein Coding and Long Noncoding RNA Gene Turnover and Neofunctionalization within *Cucurbita*
Josué Barrera-Redondo[1], Enrique Ibarra-Laclette[2], Alejandra Vázquez-Lobo[3], Yocelyn Teresa Gutierrez-Guerrero[1], Guillermo Sánchez-de la Vega[1], Daniel Piñero[1], Salvador Montes-Hernández[4], Rafael Lira Saade[5] and Luis Enrique Eguiarte[1], (1)Instituto de Ecología, UNAM, Mexico City, DF, Mexico, (2)Instituto de Ecología A.C. (INECOL), Xalapa, VL, Mexico, (3)CIByC, UAEM, Cuernavaca, MR, Mexico, (4)Campo Experimental Bajío, INIFAP, Celaya, GJ, Mexico, (5)FES Iztacala, UNAM, Tlanepantla, EM, Mexico
Whole genome duplications are an important source of evolutionary novelties that change the mode and tempo at which genetic elements evolve within a genome. The *Cucurbita* genus experienced a whole genome duplication around 30 Mya, although the evolutionary dynamics of their coding and noncoding genes have not yet been scrutinized. Here, we analyzed the genomes of four *Cucurbita* species, including the novel genome assembly of *Cucurbita argyrosperma*, and compared their gene contents with five other members of the Cucurbitaceae family to assess the evolutionary dynamics of protein coding and long intergenic noncoding RNA (lincRNA) genes after the genome duplication. We report a higher rate of protein coding gene birth-death rate in the *Cucurbita* genomes compared to the rest of the Cucurbitaceae family. The genome of *C. argyrosperma* presented significantly fast evolutionary rates in gene families associated to pollination and transmembrane transport. LincRNA families showed high levels of gene turnover throughout the phylogeny and 67.7% of the lincRNA families in *Cucurbita* have evidence of birth from the neofunctionalization of previously existing protein coding genes. Our results suggest that the whole genome duplication in *Cucurbita* resulted in faster rates of gene family evolution through the neofunctionalization of duplicated genes.

W808: **Plant long non-coding RNAs**
Non-Coding RNAs in Root Growth Adaptation to the Environment
Jeremie Bazin[1], Richard Rigo[1], Michael Moisson[2], Celine Charon[1], Julia Bailey-Serres[3], Federico Ariel[2] and Martin Crespi[1], (1)Institute of Plant Sciences Paris Saclay, GIF SUR YVETTE, France, (2)Instituto de Agrobiotecnología del Litoral, CONICET, Universidad Nacional del Litoral., Santa Fre, Argentina, (3)University of California-Riverside, Riverside, CA
Non-coding RNAs have emerged as major components of the eukaryotic transcriptome. Long non- coding RNAs (lncRNAs) act either directly or are processed to shorter miRNA and siRNAs. ncRNAs (both long and small) act through specific ribonucleoprotein complexes to modulate the expression of mRNA targets. Genome-wide RNA sequencing in roots identified many Arabidopsis lncRNAs, such as si/miRNA precursors, antisense or intergenic lncRNAs. In addition, ribosome profiling has allowed us to more precisely define the coding potential of lncRNA and to pinpoint a role of small ORF translation in ta-siRNA biogenesis. Targeted analysis of lncRNA identified that the *ALTERNATIVE SPLICING COMPETITOR (ASCO)* lncRNA interacts with Nuclear Speckle RNA Binding Proteins (NSRs) to regulate alternative splicing (AS) patterns of several mRNAs linking ASCO action and alternative splicing. We then analyzed the effect of the knockdown of *ASCO* at genome-wide level and found that only a minor subset of genes overlapped with the AS defects of the *nsra/b* double mutant. In particular, a high number of deregulated and alternatively spliced genes in *ASCO* knockdown plants were related to the response to flagellin and biotic stress. In agreement, *ASCO*-deregulated plants are more sensitive to flagellin, exhibiting a significant arrest of primary root development. Furthermore, we demonstrated that *ASCO* also interacts with PRP8a, a key component of the spliceosome, indicating that *ASCO* function on AS involves the interaction with multiple splicing factors. Our results hint the existence of a dynamic network between lncRNAs and splicing factors to modulate transcriptome diversity during development, conditioning the response to environmental cues. The evolution of the non-coding genome could be integrated into the mechanisms adapting plants to different environments.

W809: **Plant long non-coding RNAs**
Maize Inflorescence Long Non-Coding RNAs: Linking Phenotypic Variation to the Functional Non-Coding Genome
Edoardo Bertolini, Donald Danforth Plant Science Center, St. Louis, MO

W810: **Plant long non-coding RNAs**
Novel *cis*-Elements with Mammalian Enhancer-like Epigenomic Signatures Reveal a Potential Link between 3'-end Processing and Transcriptional Enhancers in Plants
Julia Chekanova, Guangxi University, Nanning, China

W811: **Plant long non-coding RNAs**
Determining the Function of the Novel Long Intergenic Non-Coding RNA *CONSERVED IN BRASSICA1* in *Arabidopsis*
Brian D. Gregory, University of Pennsylvania, Philadelphia, PA
Marianne C. Kramer[1,2], Kyle Palos[3], Eric Lyons[3], Mark A. Beilstein[3], and Brian D. Gregory[1,2]
[1]Department of Biology, [2]Cell and Molecular Biology Graduate Program, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]School of Plant Sciences, University of Arizona, Tucson, AZ 85721 USA.

Long intergenic non-coding RNAs (lincRNAs) are an emerging class of molecules that are gaining attention for their roles in many biological processes. There are over 6,000 detectable lincRNAs in *Arabidopsis* with tissue specific and stress responsive expression patterns. Despite these observations, the functions of most lincRNAs remain unknown.
Previously, we identified a protein-bound nuclear lincRNA *CONSERVED IN BRASSICA1* (*CONBR1*) that is highly conserved in the closest related crop species. To determine the function of this lincRNA, we examined the phenotype of homozygous mutants and found that mutants lacking lincRNA expression were noticeably smaller and developmentally delayed compared to wild-type plants. Additionally, *conbr1* mutant plants begin to senesce earlier than wild-type plants.
To examine the molecular function of this lincRNA, we have performed chromatin isolation by RNA precipitation (ChIRP) followed by DNA sequencing. With this technique, the lincRNA is pulled down using biotinylated probes complimentary to the lincRNA and the associated DNA is identified by high-throughput sequencing (ChIRP-seq). From ChIRP-seq, we found *CONBR1* binds to 94 genes which are enriched for proteins that are involved in phospholipid biosynthesis and cuticle development. Given the phenotype of early senescence it appears that *CONBR1* might function in cell wall maintenance and senescence. The most recent results for this project will be presented during this talk.

W8**12: Plant long non-coding RNAs**
Identification and Analysis of Salt Responsive Long Noncoding RNAs in Brassicaceae
Kyle R Palos, University of Arizona, Tucson, AZ
Space Filler

W8**13: Plant Molecular Breeding**
AE Boosting: A Neural-Network Strategy to Enhance Prediction on Wheat Yield through Spectral Information
Chun-Peng James Chen, Paul Mihalyov, Michael Pumphrey and Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA
Crop yield is known for its mild heritability and is barely consistent across environments. The variation can be dissected in several aspects, and it's believed that spectral measurements can be one of the efficient ways to alleviate such imprecision. In order to better capture the environmental factors, spectral reflectance measured in the early growing stage have been pairwise derived to vegetation indices, which show a linear correlation to crop's vitality. These indices include normalized difference vegetation index (NDVI), photochemical reflectance index (PRI) and normalized chlorophyll pigment ratio index (NCPI). However, none of these indices can properly address nonlinear variation or variation caused by higher-order combination of spectra. We therefore demonstrate a strategy that can explicitly utilize this type of information by training collected data in autoencoders (AE), an unsupervised neural network, applying features derived from AEs to boost prediction on the trait of interest. There were three years of records for wheat yield being used to examine the proposed strategy in this study. With the aid of AE boosting, couples of conventional genomic selection (GS) models, such as gBLUP and rrBLUP, can gain significant improvement by 5% ~ 7% on predicting yield from one year to another. In addition, by investigating the AE-derived features, several novel combination of spectral bands were suggested to potentially serve as vigor proxies, which is conceptually similar to those vegetation indices reported from the past studies. And the reported region of wavelength can also be further investigated by hyperspectral data, providing a promising direction to fit the environmental variation for plant breeding.

W8**14: Plant Molecular Breeding**
Functional QTL Mapping and Genomic Prediction of 3D Height Measured from a Robotic Field Scanalyzer in Wheat
Danilo H Lyra[1], Nicolas Virlet[2], Kirsty L. Hassall[1], Luzie U. Wingen[3], Simon Griffiths[3], Malcolm Hawkesford[2] and Gancho T. Slavov[1], (1)Department of Computational & Analytical Sciences, Rothamsted Research, Harpenden, United Kingdom, (2)Department of Plant Sciences, Rothamsted Research, Harpenden, United Kingdom, (3)John Innes Centre, Norwich, United Kingdom
Both fundamental and applied genetic studies increasingly rely on high-throughput phenotyping platforms. We used data from the Field Scanalyzer (LemnaTec GmbH) at Rothamsted Research (UK) to perform quantitative trait loci (QTL) scans and genomic prediction in the Chinese Spring × Paragon wheat recombinant inbred line population (N = 198 lines, M = 2100 SNPs), comparing several analytical approaches (i.e., interval/composite mapping and prediction of single-timepoint phenotypes *vs* functional mapping and GBLUP prediction of smoothed/dimension-reduced data). We identified four significant QTL (on chromosomes 2B, 5A, 5B, and 7A), with both empirical and simulation analyses demonstrating superior statistical power of functional mapping approaches compared to conventional single-timepoint analyses. Similarly, using smoothed (B-splines) and dimension-reduced (first principal component) phenotypic data (from 26 timepoints) significantly improved the predictive abilities (r = 0.41 and 0.33, respectively) compared to the timepoint with the highest r and the last timepoint (r = 0.36 and 0.16, respectively). In addition, the inclusion of significant QTLs as fixed-effect covariates in the GBLUP model resulted in substantial improvements in predictive ability. Although the statistical power of QTL detection, predictive ability, and genomic heritability increased with the number of timepoints analysed (i.e., 5, 10, and 26), the gains beyond 10 timepoints had relatively small practical significance. These results will inform the development of an integrated, semi-automated analytical pipeline with potentially broad applicability.

W8**15: Plant Molecular Breeding**
Streamlining Breeding Strategies through Data Mining of Genomics, Phenomics, and Environments
Tingting Guo[1], Xianran Li[1], Adam E. Vanous[1], Qi Mu[1], Randall J. Wisser[2] and Jianming Yu[1], (1)Iowa State University, Ames, IA, (2)University of Delaware, Newark, DE
Data mining and knowledge discovery have been attracting a significant amount of attention. In agriculture, enormous data from genomics, phenomics and environments are generated daily. With the feasibility, effectiveness, and scalability of data mining techniques, rethinking and redesigning the selection and breeding process become pertinent. Changes can be proposed to establish genotype-phenotype relationship so that efficient designs can be made to optimize genomic prediction of hybrid performance. Changes can also be made to establish genotype-

phenotype-environment relationship so that hidden patterns and specific factors underlying phenotypic plasticity can be identified and utilized to conduct in-season and on-target performance prediction.
To efficiently establish genotype-phenotype relationship, we conducted a multi-specie (maize, wheat, and rice) hybrid performance prediction study. Maize hybrids (276) were generated from diverse founder inbreds, 2,556 wheat hybrids were from an early-stage hybrid breeding system, and 1,439 rice hybrids from an established hybrid breeding system. Patterns of genomic relationships and phenotypic variation were systematically explored from clustering, graphic network analysis, and genetic mating scheme perspective to optimize training set design. Our analysis showed that optimized training set designs significantly outperformed random sampling and other methods that consider either minimizing the prediction error variance (PEV) or maximizing the generalized coefficient of determination (CD). Design optimization and pattern mining are expected to further enhance future studies of complex traits in crops.
To explore genotype-phenotype-environment relationship, we conducted a multi-environment trial for a rice population with 176 recombinant inbred lines (RILs) across 3 years. A complex flowering time variation across environments, or phenotypic plasticity, was observed. The underlying plasticity genes (*Hd1*, *Hd2*, *Hd5*, and *Hd6*) and their interactions dynamically responded to the environments their exposed. The integration of environmental index and genomics enables to predict the performance of untested genotypes in untested environment. Tracing the haplotype network and geographic distribution of these four genes of the global rice germplasm (The 3,000 Rice Genomes) indicated preferential alleles were selected for local adaptation. Knowledge discovered from data mining greatly improves our strategies in modern plant breeding.

W816: **Plant Molecular Breeding**
Dissecting a Major Spike Architecture QTL in Bread Wheat
Kai P Voss-Fels, The University of Queensland, St Lucia, QLD, Australia

W817: **Plant Molecular Breeding**
Precision Phenotyping Enables Discovery of QTLs in Walnut (*Juglans regia* L.)
Gina M. Sideli, Department of Plant Science, University of California, Davis, Davis, CA

W818: **Plant Molecular Breeding**
Distribution of *Yr15* Alleles in Wild Emmer Natural Populations and Cultivated Wheat Collections Revealed by Diagnostic Markers
Valentina Klymiuk, Andrii Fatiukha and Tzion Fahima, Institute of Evolution and the Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel
Stripe rust, caused by the fungus *Puccinia striiformis* f. s. *tritici* (*Pst*), is a destructive disease of wheat globally. One of the most promising genes conferring broad-spectrum resistance to stripe rust is *Yr15*, derived from wild emmer wheat (*Triticum dicoccoides*; WEW) accession G25. *Yr15*, located on chromosome arm 1BS, was recently cloned and designated as *Wheat Tandem Kinase 1* (*WTK1*), since it encodes a protein with kinase-pseudokinase domain structure. The available wheat reference genomes were used to develop diagnostic co-dominant KASP markers that can differentiate between functional (*Wtk1*) and non-functional (*wtk1*) alleles of *Yr15*. More than 540 wheat cultivars and breeding materials, originated from 65 countries, were screened with these allele-specific markers. *Wtk1* allele was found only in 32 introgression lines that carry 1BS chromosome segment from G25, while all the rest contained *wtk1* allele. We have used these markers to follow the distribution of *WTK1* alleles in WEW natural populations (382 accessions) across the Fertile Crescent region. Most of the tested accessions (82%) harbored the *wtk1* allele, while only 18% showed the presence of *Wtk1*. The geographic localization of WEW populations that carried *Wtk1* was assigned to a narrow region along an axis of 140 km from Mt. Carmel to Anti-Lebanon mountain range, at elevation of above 500 meters for most of the populations. The development of these markers designed based on polymorphism between the *WTK1* alleles can facilitate the introgression of *Yr15* into new varieties via marker-assisted selection, while avoiding negative linkage drag.

W819: **Plant Phenotypes**
Robotic Phenotyping at the Root-Soil Interface
Erin Sparks, University of Delaware, Newark, DE
Damage to plants that prevents them from staying upright, called lodging, can have a significant impact on cereal crop yield. In maize plants, roots that emerge from the stem above the soil, called brace roots, are proposed to play an important role in structural stability. Yet how brace roots develop, integrate environmental cues and contribute to plant stability remains a poorly understood area of plant biology. Research in our lab focuses on questions regarding the development and function of maize brace roots. To determine if brace roots promote plant stability, we used field-based flex testing to show that brace roots do significantly contribute to plant flexural strength. To understand what features of brace roots might impact this stability, we are working to define the diversity of brace root phenotypes across genotype and environment. To facilitate high-throughput brace root phenotyping, we have developed a ground-based robot tailored for imaging and analysis at the root-soil interface. We will discuss the development of this robot platform and the diversity of brace root phenotypes we have collected.

W820: **Plant Phenotypes**
Phenomics of Intact Grains and Spikes to Facilitate Selection for Grain Filling Attributes
Fernanda Dreccer[1], José Jiménez-Berni[2], Antonio Robles-Kelly[3], Ran Wei[4], Gemma Molero[5], Carolina Rivera-Amado[5], Stephen Gensemer[6], Carlos Guzmán[7], Francisco Pinto[8] and Matthew Reynolds[5], (1)The Commonwealth Scientific and Industrial Research Organisation, Clayton South Vic, Australia, (2)Instituto de Agricultura Sostenible, Córdoba, Spain, (3)Australian National University, Canbera City, Australia, (4)Data61 - Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canbera, ACT, Australia, (5)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (6)The Commonwealth Scientific and Industrial Research Organisation, GLEN OSMOND, Australia, (7)Global Wheat Program,

International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (8)International Maize and Wheat Research Center, CIMMYT, Texcoco CP, Mexico
Selecting for grain filling attributes is a challenge for plant breeders. We explored the potential for detecting differences in grain water content using a hyperspectral camera covering the range from 900 to 1700 nm. We have successfully developed a pipeline to segment and analyse images and used machine learning techniques, including feed forward and deep networks, to discriminate for grain water content allowing for the detection of changes in grain growth during grain filling.

W821: **Plant Phenotypes**

Post-Harvest Spike Phenotyping and Grain Analysis of Wheat Using X-Ray CT

Joelle Claussen[1], Norbert Woerlein[1], Anja Eggert[1], Jessica Schmidt[2], Trevor Paul Garnett[2], Norman Uhlmann[1] and Stefan Gerth[1], (1)Fraunhofer IIS, Fuerth, Germany, (2)The University of Adelaide, Glen Osmond, Australia
During the last years, X-ray technology has been applied for the non-destructive visualization of optical inaccessible structures in plants and seeds. Formerly, this technology was only used for medical imaging. Nowadays, it is used as a standard tool in industrial applications for material analysis. With X-ray computed tomography (CT) the 3D volume information of objects can be reconstructed using X-ray projections of the object from different points of view. A conical X-ray beam projects the plant on a 2D flat panel detector. The complete 3D volume information contains the exact position of each object and their morphologic structure of internal traits or the surface. The resulting spatial sampling frequency is determined by the geometrical magnification and the pixel size of the flat panel detector.
Due to the non-destructive nature of CT it is possible to analyze the inner structure of seeds. Individual seed imaging allows highest resolution in spatial sampling frequency with the disadvantage of long measurement times. There are several ways to increase the throughput in comparison to individual seed imaging. One of the options is to measure a box of seeds and do the image analysis still on individual seed level. This results in a decrease in time in the pipeline of seed analysis. Another option is to analyze the seeds prior labor intensive manual processing. Thus, this method uses the ear instead of the threshed seeds. It is possible to analyze the among of seeds in the ear and extract their morphologic structure. The developed algorithm distinguish every single seed. This enables the quantification of the position, the volume, the density – resulting in a virtual biomass – the diameter and the aspect ratio of the single seed within the wheat ear. Even though the seed is not fully developed it is possible to already detect and analyze them in an early state.
In the presentation, we present the complete imaging pipeline for boxes of seeds and for wheat ears. We demonstrate then the possibility of high throughput wheat analysis. This approach allows non-destructive trait extraction with an increased throughput compared to manual wheat ear analysis. Additionally, this approach allows the first time to analyze the seeds in the ear with non-destructive methods. This is the opportunity to have a post-harvest phenotyping setup.

W822: **Plant Phenotypes**

Phenotyping Intractable Soybean Canopy Architecture Traits

Suma Sreekanta[1], Allison Haaning[1], Kamaldeep S. Virdi[1], Anna N. Hofstad[1], Austin A. Dobbels[1], Jacob Meierhofer[2], Danica Mooney-Jones[2], Fumiaki Katagiri[3], Aaron Lorenz[4], Robert M. Stupar[1] and Gary J. Muehlbauer[1], (1)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (2)College of Biological Sciences, University of Minnesota, St. Paul, MN, (3)Department of Plant and Microbial Biology, University of Minnesota, St.Paul, MN, (4)University of Minnesota, St. Paul, MN
Canopy Architecture (CA) is a result of complex interplay between many traits. In crops such as maize and wheat, altering CA has resulted in enhanced yield. However, study of CA has been limited to a few traits because measuring CA traits has traditionally been a slow, low throughput process. Soybean is not an exception to this bottleneck. Our study focuses on better understanding soybean CA by seeking to answer two important questions 1) What traits contribute towards the overall CA and light interception? 2) What is the genetic control underlying CA? To answer these questions it is critical to be able to phenotype the CA traits accurately. We used a combination of high-throughput technologies including an unmanned aircraft system as well as inexpensive smartphone images combined with open source software to parameterize CA with more than 50 different traits. An advantage of our system is that it allows detailed examination of CA in field-grown plants. Forty different genotypes selected for strong visual variation in CA were evaluated using this platform in the field. We found that canopy coverage and light interception are highly correlated, and that height and plant shape attributes strongly contribute to canopy coverage measured at 45, 60 and 75 days after planting. The traits defining the top seven nodes constituting the crown of the soybean canopy also significantly affects light interception and canopy coverage. We were able to balance accuracy with scale of phenotyping required to study many highly plastic CA traits. With this study, we were able to quantitatively define CA in terms of traits that were either ill-defined or undefined previously. We used similar imaging and measurement techniques on a panel of 400 USDA genotypes and found the major QTL underlying canopy coverage and branch angle traits were coincident, suggesting that branch angle may be a major factor in determining canopy coverage in soybeans. Further study of these traits will allow for the modification and selection of CA in more precise and deliberate ways.

W823: **Plant Phenotypes**

Implementation of Unmanned Aerial Systems as a Tool in the Texas Maize (*Zea mays* L.) Breeding Program

Steven Anderson[1], Lonesome Malambo[2], Anjin Chang[3], Marcin Jurek[4], Sorin Popescu[2], Jinha Jung[3], Dale Cope[5], Alex Thomasson[6] and Seth Murray[1], (1)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, (2)Department of Ecosystem Science and Management, Texas A&M University, College Station, TX, (3)School of Engineering and Computer Sciences, Texas A&M University - Corpus Christi, Corpus Christi, TX, (4)Department of Statistics, Texas A&M University, College Station, TX, (5)Department of Mechanical Engineering, Texas A&M University, College Station, TX, (6)Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX
Unmanned aerial systems (UAS) accompanied with light-weight airborne sensors present the opportunity to rapidly phenotype breeding populations, aiding in dissection of genetics underlying quantitative traits in plants. In the Texas environment, maize (*Zea mays* L.) plant height (PHT) is highly correlated ($r = 0.61$) to grain yield and is hypothesized to indicate abiotic stress tolerance. Unfortunately, PHT is a resource intensive phenotype that is commonly collected at the conclusion of the growing season for maize (*Zea mays* L.). Through the implementation of

UAS technologies we are now capable of assessing PHT throughout the growth season using structure-from-motion (SfM) algorithms with very little resource inputs. Weekly UAS imaging was collected at the Texas A&M research farm for three years (2016-2018) across three genetic mapping populations the national Genomes to Fields hybrid trial across two (dryland, irrigated) and three (dryland optimal planting, irrigated optimal planting, irrigated delayed planting) environmental treatments, respectfully.

W824: **Plant Phenotypes**

Phenotyping Acclimation Capacity in Maize and Arabidopsis

Astrid Junker[1], Henning Tschiersch[1], Stephan Erbe[1], Andrea Braeutigam[2], Martin Mascher[3], Andreas Boerner[4] and Thomas Altmann[5], (1)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Stadt Seeland, Germany, (2)Bielefeld University, Bielefeld, Germany, (3)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany, (4)IPK Gatersleben, Stadt Seeland, Germany, (5)IPK Gatersleben, Gatersleben, Germany

In its natural environment plants continuously have to cope with short-term changes in environmental conditions such as light or temperature. It is known that, compared to constant environments, fluctuating conditions negatively impact plant performance and biomass production. Predictions of future climate scenarios foresee even more extreme fluctuations. Thus a detailed understanding of acclimation capacity, defined as the ability to lower the fluctuation-induced reduction in plant performance, is of utmost importance for securing yield stability. Furthermore accessions with higher acclimation capacity will enable to identify novel breeding targets for the future.

We use high throughput plant phenotyping facilities to monitor architectural and physiological changes in response to changing conditions (Junker at al. 2015, Tschiersch et al. 2017). Thereby we assessed growth dynamics and photosynthetic performance during early seedling establishment and vegetative development of about 300 IPK Genebank maize accessions under constant and changing temperature regimes (benchmarked against commercial hybrids). We were able to identify candidate accessions with superior performance in cold-tolerance and early photosynthetic efficiency whereas the latter was found to discriminate hybrids and GB accessions through a machine learning based approach. We furthermore identified variation in photosynthetic acclimation to switches in light intensity in both maize and Arabidopsis which is currently subject of an association study in a F2 cross population.

The implementation of a novel and unique plant growth and phenotyping facility at IPK will allow to extend these studies in a controlled and reproducible field-similar setup to enable the dissection of complex traits and to gain a deeper understanding of dynamic acclimation-related processes, especially in the context of early vigor.

W825: **Plant Reproductive Genomics**

Evolutionary Development of Flowers within the Ranunculaceae

Hongzhi Kong, Institute of Botany, Chinese Academy of Sciences, Beijing, China

Flowers show tremendous diversity in size, color, structure, symmetry, display, and function, etc. The mechanisms underlying the origin and diversification of the flower, however, remain largely unclear. One of the obstacles that prevent us from knowing more about the mechanisms of flower evolution is the lack of suitable systems. In the past few years, we and collaborators have developed the basal eudicotyledonous family Ranunculaceae into a model to address important evolutionary developmental questions that cannot be easily addressed by using any other existing models. In this talk, I will first introduce the advantages of using the Ranunculaceae as a model and then present our new results on four topics: 1) the molecular mechanism underlying parallel petal losses within the Ranunculaceae; 2) the molecular basis and its flexibility of the floral organ identity determination program in *Nigella damascena*, a species with spiral rather than whorled flowers; 3) the tempo, mode, and mechanisms of character evolution during petal elaboration within the genus *Nigella*; and 4) the possible mechanisms that underlie the diversification of the basic structure of the petals in the Ranunculaceae. Taking together, these results suggest that the Ranunculaceae is indeed a promising model for the study of plant developmental evolution.

W826: **Plant Reproductive Genomics**

Protein-Protein Interaction Profiles of MADS-Box Dimer Evolutionary Variants

Maria Jazmin Abraham Juarez, Biology Department, Amhert, MA

W827: **Plant Reproductive Genomics**

Regulatory Variation Controlling Flowering Time and Inflorescence Architecture in Panicoid Cereals

Andrea L. Eveland, Donald Danforth Plant Science Center, St. Louis, MO

W828: **Plant Reproductive Genomics**

Diverse Sex Chromosomes in the Salicaceae

Matthew S. Olson, Texas Tech University, Lubbock, TX

W829: **Plant Reproductive Genomics**

The Genome of *Amborella trichopoda* exhibits Ancestral Gene Order and Derived, Young Sex Chromosomes

Adam J. Bewick, University of Georgia, Athens, GA

W830: **Plant Reproductive Genomics**

Phylogenomic Analyses uncover Ancient but Highly Dynamic Moss Sex Chromosomes

Sarah Carey, University of Florida, Gainesville, FL

Sex chromosomes have evolved several times across the tree of life. The bryophytes, whose ancestor is thought to be dioecious (i.e. separate sexes), provide novel systems for understanding the evolution of ancient sex chromosomes. Given their haploid dominant nature sex in dioecious bryophytes is determined by a UV sex chromosomal system. In this system, each sex has a non-recombining chromosome (U for females and V

for males) that pair at meiosis in the monomorphic sporophyte and segregate to the male and female gametophytes. Because the sex chromosomes are transcriptionally active in the haploid stage and therefore subject to purifying selection we expect many orthologous genes will be retained between the U and V chromosomes. The moss *Ceratodon purpureus* has UV sex chromosomes that constitute ~100 megabases (MB) of the 360 MB genome. Using a combination of Illumina, PacBio, and Hi-C data we have assembled genomes of a male and a female isolate of *C. purpureus* to chromosome scale. Here we use genes annotated on the U and V sex chromosomes of *C. purpureus* and existing transcriptome data of other moss species to determine the age of moss sex chromosomes. We find that moss sex chromosomes are ancient with multiple capture events of genes throughout their evolution. We also show that moss sex chromosomes are dynamic. Most notably, when we map orthologous genes of the multiple capture events to the U and V sex chromosomes we find there is no sign of evolutionary strata, suggesting genes are moved soon after they are captured.

W831: **Plant Transgene Genetics**
Genetic Improvement of Potato by Innate® and Gene Editing Technologies
Hui Duan, J. R. Simplot, Boise, ID

W832: **Plant Transgene Genetics**
Recombinase Technology for Gene Stacking; From Microbes to Plants
James Thomson[1], Leyla Hathwaik[1], Upul Hathwail[1], Troy Weeks[2], Nic Nottingham[2], Manmeet Singh[2], Hui Duan[2] and Roger Thilmony[1], (1)USDA-ARS, Albany, CA, (2)J. R. Simplot, Boise, ID
Recombinase-mediated genetic engineering provides a favorable direction for precise gene stacking. The process termed Recombinase Mediated Cassette Exchange (RMCE) can be applied to any genome or vector system. Recent work from our labs has demonstrated the use of RMCE for T-DNA building directly into the *Agrobacterium* virulence plasmid and its concomitant transfer to the plant genome with high fidelity. This stacking system (termed GA*A*NTRY) appears to have an exceptionally large capacity for gene stacking limited only by the biology of the microbes T-DNA transfer system. The initial GA*A*NTRY system was built, tested and published within the *Agrobacterium rhizogenes* strain ArPORT1. This strain appears to work very well in most monocots works in some dicot species. To further the utility of this system, we have generated the *Agrobacterium tumefaciens* JGT105 GA*A*NTRY strain (a derivative of EHA105). Here we present data demonstrating the successful use of the GA*A*NTRY gene stacking system with the JGT105 strain and its successful application in both Arabidopsis and potato.

W833: **Plant Transgene Genetics**
Development of Transgene-Free Gene Editing Technology
Yunde Zhao, Section of Cell and Developmental Biology, University of California, San Diego, San Diego, CA

W834: **Plant Transgene Genetics**
Using Agrobacterium-Mediated Transient Cas9 and sgRNA Expression to Produce Genome-Edited Perennial Plants that are Transgene-Free
Yi Li, University of Connecticut, Storrs, CT
Developing genome-edited and transgene-free asexually propagated perennial crop plants is challenging but highly desirable. We have developed a highly useful method using Agrobacterium-mediated transient Cas9 and sgRNA expression to create transgene-free mutant plants without the need for sexual segregation. We have also developed a rapid, cost-effective, and high-throughput mutant screening protocol based on Illumina sequencing followed by high-resolution melting (HRM) analysis. Using tetraploid tobacco as a model species and the phytoene desaturase (PDS) gene as a target, we successfully created and expediently identified transgene-free pds mutant plants that are transgene-free. Our method may provide a tool to produce genome-edited and transgene-free plants that are heterozygous, perennial or difficult to regenerate.

W835: **Plant Transgene Genetics**
Expanding the Capabilities for Plant Genome-Editing and Synthetic Biology
Degao Liu[1], Henrique De Paoli[2], Hua Cheng[1], Brian Mendoza[3], Linling Li[1], Rongbin Hu[1], Mei Chen[4], Cong Trinh[3], Gerald Tuskan[1] and **Xiaohan Yang**[1], (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)Lawrence Berkeley National Laboratory, De Paoli, CA, (3)University of Tennessee, Knoxville, TN, (4)Southwest University of Science and Technology, Mianyang, China
Genome-editing and synthetic biology have great potential in basic and applied plant science research. The CRISPR-Cas9 systems have been widely used for genome-editing in plants. However, the technical advancement in application of the CRISPR/Cas9 technology in plants is lagging behind that in animals and microbes. To enhance the capability of genome-editing in plants, we demonstrate for the first time the CRISPR-mediated cytidine deaminase editor as an efficient technology for disrupting plant genes through induction of pre-mature STOP codons (iSTOP). Also, we show that the adenine deaminase editor is effective for multiplex base-editing in dicot plants. Importantly, assembling DNA parts into gene constructs or genetic circuits is critical for synthetic biology research. Recently, we developed a high-throughput platform for rapid, flexible DNA assembly from universal libraries to generate multi-gene constructs. It features a pre-defined three-nucleotide (TNT) signature and a buffer system for a quick one-pot reaction. This cloning system automatically maintains the open reading frames of protein-coding sequences and does not require linkers, adaptors, sequence homology, amplification or mutation of DNA fragments in order to work properly. This new technology is useful for synthetic biology research in a wide range of organisms including plants, microbes and animals.

W836: **Plant Transgene Genetics**
Speed Breeding to Accelerate Crop Research and Breeding
Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

W837: **Polyploidy**
Whole Genome Duplication Shapes Pollinator-Mediated Assortative Mating and Herbivore Diversity
Robert Laport, Rhodes College, Memphis, TN

W838: **Polyploidy**
Subgenome Merger in *Brassica* Hybrids Supports a Novel Form of Hybrid Speciation
Annaliese S Mason, Justus Liebig University, Giessen, Germany

W839: **Polyploidy**
Comparative Genomics of Two Carnivorous Sundew Plants (Drosera)
Victor A. Albert, University at Buffalo, Buffalo, NY
Recent studies of carnivorous plant genomes (*1-3*) have focused on genome-scale issues in the evolution of carnivory. Among other topics, research has centered on functional enrichments in duplicate gene space (*4*). The tiny 100 Mb genome of the humped bladderwort (*Utricularia gibba* (*2-3*)) is at least octoploid relative to grapevine, whereas the large, at least 1.6 Gb genome of the Australian pitcher plant (*Cephalotus follicularis* (*1*)) has not duplicated past the *gamma* triplication shared with *Vitis*. These two species represent two independent derivations of the carnivorous habit. We recently expanded genome sequencing for a third independent lineage of carnivorous plants, the genus *Drosera* (sundews, Caryophyllales). Nearly all *Drosera* species are thought to bear holocentric chromosomes, wherein kinetochores attach throughout chromosome arms. We used Dovetail Chicago and HiRise Hi-C data to perform chromosome-level scaffolding of PacBio draft assemblies for *D. regia* (N50 = 15.8 Mb) and *D. capensis* (N50 = 11.2 Mb), species reported to be monocentric and holocentric, respectively. *D. regia* and *D. capensis* harbor long tandem arrays of structurally similar, presumed centromeric repeats of length 171 and 160 bp, respectively. The arrays are broadly distributed across scaffolds in both species, unexpectedly in *D. regia*, and even more so than in *D. capensis*. Syntenic analyses revealed no post-*gamma* polyploidies prior to their species split, after which an independent genome triplication occurred in *D. regia*, followed by a *D. capensis*-specific genome duplication. Dating of the species split between these two morphologically stereotyped, sticky-finger trappers indicated a split ~80 million years before present (Mybp), based on Ks values for orthologous syntelogs and a calibration of ~120 Mybp for the *gamma* event (*5*). Among tandemly duplicated *Drosera* genes, significantly enriched GO categories include defense response and secondary metabolism, a finding congruent with the hypothesis that defense-related genes have been co-opted for plant carnivory (*6*). One gene family of particular interest for carnivory is the cysteine proteases, which are tandemly expanded in both *Drosera* species as well as in *U. gibba*. However, the *Drosera* gene family expansion, which in part predates evolution of the genus (because Venus flytrap – *Dionaea muscipula* – cysteine proteases are nested within), is entirely independent. Similarly to the situation in *U. gibba* (*2*), however, tandem expansion of genes encoding these proteolytic enzymes was potentiated by polyploidy events.
(1) Fukushima K, et al. Nature Ecology & Evolution, 2017. 10.1038/s41559-016-0059.
(2) Lan T, et al. Proceedings of the National Academy of Sciences, 2017. 10.1073/pnas.1702072114.
(3) Ibarra-Laclette, et al. Nature, 2013. 10.1038/nature12132.
(4) Renner T, et al. "Carnivorous plant genomes". Carnivorous Plants: Physiology, Ecology, and Evolution. Oxford University Press, 2018. 10.1093/oso/9780198779841.003.0011
(5) Jiao Y, et al. Genome Biology, 2012. 10.1186/gb-2012-13-1-r3
(6) Renner T, Specht CD. Molecular Biology and Evolution, 2012. https://doi.org/10.1093/molbev/mss106.
Coauthors
Tanya Renner (The Pennsylvania State University; PSU)
Tianying Lan (University at Buffalo; UB)
Sitaram Rajaraman (University of Helsinki)
Luis Herrera-Estrella (Texas Tech University; LANGEBIO, CINVESTAV)
Enrique Ibarra-Laclette (INECOL, Mexico)
Martín Mata-Rosas (INECOL)
Chunfang Zheng (University of Ottawa; UO)
David Sankoff (UO)
Qiao Wen Tan (Nanyang Technological University; NTU)
Chloe Drummond (PSU)
Stephan C. Schuster (NTU)
Todd P. Michael (JCVI)
Jarkko Salojärvi (NTU)

W840: **Polyploidy**
Premeiotic Endomitosis Facilitates Recombinant Apomixis in Polyploid Land Plants
Amanda Grusz, University of Minnesota Duluth, Duluth, MN

W841: **Polyploidy**
Intergeneric Allopolyploidy in the Moss Family Funariaceae Revealed through Targeted Sequencing of Gametophytes
Matthew G Johnson, Department of Biological Sciences, Texas Tech University, Lubbock, TX

W842: **Polyploidy**
Ecological Adaptation of Polyploidy in Response to Environmental Change

Na Wei[1], Richard Cronn[2], Aaron Liston[3] and Tia-Lynn Ashman[1], (1)University of Pittsburgh, Department of Biological Sciences, Pittsburgh, PA, (2)USDA Forest Service Pacific Northwest Research Station, Corvallis, OR, (3)Oregon State University, Department of Botany and Plant Pathology, Corvallis, OR
Polyploidy, or whole genome duplication often with hybridization, is common in eukaryotes and is thought to drive ecological and evolutionary success especially in plants. The mechanisms of polyploid success in ecologically relevant contexts, however, remain largely unknown. Here we conducted an extensive test of functional trait divergence and plasticity in conferring polyploid fitness advantage in heterogeneous environments, by growing clonal replicates of a worldwide genotype collection of six allopolyploid and five diploid wild strawberry (*Fragaria*) taxa in three climatically different (i.e. cool coastal, temperate valley and arid montane) common gardens. Among leaf functional traits, we detected divergence in trait means but not plasticities between polyploids and diploids, suggesting that increased genomic redundancy in polyploids does not necessarily translate into greater trait plasticity in response to environmental change. Across the heterogeneous garden environments, however, polyploids exhibited fitness advantage, supporting a 'jack-and-master' hypothesis for polyploids. Specifically, polyploids benefit from stronger positive effects and weaker negative effects of trait means on fitness, and from stronger adaptive effects of trait plasticity, relative to diploids. Our findings elucidate essential ecological mechanisms underlying polyploid adaptation to heterogeneous environments, and provide important insight into the prevalence and persistence of polyploid plants.

W843: **Polyploidy**
Genome-Guided Phylo-Transcriptomics for Resolving a Hexaploid Lineage
R. Shawn Abrahams, University of Missouri, Columbia, MO
For polyploid lineages, phylogenomic methods that can account for the effect of paralogous genes are necessary in understanding species relationships, though most methods cannot. The tribe Brassiceae of the Brassicaceae shares a hexaploidy event, an effective triplication of their ancestral genome from a series of genome duplication events. Despite being the most economically important tribe of the mustard family, much of the evolutionary relationships remain obscured, with a history of conflicting molecular species trees and polyphyletic genera. In this study, we use a unique genome-guided phylo-transcriptomic method that uses genome derived synteny information to improve orthology detection over standard sequence similarity approaches. Using this method, we clarify clade relationships and detect the putative nodal placement of the tribal whole genome duplication events. We also use in-group synteny to identify putative signatures of an ancient introgression between major crop and orphan crop lineages of the tribe.

W844: **Population and Conservation Genomics 1**
Population Genomic Perspectives on Ice Age Mammal Biogeography in Southeast Alaska
Charlotte Lindqvist, University at Buffalo, Buffalo, NY
The Alexander Archipelago in Southeast Alaska holds a highly diverse fauna with a large number of endemic taxa and unique genetic lineages, suggesting a long history of isolation. Separation from the remainder of North America by the nearly impenetrable Coast Range Arc, a fragmented landscape, and a dynamic climatic history likely contributed to this diversity. For example, it has been suggested that ice-free areas along the North Pacific Coast may have served as refugia for plant and animal populations during the peak of the Ice Age, when ice sheets extensively curbed biotic exchange between the Old and New Worlds for thousands of years. Such ice-free areas or early deglaciation may have facilitated the postglacial colonization of the Americas, including by the first humans. Based on nuclear and mitogenomic analyses of both modern and ancient samples of several mammal species, in addition to radiocarbon dating of cave fossils and exposure dating of rock surfaces, we are testing the existence of ice-free refugia and studying the diversification and postglacial migration of animals following the Ice Age. For example, we have estimated using radiocarbon and exposure dating that the culmination of maximum ice extent in SE Alaska occurred ~20-17 thousand years ago. We find that pre-glacial bear lineages did not survive in local refugia during the Ice Age, but instead that bears living in the Alexander Archipelago today repopulated the islands later. This research has broad significance for assessing the impact of late Pleistocene climate change on mammal diversification and historical biogeography of the region, including the peopling of the Americas.

W845: **Population and Conservation Genomics 1**
A Large-Scale Study of Structural Variation in 145 Dogs Illustrates the Selection Patterns and Evolutionary Role of Copy-Number Variation
Jarkko Salojärvi, Nanyang Technological University, Singapore, Singapore
Copy-number variation (CNV) and mobile element insertions (MEI), such as SINEs and LINEs, are the major contributors to the structural variation between individuals. Here we called CNVs from whole genome sequencing data of 145 dogs and eight wolves using a hybrid pipeline incorporating several CNV calling software, and subsequently estimated MEIs in the same set, thus making it possible to study the selection patterns and co-occurrence dynamics between the different variant types.
Overall, the results support findings in human studies, indicating different selection pressures for the different CNV types. The MEIs and CNVs showed significant co-occurrence, with over 32% of CNVs being located in the proximity of MEIs. The presence of parent-progeny trios in the data allowed us to study the inheritance patterns and de novo mutation rates of the CNVs. Wolf individuals were used to polarize the site frequency spectrum and thus estimate the CNVs that have reached fixation after domestication. Finally, we addressed the impact of the CNVs to evolution in order Carnivora through a comparative analysis of occurrence patterns of CNVs and genomic variation in eight other species.

W846: **Population and Conservation Genomics 1**
Population Genomics of Environmental Change across Multiple Scales in the American Pika
Michael Russello, University of British Columbia, Kelowna, BC, Canada

W847: **Population and Conservation Genomics 1**
Demography or Selection on Linked Cultural Traits or Genes? Investigating the Driver of Low mtDNA Diversity in the Sperm Whale using complementary Mitochondrial and Nuclear Genome Analyses

Phillip A. Morin, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA
Mitochondrial DNA has been heavily utilized in phylogeography studies for several decades. However, underlying patterns of demography and phylogeography may be misrepresented due to coalescence stochasticity, selection, variation in mutation rates, and cultural hitchhiking (linkage of genetic variation to culturally transmitted traits affecting fitness). Cultural hitchhiking has been suggested as an explanation for low genetic diversity in species with strong social structures, counteracting even high mobility, abundance and limited barriers to dispersal. One such species is the sperm whale, which shows very limited phylogeographic structure and low mtDNA diversity despite a worldwide distribution and large population. We use analyses of 175 globally distributed mitogenomes and three nuclear genomes to evaluate hypotheses of a population bottleneck/expansion versus a selective sweep due to cultural-hitchhiking or selection on mtDNA as the mechanism contributing to low worldwide mitochondrial diversity in sperm whales. In contrast to mtDNA control region (CR) data, mitogenome haplotypes are largely ocean-specific, with only one of 80 shared between the Atlantic and Pacific. Demographic analyses of nuclear genomes suggest low mtDNA diversity is consistent with a global reduction in population size that ended approximately 125,000 years ago, correlated with the Eemian interglacial. Phylogeographic analysis suggests that extant sperm whales descend from maternal lineages endemic to the Pacific during the period of reduced abundance, and have subsequently colonized the Atlantic several times. Results highlight the apparent impact of past climate change, and suggest selection and hitchhiking are not the sole processes responsible for low mtDNA diversity in this highly social species.

W848: Population and Conservation Genomics 1

Evolution in the Extremes: Insights into the History of an Antarctic Midge

Hilary A. Edgington, The Ohio State University, Wooster, OH
Characterizing the evolutionary history of species in extreme environments can provide insights into broad questions of invasion and adaptation, in addition to clues about climatic or geological changes tied to the species' history. While a great deal of research has focused on the species inhabiting the extreme environment of Antarctica, the population genetics of most species inhabiting the continent has not been studied. We used a combination of population and landscape genetics, approximate Bayesian computation, and phylogenetics to investigate evolutionary history of *Belgica antarctica*, a wingless midge endemic to Antarctica. *B. antarctica* is a notable system for a number of reasons, including its physiological adaptations to the frozen environment and its exceptionally small genome.
Our study included genome-wide SNP data from eleven populations of *B. antarctica* located on islands throughout the Antarctic Peninsula. Contrary to expectations for island populations, we found an excess of heterozygosity in nearly all populations (F_{IS} = -0.354 - 0.031), coupled with low to moderate structure between populations within and among islands (F_{ST} = 0.007 - 0.160). This finding is especially surprising in this wingless species, for which ocean channels between islands were expected to present an effective barrier to migration. These results will inform not only our understanding of the population dynamics for species inhabiting these environments, but ongoing efforts to understand variation in physiological mechanisms allowing Antarctic species to adapt to that extreme environment.

W849: Population and Conservation Genomics 1

Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of the Sea Otter and Giant Otter

Annabel Beichman, University of California, Los Angeles, Los Angeles, CA
Despite its recent invasion into the marine realm, the sea otter (*Enhydra lutris*) has evolved a suite of adaptations for life in cold coastal waters, including limb modifications and dense insulating fur. This uniquely dense coat led to the near-extinction of sea otters during the 18th-20th century fur trade, leading to an extreme population bottleneck. We used the *de novo* genome of the southern sea otter (*E. l. nereis*) to reconstruct its evolutionary history, identify genes influencing aquatic adaptation, and detect signals of population bottlenecks. We compared the genome of the southern sea otter to the tropical freshwater-living giant otter (*Pteronura brasiliensis*) to assess common and divergent genomic trends between otter species. We found signals of positive selection in genes related to aquatic adaptations, particularly limb development in the sea otter and polygenic selection on genes related to hair follicle development in both otter species. We found extensive pseudogenization of olfactory receptor genes in both the sea otter and giant otter lineages, consistent with patterns of sensory gene loss in other aquatic mammals. These findings illuminate genetic mechanisms underlying aquatic adaptations in a species that has recently occupied marine environments, including positive selection on single genes, polygenic selection, and gene loss. At the population level, the southern sea otter showed extremely low genomic diversity, signals of recent inbreeding, and a demographic history marked by population declines that pre-date the fur trade and have resulted in an increase in putatively deleterious variants that could impact the future recovery of the population.

W850: Population and Conservation Genomics 1

Whole Genome Sequence Analysis of West African Taurine reveals their unique Trypanotolerant Adaptation

Abdulfatai Tijjani[1,2], Hassan Musa[3], Eui-Soo Kim[4], Bin Zang[4], Dajeong Lim[5], Tad S. Sonstegard[4], Oyekanmi Nash[2] and Olivier Hanotte[1,6], (1)Cells Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, (2)Centre for Genomics Research and Innovation, National Biotechnology Development Agency (NABDA), Abuja, Nigeria, (3)Faculty of Medical Laboratory Sciences, University of Khartuom, Khartuom, Sudan, (4)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN, (5)National Institute of Animal Science, Suwon, Korea, Republic of (South), (6)International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia
The West African taurine represent a unique group of indigenous African cattle population. They possess the ability to thrive in their local production environment where trypanosomiasis is endemic. In this study, we investigated the signatures of selection across the genome of two trypanotolerant West African taurine breeds (WAT), the shorthorn Muturu and the longhorn N'Dama, using six selection scan tests. Each of the trypanotolerant cattle was compared to two groups of trypanosusceptible cattle populations (African zebu and European taurine). Functional annotation of genes within candidate regions reveals genes involved in pathways, e.g. T cell and B cell activation and Natural Killer Cell mediated cytotoxicity, that are relevant to trypanosomiasis disease progression. The list of common genes between Muturu and N'Dama were investigated further. Our findings identify *PTPN6* involved in several of these pathways. Protein-protein network indicates its pivotal role through its interactions with bovine MHC class II genes and other genes, and therefore a possible central role of the gene in the initiation and control of a cascade of biological processes necessary to confer protective immunity linked to trypanotolerance. The importance of *PTPN6* is

further supported by the presence of an unique WAT haplotype closed to fixation. Our results support that *PTPN6* gene is linked to the trypanotolerance status of the WAT calling for further functional investigation of the gene.

W851: Population and Conservation Genomics 2
Comparing eDNA with Conventional Methods for Bio-Monitoring of Fish Communities
Louis Bernatchez, Université Laval, Québec, QC, Canada
Accurate data on distribution and abundance are critical for conservation and management of biodiversity. In aquatic ecosystems, several inventory methods, such as gillnet, trawler and hydro-acoustic surveys, are widely used to estimate those parameters. Despite their positive assets, these methods are not without constraints and limitations. For instance, they can be invasive, costly in terms of material and human resources, may cause unwanted mortality in communities studied as well as being subject to size and species selection bias. Others such as hydro-acoustic methods may leave uncertainties regarding species composition. Environmental DNA (eDNA) analysis, which consists of detecting DNA traces released by species in their environment, could be used as a non-invasive, less costly, and perhaps more accurate alternative or in complement to conventional bio-moniytoring methods. Yet, eDNA methods potentially come with their own caveats and therefore, there is a need for rigorous comparisons with conventional inventory methods. In this presentation, we evaluate the pros and cons between eDNA (both single species and eMetabarcoding approaches) and three conventional methods (gillnet, trawler, hydro-acoustic) for monitoring freshwater and marine fish communities in terms of species identification, richness, relative abundance and species abundance. The results will be discussed in the context of their applications for bio-monitoring in different contexts including environmental assessment, commercial fisheries and whale conservation.

W852: Population and Conservation Genomics 2
Evaluating Faunal Diversity with Environmental DNA
Kevin Leempoel[1], Trevor Hebert[2] and Elizabeth Hadly[1,2], (1)Stanford University, Stanford, CA, (2)Jasper Ridge Biological Preserve, Stanford University, Stanford, CA
Environmental DNA (eDNA) is a particularly promising approach for the rapid and sensitive monitoring of biodiversity. However, a number of obstacles must be addressed before it can be used for this purpose, especially for terrestrial mammals. In fact, the influence of environmental conditions, experimental design and species characteristics on the spread and persistence of DNA released by animals into the environment is poorly understood. In addition, comparisons between eDNA and current monitoring methods are non-existent.
To address some of these unknowns, we tested our ability to detect the presence of terrestrial mammals using eDNA analysis of soil samples against confirmed species observations from nearby camera traps. By doing this project in Jasper Ridge Biological Preserve (Stanford, CA), we took advantage of the camera traps installed there that have been monitoring wildlife for the last 9 years. At the same time, we compared several parameters, including 2 sampling designs, 2 DNA extraction kits and 2 metabarcodes of different length.
Our results show that eDNA analysis is not only effective at detecting large mammals but also species that are often too small to trigger cameras. We also report that a phosphate buffer-based extraction gave similar results to a commercial kit in terms of species detected but for a fraction of the price. Finally, the longest metabarcode (219bp) proved to be less sensitive than a shorter one (67bp), while offering a much higher taxonomic resolution. These results demonstrate that eDNA is relevant for biodiversity monitoring, but require many more validation studies.

W853: Population and Conservation Genomics 2
Population Genomics of Environmental Change across Multiple Scales in the American Pika
Michael Russello, University of British Columbia, Kelowna, BC, Canada
The ecological effects of climate change have been shown in most major taxonomic groups; however, the evolutionary consequences are less well-documented. Adaptation to new climatic conditions offers a potential long-term mechanism for species to maintain viability in rapidly changing environments, but mammalian examples remain scarce. The American pika (*Ochotona princeps*) has been impacted by recent climate-associated extirpations and range-wide reductions in population sizes, establishing it as a sentinel mammalian species for climate change. To investigate evidence for local adaptation and reconstruct patterns of genomic diversity and gene flow across rapidly changing environments, we used a space-for-time design and restriction site-associated DNA sequencing to genotype American pikas along two steep elevational gradients at 30,966 SNPs and employed independent outlier detection methods that scanned for genotype-environment associations. We identified 338 outlier SNPs detected by two separate analyses and/or replicated in both transects, several of which were annotated to genes involved in metabolic function and oxygen transport. Additionally, we found evidence of directional gene flow primarily downslope from high-elevation populations, along with reduced gene flow at outlier loci, suggesting elevational range contractions in American pikas will likely be from local extirpation rather than upward movement of low-elevation individuals. We are extending this design to sites across the North American range of *O. princeps*, with analyses to be aided by a new, high-quality genome assembly that is currently underway. Overall, these findings are of particular relevance for future conservation and management of American pikas and other elevationally-restricted, thermally-sensitive species.

W854: Population and Conservation Genomics 2
The Devils' Cancer: Genomics of Rapid Evolution and Adaptive Potential in the Face of Epidemic Disease
Paul Hohenlohe[1], Amanda Stahlke[1], Hamish McCallum[2], Menna Jones[3], Rodrigo Hamede[4] and Andrew Storfer[5], (1)University of Idaho, Moscow, ID, (2)Griffith University, Nathan, QLD, Australia, (3)University of Tasmania, Hobart, Australia, (4)University of Tasmania, Hobart, TAS, Australia, (5)Washington State University, Pullman, WA
Devil facial tumor disease (DFTD) is a transmissible cancer that threatens the persistence of Tasmanian devils (*Sarcophilus harrisii*). We have detected evidence for rapid evolution in response to DFTD as well as a genetic basis to disease-related phenotypes. Recently a second independently derived transmissible cancer was discovered in devils, raising the hypothesis that this is a recurring selective force, and that the devil lineage has faced multiple transmissible cancers in the past. Candidate loci responding to selection from the current DFTD show only a slightly greater rate of positive selection in the past compared to the rest of the genome, although historic selection patterns point to functional categories of genes, particularly involved with cancer and immune function, that may have responded to previous epidemics. These results hold

important implications for management of devil populations for the conservation of adaptive potential in the face of transmissible cancer and other threats.

W855: Population and Conservation Genomics 2

Local Genetic Adaptation in United States *Bos taurus* Beef Cattle

Troy N. Rowan[1], Robert D. Schnabel[2] and Jared E. Decker[1], (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Division of Animal Sciences, Informatics Institute, Columbia, MO

Animals that are poorly adapted to their local environments cost the beef industry more than a billion dollars each year. This project aims to categorize genomic variation that contributes to successful animals in particular environments, as well as to better understand gene-by-environment interactions in cattle. Using an evolutionary approach to study artificial selection, we create deliverable results for the beef industry while providing important insights to adaptation biology.

Genotypes were phased with Eagle v2.4 and imputed with Minimac3 up to ~850,000 SNPs for over 13,000 Gelbvieh, 12,000 Simmental, and 22,000 Red Angus cattle from 9 distinct regional environments of the United States. We performed FLK selection scans to identify variants under strong, region-specific selection. In addition, we used 30-year mean temperature, precipitation, and elevation measurements as phenotypes in univariate and multivariate genome-wide association analyses to identify loci directly associated with these environmental gradients, and presumably under selection. Birthdate selection mapping, in conjunction with environmental variables, was used to observe significant allele frequency shifts over time, and their relationship to climate variables. To control for within-family selection, we use the SmartPCA algorithm and a PC-based test for selection to identify loci responsible for familial, but not region-specific selection in these populations. Loci discovered in regional selection scans will be used to calcualte region-specific genomic predictions via genomic feature selection. These predictions will provide beef producers with a new breeding and selection tool to maximize efficiency specific to different environmental stressors throughout the United States.

W856: Population and Conservation Genomics 2

Your Daily GWATS: A Time-Series Aware GWAS to Detect Natural Climate Adaptations in North American *Populus trichocarpa*

Jared Streich[1], David Kainer[2], Jonathan Romero[3], Ashley Cliff[3] and Daniel Jacobson[4], (1)Oak Ridge National Labs, Oak Ridge, TN, (2)Oak Ridge National Laboratory, Canberra, TN, (3)University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, TN, (4)Oak Ridge National Laboratory, Oak Ridge, TN

We created a time-series aware association approach called GWATS (Genome Wide Association Time-series Studies) to detect climate adaptive alleles. GWATS involves running a GWAS analysis using time-series phenotype data, in this case for each day of the year. BioClim has been an ideal data series for species distribution modeling and finding adaptive alleles but lacks a true seasonal component and leaves most of each seasonal period unexamined. Instead, from raw monthly climate data, we interpolated 365 daily values of 26 climate/environment layers in the 265 locations where the 970 *Populus trichocarpa* individuals of our GWAS population were sourced. GWAS was performed for each climate variable on each of 365 days using ~10M genome-wide SNPs. Additionally, we ran GWATS on gene presence/absence using genes in 95% or less in all individuals, as well as a traditional population structure analysis using both SNPs and gene presence/absence to discover demographic processes underlying *P. trichocarpa*. Geographically isolated alleles tend to coincide with adaptive alleles making them difficult to distinguish from false positives. However, true positive p-values become distinct from false positives by using a Fourier Transform to analyze co-variate variation and output similarity matrix calculations across the time-series. Further, using a machine learning algorithm called iRF (iterative Random Forest) we can filter complex climate phenotypes into epistatic interactions across multiple suites of alleles ranging as high as five or more orders of epistasis. Using GWATS we detected hundreds of candidate climate adaptive loci for Solar Radiation Stress, Precipitation Stress, Temperature Stress, Drought Stress and many more.

W857: Population and Conservation Genomics 2

Genome Assembly and Population Genomic Analysis of *Vanessa tameamea*, the Island Endemic Kamehameha Butterfly

Sheina Sim[1], William P. Haines[2], Angela N. Kauwe[1], Julian R. Dupuis[3], Daniel Rubinoff[4] and Scott Geib[1], (1)USDA-ARS, Hilo, HI, (2)University of Hawaii at Manoa, Honolulu, HI, (3)University of Hawaii, Hilo, HI, (4)University of Hawaii, Honolulu, HI

The Kamehameha butterfly, *Vanessa tameamea*, a species endemic to the Hawaiian Islands, provides a unique study system for the study of allopatric divergence. Though it is not currently endangered, it is a threatened species with a population level that is on the decline likely due to habitat loss. To lay the foundation for a population genomic study in this species, we generated a high-quality chromosome-scale reference assembly. Using this reference, we performed a population genetic analysis using genome-wide SNP loci to assess the population structure of *V. tameamea* across its geographic distribution. Illumina sequencing and assembly of a 10x Chromium library and further assembly using HiC sequences resulted in a reference assembly with a total length of 324.7 Mb in 31 scaffolds with a scaffold N50 of 12.3 Mb. The HiC chromosome assembly was validated using linkage information obtained through analysis of genome-wide SNP genotypes of five mapping populations. A principal component analysis revealed a population genetic structure based on the island in which individual butterflies were collected with the distal islands Kauai and Hawaii distinct from the central islands Oahu, Molokai, and Maui Nui with Kauai and Hawaii butterflies harboring few but private or nearly private alleles. A structure analysis revealed that only Kauai was distinct from the rest of the islands, which indicates more gene-flow between the southernmost islands. This study represents a model for developing resources in species of conservation concern, and can be used to inform the practices employed to conserve this threatened island endemic.

W858: Potato Genomics

Dissecting Genes That Contribute to Yield in Potato

Shelley Jansky, USDA-ARS, Madison, WI, Alex Marand, University Of Georgia, Athens, GA and Jiming Jiang, Michigan State University, East Lansing, MI

Yield is the most important trait in plant breeding programs. However, although potato is a major food crop, breeders have not been able to identify genes associated with the components of yield – tuber number and tuber size. This is largely due to the genetic nature of the potato of worldwide commerce, which is a heterozygous autopolyploid. To overcome these limitations, potato breeders have been exploring the possibility of converting the crop into a diploid inbred line-based hybrid crop. Success will depend on overcoming two critical obstacles. First, inbred lines must be created in a crop with a gametophytic self-incompatibility system. Second, it will be necessary to demonstrate that high yields can be achieved at the diploid level. In this study, we crossed a heterozygous cultivated diploid with an inbred wild potato homozygous for a dominant self-incompatibility inhibitor. The segregating population, comprised of 90 clones, was phenotyped for vine and tuber traits. Tremendous yield variation was observed, with some clones producing yields similar to those of commercial tetraploid cultivars. Tuber number and tuber weight were highly heritable and independent of each other. Mapping based on 1.5 million short nucleotide variants revealed 14 quantitative trait loci associated with yield. Candidate genes were identified among the 28 tuber/stolon-specific genes underlying the QTL. This study is foundational for the development of F1 hybrid potato varieties.

W859: Potato Genomics

SNP Discovery Driven By Maximizing Allelic Richness from a Genetic Diversity Analysis of a European Potato Breeding Germplasm

Aurélie Evrard, Florimond-Desprez Veuve et Fils SAS, Cappelle en pévèle, France

Potato (*Solanum tuberosum* L.) is the fourth most important crop in the world. New potato breeding tools have recently become available such as genotyping arrays (SolCap 8303, SolSTW). Using re-sequencing of diploid, wild and tetraploid potatoes allowed detection of thousands of high-quality SNP to be printed on arrays. Most of the genotypes selected to build a microarray are chosen randomly or according to the standards of a market classes. Therefore, SNP from the SolCap 8303 identified from North American cultivars are predominantly targeting the processing industry whereas SNP from the discovery panel of Uitdewilligen *et al.,* 2013 are mostly from wild species progenitors carrying genomic introgressions. Although, these microarrays are of general use for research purposes they might not be of interest from a European breeder point of view as they might lack genetic diversity. Moreover, it was shown that some public SNP (called in tetraploid mode) showed no polymorphism or present low minor allele frequency in a given potato population. Here, we present a strategy where we selected a potato core collection to maximize genetic diversity in order to produce a microarray with high allelic diversity. Using the SolSTW array, we perform a genetic diversity analysis on 300 varieties from our germplasm. This analysis revealed that the distribution of the allelic variability regroups potatoes according to their market classes (as it was previously shown by Uitdewilligen et al., 2013). By developing our own core collection algorithm, we were able to select the best set of individuals that will maximize the allelic richness and therefore increase the number of SNP discovery. Re-sequencing and mapping of this core collection onto the reference genome indicate that we have called 20 times more SNP than in published SNP discovery pipelines. A comparison allowed us to determine that approximately a quarter of the public SNP were not included in our SNP pool.

W860: Potato Genomics

Potato Pangenome Consortium

Richard Finkers, Wageningen UR Plant Breeding, Wageningen, Netherlands

Cultivated potato is what we call a highly heterozygous, self-incompatible outbreeder – in practice that makes it impossible to produce true breeding lines and so genetic improvement is a complex and lengthy process. With the availability of the reference genome sequence of the doubled monoploid clone DM, potato genomic assisted breeding made a huge step forward. However, potato is an auto-tetraploid, each potato cell contains four nearly identical copies of each chromosome and during a cross, each parent contributes two copies to each offspring at random. With four nearly identical copies, the assembly and phasing of the four copies extremely difficult for traditional technologies.

This complexity also has impact on the implementation of marker assisted breeding in Potato. Though markers are available for a set of qualitative traits, markers for many of the more complex quantitative traits are lacking or show limited usability. Though it is know that copy number variation exists within one cultivar, we have only limited knowledge of the extend of this.

To further understand the structural (and functional) variation in tetraploid potato, we used the denovomagic 3.0 technology, to develop phased assemblies of, initially, two potato cultivars. Outcomes of this effort will be discussed. Furthermore, an International consortium with 12 Academic institutions and 2 industrial partners and NRGene has been established to further develop this effort in to developing phased assemblies of, in total, six potato cultivars. These six potato cultivars will be used as a basis for developing the tetraploid potato pan-genome.

W861: Potato Genomics

Allelic Diversity of the Complex Earliness Locus in European Tetraploid Germplasm

Christian Bachem, Laboratory of Plant Breeding, Wageningen University & Research, Wageningen, Netherlands, Lorena Ramirez-Gonzales, Plant Breeding, Wageningen UR, Wageningen, Netherlands, Richard Finkers, Wageningen UR Plant Breeding, Wageningen, Netherlands, Li Shi, Wageningen UR, Wageningen, Netherlands and Richard G.F. Visser, Wageningen University & Research, Plant Breeding, Wageningen, Netherlands

Potato tuber formation is regulated via a pathway that is very similar to floral induction in all investigated angiosperms systems. Potato tubers, in wild Andean potato, are formed under short day conditions. Adaption to long-day tuberisation prevalent in the growing seasons of northern latitudes has been facilitated by a mutation in a regulatory protein belonging to the Cycling DOF transcription factor family *StCDF1*. This mutation represents a transposon induced truncation of the StCDF1 protein that eliminates the FKF1 binding site, allowing the protein to evade diurnal proteolytic degradation. Natural allelic variation in the *StCDF1* locus, correlates with phenotypic variation of the earliness and plant maturity traits. We have studied this allele variation in a number of *de novo* assembled diploid and tetraploid sequenced genotypes and analysed segregation and dosage effects on the phenotypes related to earliness. Furthermore, closer analysis of this locus has revealed a second overlapping non-coding transcript transcribed in the opposite direction in some *StCDF1* alleles, similar to the one that has also been identified in *A. thaliana,* that we have named StFLORE. StFLORE is a long non-coding RNA that appears to be regulated by StCDF1 as well as by ABA. We will discuss the implications of the complex StCDF1 locus on traits associated with earliness and other important agronomic traits related to abiotic stress.

W862: **Potato Genomics**
Diploid Potato Genomics
Martina Stromvik, Department of Plant Science - McGill University, Ste. Anne de Bellevue, QC, Canada

W863: **Potato Genomics**
Frame Shift and Missense Mutations in SGT2 and Game 4 Lead to Low Glycoalkaloid Content and Altered Expression of Glycoalkaloid Biosynthesis Genes in Diploid Potatoes
Bourlaye Fofana[1], Ashok Somalraju[1], Jason McCallum[1], Kaushik Ghose[2] and David Main[1], (1)Agriculture and Agri-Food Canada, Charlottetown, PE, Canada, (2)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX
Potato is the third most consumed food crop by humans after rice and wheat. In the potato breeding continuum, breeders face the challenge of reducing anti-nutritional factors such as steroidal glycoalkaloids (SGAs) through recurrent crosses. Despite all efforts, and depending on variety and postharvest handling conditions, some released potato varieties still show high levels of SGA. We developed an EMS mutagenized diploid potato population and phenotyped 246 mutant lines, 21 wild types and 3 commercial varieties for the SGA trait. Sixteen percent of the mutant lines showed lower SGA content compared to the wild type potato lines and the commercial varieties. An ampliseq sequencing was conducted using a panel of 9 target genes including SGT1, SGT2, SGT3, Game 4a, Game 4b, Game 6, Game 7, Game 11, and Game12 to understand the mutational events underlying the low SGA phenotypes. An RNAseq transcriptomic analysis was further conducted in a mutant line showing a low SGA content and in its wild type counterpart to identify the SGA biosynthetic genes whose expression might be altered in mutant lines. Frame shift and missense mutations leading to loss-of-function were identified in SGT2 and Game 4 genes of the EMS-induced mutant lines and correlated alteration of SGA biosynthesis gene expression was observed in the mutant potato lineThe data will be presented and further discussed in the context of generating a large potato pre-breeding germplasm for improving quality and agronomic traits in potato.

W864: **Poultry 1**
Genetics of Response to Heat Stress in Mature Laying Hens
S.J. Lamont, Department of Animal Science, Iowa State University, Ames, IA
High ambient temperature can have severe negative consequences on production in laying hens. A genome-wide association study was conducted using 600k SNP genotype data and production traits including feed intake, body weight, digestibility, egg quality, and blood components of commercial white egg-laying hens before and during a 4-week heat exposure. Several phenotypes at various times had heritability estimates greater than zero. Quantitative trait loci were identified for many phenotypes and were generally different for the same trait measured before and after heat exposure. Estimated effects were low, suggesting highly polygenic control of physiological response to heat stress in mature laying hens. The measurable heritabilities, however, indicate the existence of genetic control and the feasibility of changing layer production traits under heat stress through genetic selection using genomic markers or physiologic biomarkers.

W865: **Poultry 1**
Effect of Heat Stress on the Development of *Eimeria* spp in Broiler Chickens
Gustavo H Schneiders[1], Alberta L Fuller[2], Romdhane Rekaya[2] and **Samuel E. Aggrey**[2], (1)Unversity of Georgia, Athens, GA, (2)University of Georgia, Athens, GA
Both heat stress (HS) and *Eimeria* parasite infection are stressors that cause significant economic losses in broiler chickens. These stressors compromise intestinal integrity and independently or interactively predispose broiler chickens to immunosuppression rendering them susceptible to other enteric diseases. We conducted an experiment to determine the consequence of HS and *Eimeria* infection in broiler chickens. Eimeria oocysts were detected in the excreta of chickens infected and raised in thermoneutral environment (25^{o}C), however, no oocysts were detected in infected birds raised under heat stress (35^{o}C) suggesting a potential inhibition of oocysts development in *Eimeria* infected birds under chronic HS. We therefore investigated the effect of temperature on viability of *E. tenella in-vitro* to ascertain the relationship between temperature and development and viability. Viability and morphology were assessed by imaging flow cytometry with 10,000 events acquired per sample. Sporozoite count was accessed by optical microscopy. Incubation of Eimeria at 50˚C induced a significant drop in sporozoite viability marked by increase in propidium iodide and decrease in fluorescein diacetate fluorescence emissions. Incubating at 50^{o}C for 60 min reduced viability to 81%, however, at 120 min post-incubation, viability was only 3%. Additionally, incubation at 50˚C affected sporozoite counts and shape. Sporozoites extracellularly treated at 55˚C show inability to invade MDBK cells. Intracellular development of merozoites was significantly reduced by a slight increase in 2˚C in the optimal temperature of incubation, suggesting that an increase in temperature alone could reduce the infection rate *in vivo* and have a significant effect in the outcome of *E. tenella* infection in heat-stressed chickens. In another *in-vivo* experiment where live birds were exposed to HS, about 80% of the birds did not support the development of merozoites in the caeca, while the remaining 20% had about 110-fold reduction in merozoites. Since HS reduces oocyst counts in *Eimeria* infected chickens, we can therefore conclude that, under HS, the viability of *E. tenella* sporozoites is significantly reduced thereby affecting its ability to recycle.

W866: **Poultry 1**
Improving Food Security in Africa by Enhancing Resistance to Newcastle Disease and Heat Stress in Chickens (Genomics to Improve Poultry Innovation Lab)
Huaijun Zhou, Animal Science, University of California, Davis, CA

W867: **Poultry 1**
Change in Gene Expression of Major Receptors Involved in the Neuroendocrine Regulation of Stress in Poultry
Wayne Kuenzel, Hakeem Kadhim and Seong W. Kang, University of Arkansas, Fayetteville, AR
The nucleus of the hippocampal commissure (NHpC), containing corticotropin-releasing hormone (CRH) neurons, has been proposed to be part of the classical hypothalamic pituitary adrenal (HPA) axis of poultry. We tested broilers exposed to food deprivation (FD) and determined gene

expression of CRH neurons, its major receptors, CRH-R1 and CRH-R2, in the NHpC compared to the main hypothalamic structure, the paraventricular nucleus (PVN), known to regulate the stress response of the HPA axis in most vertebrates. Control birds fed ad libitum and broilers subjected to FD had blood, brains and anterior pituitaries sampled (n = 12/group) at 0, 1, 2, 3, 4, and 8h of FD. Plasma corticosterone (CORT) was determined by radioimmunoassay. Brain samples were cut as cross-sections using a cryostat. The NHpC and PVN were micro-dissected from each cross-section so that the entire extent of each structure was sampled. RNA extractions of brain structures and anterior pituitary glands were followed by real-time, quantitative PCR to determine mRNA. Results showed that CRH mRNA in the NHpC peaked before PVN mRNA. CRH and its major receptor CRH-R1 mRNA in the NHpC were negatively regulated over the 8h sampling period resulting in CRH gene expression being significantly increased from controls only from 1 to 4h of FD. In contrast, CRH mRNA and its receptor CRH-R1 mRNA in the PVN were positively regulated throughout, resulting in both CRH and CRH-R1 showing significantly elevated mRNA throughout the entire FD period compared to controls. CRH-R2 showed a different pattern of expression. In the NHpC and PVN, CRH-R2 showed no significant change from controls at 1h of FD. Thereafter, CRH-R2 mRNA showed significant increases in both the NHpC and PVN from 2 to 8h of FD compared to controls. Overall current data suggest that CRHmRNA in the NHpC may prime early activation of the HPA axis in birds and shortly thereafter gene expression of the CRH neurons return to unstressed control levels. The steady increase in CRH mRNA in the PVN and its positive feedback with its CRH-R1gene expression sustained elevated CRH mRNA and ultimately significantly elevated CORT levels thereby maintaining the FD stress response. Supported by a grant from the Arkansas Biosciences Institute, funding from the Division of Arkansas and an HCED fellowship from Iraq to HK.

W868: **Poultry 1**
The Search for the Missing Sequence of the Chicken Genome
Wesley Warren, University of Missouri, Columbia, MO

W869: **Poultry 1**
Quantitative Genetic Analysis of Reproductive Traits on a Male Line of Turkey
Guilherme J. M. Rosa, University of Wisconsin-Madison, Madison, WI

W870: **Poultry 1**
DNA Sequence and Haplotype Variation Analyses of Functionally Important Turkey Genes and Update of VT's NIH-Funded Research Education Efforts
Edward Smith, Virginia Tech, Blacksburg, VA

W871: **Poultry 1**
Molecular and Immunogenetics of Marek's Disease Resistance
Hans Cheng, USDA, ARS, ADOL, East Lansing, MI

W872: **Poultry 1**
Directed Genome Evolution to understand Phagocyte Survival in a BCO Hypervirulent Isolate
Abdulkarim Shwani[1], Sura Zaki[1], Sohita Ojha[2,3] and **Douglas D. Rhoads**[4], (1)University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR, (2)Langston University, Langston, OK, (3)University of Arkansas, Fayetteville, AR, (4)University of Arkansas, Cell and Molecular Biology Program, Fayetteville, AR
The aim of our study was to identify bacterial genes responsible for survival in and killing of chicken macrophage. We have been investigating the pathogenicity of an isolate of *Staphylococcus agnetis* obtained from lame birds on our research farm. The isolate was from femoral head necrosis samples from broilers induced for lameness by growth on suspended wire flooring. All evidence supports the development of a hypervirulent strain through repeated experiments inducing high incidence of lameness; primarily Bacterial Chondronecrosis with Osteomyelitis (BCO). Our isolate 908 is capable of inducing 50% BCO lameness by 56 days in birds on litter, when administered as a single dose in drinking water at 20 days of age. The BCO infection can be spread through the air to other birds in the same facility. We compared our chicken isolate 908 to *S. agnetis* isolates from mastitis in dairy cattle for bacterial killing assays in an immortalized chicken macrophage line. The cattle isolates are rapidly killed but isolate 908 not only survives after phagocytosis, it kills the macrophage within 2 days; even at low multiplicity of infection (MOI). We have sequenced and assembled genomes for 9 cattle isolates and isolate 908. Phylogenetic trees demonstrate that isolate 908 clusters within the cattle isolates. We transformed DNA from isolate 908 into a closely related cattle isolate (1379) and passaged the transformed library through chicken macrophage to select for survival and macrophage-killing. Repeated transformations have demonstrated that the survival and killing activity appears to reside on one or more episomes in isolate 908. Further work will be focused on the specific episome and gene(s) responsible for this virulence determinant. Our model is that isolate 908 is hypervirulent in chickens because it has obtained determinants that allow it to survive phagocytosis and colonize weak areas in the chicken vascular system, specifically the rapidly growing proximal growth-plates of the long leg bones, leading to BCO lameness. Defining these virulence systems will help us design management and selection strategies to reduce BCO lameness in broilers.

W873: **Poultry 1**
Improving Bone Health and Protection against Bacterial Chondronecrosis with Osteomyelitis using Micronutrients and Genetic Selection
Adnan Alrubaye, University of Arkansas/Fayetteville, Fayetteville, AR

W874: **Poultry 1**
Mitochondria Changes and MDV Infection in Chicken

Jiuzhou Song, University of Maryland, College Park, MD

W875: **Poultry 1**
Epigenetic Mapping of the Chicken Genome: Implications for Disease Resistance and Production
Yvonne Drechsler, Western University of Health Sciences, Pomona, CA

W876: **Poultry 1**
Expression of Host Defense Peptides in the Chicken Intestine and Yolk Sac
Eric A. Wong[1], Haihan Zhang[1], Shengchen Su[1], Kate B. Miska[2], Raymond Fetterer[2], Mark C. Jenkins[2] and S.J. Lamont[3], (1)Virginia Tech, Blacksburg, VA, (2)Beltsville Agricultural Research Center, Beltsville, MD, (3)Department of Animal Science, Iowa State University, Ames, IA
Host defense peptides are part of the innate immune system and include the avian beta defensins (AvBD), cathelicidins and liver expressed antimicrobial peptide 2 (LEAP2). The yolk sac and intestine provide a first line of defense against pathogens present in yolk or feed, respectively. Objective 1 profiled the developmental expression of AvBD mRNA in the yolk sac. Expression of AvBD1, 2, 7, and 10 mRNA was low at embryonic day 7 (e7), increased to e9 through e13 and then declined to e19. Using in situ hybridization, AvBD10 mRNA was found to be expressed in endodermal epithelial cells, while AvBD1, 2, and 7 mRNA were expressed in heterophils. Objective 2 compared the expression profile of AvBD mRNA in the intestine of Ross broilers and two lines of the disease-resistant Fayoumi chickens (M5.1 and M15.2) in response to an *Eimeria maxima* challenge. *E. maxima* challenge caused a downregulation of AvBD1, 6, 10, 12, and 13 mRNA in the jejunum of the two Fayoumi lines, but no change in the Ross broilers compared to non-challenged chickens. In the duodenum, there was upregulation of AvBD10 mRNA in the Ross and Fayoumi line M5.1 and upregulation of AvBD10, 11, 12, and 13 mRNA in Fayoumi line M15.2. LEAP2 mRNA was downregulated in the duodenum and jejunum of Ross and Fayoumi line M5.1, but not in Fayoumi line M15.2. In summary, AvBD mRNA showed development- and cell-specific expression in the yolk sac and tissue- and line-specific expression in the intestine in response to an *Eimeria* challenge.

W877: **Poultry 1**
HPIdb 3.0 : A Database for Host-Pathogen Interactions
Bindu Nanduri, Mississippi State University, Mississippi State, MS

W878: **Poultry 1**
Progress Toward Revealing MHC-Y Diversity and Function in Chickens
Marcia M. Miller, Beckman Research Institute, City of Hope, Duarte, CA

W879: **Poultry 1**
Fighting Fat with Fat: Fatty Acids and Adipose Deposition in Broiler Chicks
Brynn Voy, University of Tennessee, Knoxville, TN

W880: **Poultry 1**
Metabolic Risk Factors of Wooden Breast Disease in Commercial Broiler Chickens
Behnam Abasht, University of Delaware, Newark, DE

W881: **Poultry 1**
The Turkey Selenoproteome: Genes and Regulation of Transcript Expression by Selenium Deficiency and High Se Status
Roger A Sunde, University of Wisconsin, Madison, WI and Rachel M Taylor, University of Wisconsin-Madison, Madison, WI
Selenium is an essential trace element but is also toxic at high concentrations. Metabolism of inorganic Se into selenoproteins requires 6 unique gene products, including a selenocysteine tRNA, and each eukaryotic selenoprotein transcript requires an in-frame UGA plus a 3'UTR stem-loop called a Selenocysteine Insertion Sequence. Because of this selenium-specific signature, the complete selenoproteome of an organism can be determined: humans and pigs have 25, rodents 24, and chickens and turkeys have 24 selenoproteins. Relative to humans, avians lack 3 selenoprotein genes but have 2 unique selenoproteins.
Specifically, the current turkey Annotation 102 correctly annotates and fully-assembles 15 selenogenes. Annotations of 6 selenogenes are missing portions of the 5' UTR and/or N-terminal sequences, two (SELENOP1, TXNRD1) are miss-assembled and incomplete, and one selenogene (SELENOW) is not annotated and only present in unplaced scaffold sequences.
To study Se regulation of selenoenzyme activity and transcripts in turkey poults, we fed graded Se levels from Se-deficient to high–Se (0 – 5 μg Se/g as selenite) in vitamin E-adequate diets for 28 days. Relative to rodents, turkeys have $1/10^{th}$ the level of GPX1 and 6X the level of GPX4 in liver, and the secreted selenoprotein transcripts (GPX3, SELENOP1) are well-expressed in multiple tissues. Poults supplemented with <0.05 μg/g had reduced growth, but with no effect of high Se. Se biomarkers responded hyperbolically to increasing dietary Se and reached plateaus at or before 0.4 μg/g. Liver and kidney Se, plasma GPX3 activity, and activities of tissue GPX1 and GPX4 decreased to <10% in Se deficiency and reached plateau levels by 0.4 μg/g. In the same tissues, ≤6 out of 24 selenoprotein transcripts were downregulated to <50% by Se deficiency and no transcript was altered by high Se. Except for liver Se, none of these Se status biomarkers increased >2X Se-adequate levels in poults fed up to 5 μg/g diet, whereas liver Se increased to 5.6X. Thus the dietary Se requirement of growing turkey poults is 0.4 μg Se/g diet, double the current NRC requirement; turkeys appear resistant to excess dietary Se, suggesting that FDA Se supplementation limits can be safely raised. (USDA Hatch 1013496)

W882: **Poultry 1**

Epigenetic Effects of a Diet High in Methyl Donors across Multiple Generations of Japanese Quail
Christopher M. Ashwell, North Carolina State University, Raleigh, NC

W883: **Poultry 1**
Omic Analysis of Chicken Liver Post-Hatch Development
Carl J. Schmidt, Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

W884: **Poultry 2**
NRSP8 Bioinformatics Update
James M. Reecy, Iowa State University, Ames, IA

W885: **Poultry 2**
Molecular Understanding of Different Forms of Dwarfism in Chicken
Richard P.M.A Crooijmans, Wageningen University & Research, Wageningen, Netherlands

W886: **Poultry 2**
Lessons Learned from a Naturally Evolved System to Enrich a Functional Microbiome: The Wonders of Mammalian Milk
David Mills, Food Science & Technology, University of California, Davis, CA
Human milk contains numerous components that shape the microbial content of the developing infant gastrointestinal tract. A prominent feature of milk is an array of oligosaccharides and glycoconjugates that serve a passive immune function by sequestering and deflecting pathogens while simultaneously enriching a protective often dominated by bifidobacteria. Recent research suggests the timing of establishment, and proper function of, the neonate gut microbiota is critical for infant development. This community is initially established through environmental transfer to the gut and subsequently shaped by diet (milk) and host genetics. Once established, infant gut communities dominated by bifidobacteria exhibit low residual milk glycans and higher levels of short chain fatty acids in the feces, suggesting a strongly saccharolytic activity. The mechanistic basis for milk glycan consumption by bifidobacteria has been the subject of active research. Different infant-borne bifidobacteria contain specific glycosidases and transport systems required to utilize milk oligosaccharides and glycoconjugates. In aggregate, these studies suggest a co-evolutionary relationship between mammalian milk glycans, infant-borne bifidobacteria and the infant host resulting in a programmed enrichment of a protective bifidobacterial-dominant community during a critical stage of infant development. Disruption of this programmed enrichment, by poor environmental transfer, antibiotic use, or infection, can lead to a “poorly functioning” microbiota that may pose a risk for negative health outcomes. Analysis of this naturally evolved system provides insight on effective pre- and probiotic tools that support and ensure a protective gut microbiota for any animal.

W887: **Poultry 2**
Porcine Microbiome
Glenn Zhang, Oklahoma State University, Stillwater, OK

W888: **Poultry 2**
Poultry Microbiome
Timothy J Johnson, University of Minnesota, Saint Paul, MN

W889: **Poultry 2**
Neal A. Jorgenson Travel Award Winner
Ganrea Chanthavixay, Animal Science, University of California, Davis, CA

W890: **Poultry 2**
Neal A. Jorgenson Travel Award Winner
Lei Liu, University of Minnesota, St. Paul, MN

W891: **Pragmatic solutions for scaling your analysis: Machine Learning, Imaging, Containers, Clouds and APIs**
Discussion Leader
Nirav Merchant, CyVerse, Tucson, AZ

W892: **Proteomics**
Decoding the Dynamic Proteome of Plant Peroxisomes
Jianping Hu, Plant Research Laboratory and Department of Plant Biology, Michigan State University, East Lansing, MI
Peroxisomes are morphologically and metabolically dynamic organelles that house a suite of biochemical pathways, such as the degradation of hydrogen peroxide, catabolism of fatty acids and derivatives, photorespiration, and the biosynthesis of plant hormones. Plant peroxisomes had been classified into three major subtypes: leaf peroxisomes, glyoxysomes in seeds and germinating seedlings of oilseeds, and gerontosomes in senescent tissues. To catalogue the dynamic peroxisome proteome and fully understand the importance of these organelles in plants, we performed mass spectrometry (MS)-based proteome analysis of peroxisomes isolated from green leaves, etiolated seedlings and dark-treated senescent tissue of Arabidopsis, followed by fluorescence microscopy-based protein targeting verification and mutant analysis. We discovered dozens of novel peroxisomal proteins that revealed new aspects of plant peroxisome function and metabolism. The core proteome is conserved

between peroxisomal subtypes, suggesting that these variants can all be simply called “plant peroxisomes” to avoid confusion. In silico analysis of the rice genome suggested strong similarities of the peroxisome proteome between diverse plant lineages, but it is highly likely that significant functional differences exist in peroxisomes between oil seeds such as like Arabidopsis and starchy seeds such as rice. The plant peroxisome proteome appears to be more complex than those from other eukaryotes, and many proteins and biochemical pathways are plant specific. Our work may provide a knowledge base for improving the production, quality and stress tolerance of crops.

W893: Proteomics
Proteome Rebalancing in Camelina Seeds
Monica Schmidt, University of Arizona, Tucson, AZ
Oilseed crops are global commodities for their desired oil and protein seed content. Increasing either of these seed constituents without altering the other has been a desired goal. We have engineered the oilseed *Camelina sativa* to exhibit increased protein content with only a slight decrease in oil content. The introduced seed-specific expression cassettes consisted of a bacterial codon-optimized chloroplast targeted phytoene synthase gene with/without an RNAi cassette directed to suppress the storage protein 2S albumin. The introduced phytoene synthase cassette produces enhanced b-carotene content of an average 275 +/- 6.10 mg/mg dry seed and an elevated overall average 44 +/- 1.02 % protein/ 25 +/- 0.24 % oil seed composition. Stacking an RNAi to suppress the major 2S storage protein resulted in seeds that contain an average 54 +/- 1.67 % protein / 25 +/- 0.70 % oil compared to control seeds with 35 +/- 0.87 % protein/ 26 +/- 0.45 % oil seed composition showing that *Camelina* rebalances its proteome within an enlarged protein content genotype. Additionally, the b-carotene trait, independent of the RNAi2S trait, resulted in a reduction in seed carbohydrates and an overall enlargement of seed size. The enhanced b-carotene trait functions as an orange visual selectable marker that is readily distinguished from nontransgenic seeds. The use of GRAS (generally regarded as safe) b-carotene as a visual marker in a floral dip transformation system, such as *Camelina*, eliminates the need for costly regulatory and controversial antibiotic resistance markers. b-carotene enhanced RNAi2S suppressed *Camelina* seeds can be further developed as a rapid heterologous protein production platform in a nonfood crop leveraging its enlarged protein content and visual marker.

W894: Proteomics
A Proteomics-Based Approach to Analyze the Localization, Composition, and Dynamics of Endogenous Protein Complexes
Dan Szymanski[1], Youngwoo Lee[1], Zach McBride[1], Thomas W. Okita[2], Donglai Chen[1] and Jun Xie[1], (1)Purdue University, W. Lafayette, IN, (2)Washington State University, Pullman, WA
Protein complexes assemble in cells to carry out activities that could never be achieved by a single polypeptide. Oligomeric protein machines coordinate information flow, control metabolic flux, and direct intracellular transport and cell wall secretion. Furthermore, the activities of protein complexes change to enable the plant to respond to developmental or environmental cues. However, despite the importance and ubiquity of protein oligomerization in biological systems, broad knowledge about protein complex composition and regulation in general is lacking. There is a strong need to develop new technologies that enable systems-level analyses of protein complexes. Here we describe the use of label-free protein correlation profiling of cell fractions that are separated using multiple orthogonal chromatographic separations. The profile data are being used to broadly analyze the abundance, localization, and oligomerization states of tens of thousands of proteins. The method is robust, and has been used to analyze soluble and membrane-associated protein complexes in leaves (Arabidopsis and soybean), developing rice aleurone, and purified cotton fibers. One goal of this project is to discover how cellular pathways and protein complexes respond to metabolic stress, and this method identified a small number of protein complexes that dramatically rearrange in response to metabolic stress. The method has also been used to predict protein complex composition in Arabidopsis leaves and the rice aleurone based on the principle of co-elution of stable protein complex subunits. These protein complex composition predictions and their validation using known complexes, coIP, and proteomic profiling of a predicted complex subunit will be described.

W895: Proteomics
Efficacy of Phytochemicals on *Campylobacter jejuni* Biofilms on Common Food Processing Surfaces and their Effect on Proteome of *C. jejuni*
Basanta R. Wagle[1], Abhinav Upadhyay[1], Indu Upadhyaya[2], Komala Arsi[1], Sandip Shrestha[1], Rohana Liyanage[1], Dan J. Donoghue[1] and **Ann M. Donoghue**[3], (1)University of Arkansas, Fayetteville, AR, (2)Tennessee Tech University, Cookville, TN, (3)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^{o}C and 37^{o}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.
Funded by USDA-NIFA-OREI-2017-51300-26815.

W896: **Proteomics**

Mitochondrial Proteomics and Transcriptional Analysis of Cytoplasmic Male Sterility in Sugar Beet using iTRAQ and qRT-PCR

Tianjiao Liu, Harbin Institute of Technology, Harbin, China

W897: **Proteomics**

Proteome Changes during Kernel Development in Wheat

Sintayehu D. Daba, Purdue University, West Lafayette, IN

Background: Grain yield is the top priority in any cereal breeding program. The final yield is the result of the combined action of genes operating at the different plant growth period, particularly tillering, spike formation, and kernel development periods. Proteome maps during the crop growth stage provide a powerful tool to investigate the mechanism yield development. This study characterizes the dynamics of protein abundance in the samples taken at seven stages.

Methodology: We applied label-free proteomic analysis on kernel sampled at three development stages (5DAA, 10DAA, and 15DAA) in comparison with ovary sample. The kernel development was also quantified in five development stages (5DAA, 10DAA, 15DAA, 20DAA, and 25DAA).

Results: Fresh and dry weights increased consistently. The moisture content drops sharply, particularly after the 10th day after anthesis. The first 10 days after anthesis was a period of increase in grain size and then period after 10DAA marked accumulation of photoassimilates such as carbohydrate and protein. The results from 3D-plot of principal components indicated the validity of the quantification procedure. Our comparison of the protein profiles of kernel development indicated several proteins showing significantly high or low expression in the three kernel development periods. Particularly, 76% of the proteins highly expressed at 15DAA were related to storage proteins.

Conclusion: The results from this study indicated that process of kernel development in wheat involves an initial kernel size increase in the first 10DAA, which is followed by extensive accumulation of photoassimilates displacing moisture.

W898: **QTL Cloning**

Deep Learning of Genomic Variation and Regulatory Network

Hai Wang, USDA ARS, Ithaca, NY

W899: **QTL Cloning**

Analysis of the Recombination Landscape of Hexaploid Bread Wheat Reveals a QTL and Genes Controlling Recombination and Gene Conversion Frequency

Anthony Hall[1], Ryan Joynson[1], James D. Higgins[2], Luzie U. Wingen[3], Bernardo J. Clavijo[1], Paul C. Bailey[1], Jonathan Wright[1], Neil Hall[1], Simon Griffiths[3], Thomas Brabbs[1] and Laura-Jayne Gardiner[1], (1)Earlham Institute, Norwich, United Kingdom, (2)University of Leicester, Leicester, United Kingdom, (3)John Innes Centre, Norwich, United Kingdom

Sequence exchange between homologous chromosomes through crossing over and gene conversion is highly conserved among eukaryotes, contributing to genome stability and genetic diversity. Lack of recombination limits breeding efforts in crops, therefore increasing recombination rates can reduce linkage-drag and generate new genetic combinations. We use computational analysis of 13 recombinant inbred mapping populations to assess crossover and gene conversion frequency in the hexaploid genome of wheat (Triticum aestivum). We observe that high frequency crossover sites are shared between populations and that closely related parental founders lead to populations with more similar crossover patterns. We demonstrate that gene conversion is more prevalent and covers more of the genome in wheat than in other plants, making it a critical process in the generation of new haplotypes, particularly in centromeric regions where recombination is sparse. We identify QTL for altered gene conversion and crossover frequency and confirm functionality for a novel RecQ helicase gene that belongs to an ancient clade that is missing in some plant lineages including Arabidopsis. This is the first gene to be demonstrated to be involved in gene conversion in plants. Harnessing the RecQ helicase has the potential to break linkage-drag utilizing widespread gene conversions.

W900: QTL Cloning

An Ethylene-Gibberellin Relay Co-Opts the Green Revolution Gene to Allow Rice Adaptation to Submergence

Takeshi Kuroha, Tohoku University, Sendai, Japan, Susan McCouch, Cornell University, Ithaca, NY and Motoyuki Ashikari, Nagoya University, Nagoya, Japan

As sessile organisms, plants must have an arsenal of response mechanisms to survive the various environmental stresses they encounter. A group of rice varieties, known as deepwater rice, has specifically adapted to survive periodical deep floods. In response to rising water levels, deepwater rices elongate their internodes to keep above the water surface and avoid anoxia. Here, we identify the gibberellin biosynthesis gene, *SD1*(also known as the Green Revolution gene), as a key gene responsible for submergence-inducible internode elongation using GWAS and genetic linkage analysis. The deepwater rice-specific haplotype (DWH) contributes to transactivation of *SD1* by direct binding of OsEIL1a, an ethylene-mediated transcription factor, under submergence. This transactivation is accomplished by direct binding of OsEIL1a, ethylene-mediated transcription factor, to a *cis*-regulatory element on the promoter region.The SD1 protein in deepwater rice contributes to the increase of active gibberellins, preferentially GA_4, which drastically promotes internode elongation. We further demonstrate that the DWH was derived from standing variation in wild rice and selected for deepwater rice cultivation in Bangladesh. Our findings elucidate the molecular mechanism of the deepwater response and the evolutionary history of deepwater rice, and shed light on intrinsic complexity and molecular plasticity of plant adaptation strategies.

W901: QTL Cloning

Seven in Absentia Gene underlies Increases in Biomass and Yield in Wheat in Hot Climates

Delphine Fleury, University of Adelaide, Glen Osmond SA, Australia

W902: QTL Cloning

Towards Cloning and Application of QTLs Associated with Steeper and Longer Root Systems to Develop Climate-Resilient Rice

Yusaku Uga, Institute of Crop Science, NARO, Tsukuba, Japan

Climate changes have caused droughts and soil degradation worldwide in recent time. To produce crops that can be sustainably grown under these difficult conditions, it is imperative to develop crop plants that are resilient to various environmental stressors. Genetic adaptations of the root system architecture are necessary to facilitate growth in water- and nutrient-deficient soils. Our research group identified and characterized several quantitative trait loci (QTLs) associated with steeper and longer root systems in rice. Recently, we performed a fine mapping of two QTLs for root length (*QRO1* and *QRO2*), which help to expand rootzone for increased active water and nutrient uptake. The use of root-related QTLs for marker-assisted breeding selection based on an ideal root system may allow us to develop climate-resilient rice. For this, we introgressed seven root-related QTLs from Kinandang Patong, an upland rice, into Colombian lowland varieties. As a preliminary result, several of the modified lines produced higher yields than parental varieties under low-input conditions as well as in normal paddy soils, suggesting improved resource use efficiency. To easily identify QTLs for 3-dimensionally-developed roots underground, furthermore, we are developing a high-throughput 3D non-destructive root phenotyping platform.

W903: QTL Cloning

Unravelling the Physiological Basis of Drought Tolerance Regulated by "QTL-Hotspot" Region in Chickpea (*Cicer arietinum* L.).

Rutwik Barmukh, ICRISAT, Telangana, India

W904: Quinoa and close relatives

Associating Quinoa Phenotypes and Genes to Reduced Yield Under Heat

Jose Tovar[1], Elizabeth Castillo[1] and **Malia Gehan**[2], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)Donald Danforth Plant Sciences Center, St. Louis, MO

Global temperatures are predicted to rise 1.5 to 5°C by the end of the century, making crop adaptation to heat an increasing agricultural priority. Quinoa (*Chenopodium quinoa*) produces highly nutritious grain under a range of mid to low temperatures, but is not adapted to heat, which limits its cultivation. The purpose of this study is to identify phenotypic changes and genes associated with quinoa seed yield under heat. To identify whether roots or shoots of quinoa drive yield-related phenotypic changes, the quinoa accession PI 614886 was submitted to heat in roots, shoots or both roots and shoots. Seed yield decreased by ~80% in both treatments with heated shoots, while no statistically significant reduction in yield was observed in the treatment with only roots heated. To assess the phenotype of seed produced from heat-treated plants, a low-cost high throughput phenotyping system was used. Seed phenotypic analysis revealed yield loss is due to significant reductions in seed size and in the number of seeds produced per plant. To investigate the causes of reduced seed number and size, quinoa inflorescence and flower phenotypes were studied. Heat-treated quinoa inflorescences have delayed development and bear significantly fewer fruit than control plants. Flowers fail to open in shoot heated plants, which potentially limits pollination. Preliminary analysis of pollen viability did not show significant differences between control and heat-treated plants. RNA-seq analysis of gene expression identified 394 genes differentially expressed in treatments with reduced yield, including transcription factors previously associated to flower closing and flowering time in *Arabidopsis*. Ongoing studies focus on assessing the effect on yield of inflorescence and flower development under heat using visible and hyperspectral imaging.

W905: Quinoa and close relatives

Flowering Time Regulation in Quinoa and Related Species of the Amaranthaceae Family

Christian Jung[1], Dilan Sarange[2] and Nazgol Emrani[1], (1)Christian Albrechts University of Kiel, Kiel, Germany, (2)Plant Breeding Institute, Christian-Albrechts-University of Kiel, Kiel, Germany

Quinoa (*Chenopodium quinoa* Willd) belongs to the family Amaranthaceae native to the Andean region of South America. Thus, a phenological development as a short-day plant is expected. The primary requirement for quinoa cultivation in Northern Europe is the day-length adaptation through modification of flowering time. Therefore, our study aims to understand the flowering time regulation in quinoa by harnessing the knowledge gained from its close relatives, sugar beet (*Beta vulgaris*) and *Chenopodium album*. We are performing field trials at the University of Kiel (Northern Germany) with 334 quinoa accessions from different geographical regions to measure agronomically important characters for a Genome-Wide Association Study (GWAS) in collaboration with the King Abdullah University of Science and Technology (Prof. Mark Tester). We observed a broad variation not only in flowering time but also in other morphological and phenological traits, which provides a promising potential for quinoa breeding programs in Germany. We will combine GWAS, QTL mapping and expression data, to identify major flowering time genes in quinoa. Based on the field data, we selected 4 early- and late-flowering accessions for a growth experiment under controlled conditions in the climate chamber. We are studying the expression pattern of candidate genes selected by their sequence homology with sugar beet flowering time genes *BTC1*, *BvBBX19, BvFT1,* and *BvFT2* under different photoperiod regimes. We observed different photoperiod responses among quinoa accessions. Under short-day conditions (8h light), two accessions showed accelerated flowering response while under long-day conditions (16h light) flowering was delayed in both accessions. We also identified two accessions with day-neutral flowering behavior. As a next step, we will compare the expression patterns of candidate genes between photoperiod-sensitive and –insensitive accessions. Moreover, we will use these genes for a candidate gene association mapping with the Quinoa core collection to link sequence variation with phenological development.

W9**06: Quinoa and close relatives**

Breeding Quinoa for Nutrition, Flavor and Functionality

Kevin Murphy, Washington State University, Pullman, WA

W9**07: Quinoa and close relatives**

Genome-Wide Association Study for Agronomic Quality Traits in *C. quinoa*

Hayley Hansen, Brigham Young University, Provo, UT

Genome-wide association studies (GWAS) incorporate genome-wide variant data with phenotypic data to identify genetic variants associated with traits of interest. A total of 479 accessions of *C. quinoa* genotyped using tGBS®-generated reads (Freedom Markers) were aligned to a

chromosome-scale reference assembly of *C. quinoa*. SNPs were called using reads that aligned to a single location in the reference genome and were filtered to include only those sites at which at least 50% of the samples were genotyped, resulting in a set of 198,288 high quality SNPs. Each SNP was supported on average by 31 reads/SNP/genotyped sample. Data was collected by Dr. Luz Gomez (Universidad Nacional Agraria La Molina, Lima, Peru) measuring 24 phenotypic traits for each of the accessions sequenced. Traits include seed weight, percentage of protein, saponin content, and percentage of leaves affected by downy mildew – a major disease affecting yield within the Andean region. We report preliminary association results using the rrBLUP package in R, utilizing a mixed linear model that incorporates the first 3 components from principle component analysis and a kinship matrix.

W908: **Quinoa and close relatives**
The Development of Genetic and Genomic Resources for *Atriplex hortensis*
Spencer Philip Hunt, Brigham Young University, Provo, UT
Atriplex hortensis (orach) is a highly nutritious leafy plant that is ubiquitous throughout the world. Unfortunately, a current sparsity in genetic and genomic resources has made it difficult to fully understand its diversity, taxonomic relationships and future potential. Here, I report the assembly of the first high quality reference genome for orach which was produced using Oxford Nanopore's MinION sequencing technique in conjunction with Illumina read data. I also report on the general genomic comparison of orach to other species in the genus of Chenopodium. These first genomic resources will serve as an important first step towards better understanding and improving orach.

W909: **Quinoa and close relatives**
Why do Amaranths have Betalains?
Michael K. Deyholos, University of British Columbia, Kelowna, BC, Canada and Dinesh Adhikary, University of British Columbia, Okanagan, Kelowna, BC, Canada

W910: **Resources and Programs for Undergraduate Education in Genomics**
Concepts before Coding: RNA-Seq Data Analysis for Undergraduate Students with Little Computational Experience using the CyVerse Discovery Environment
Judy A. Brusslan, California State University, Long Beach, Long Beach, CA
Mastering a computer language can often seem an insurmountable wall that prevents students from appreciating the power and logic of bioinformatics analysis. The CyVerse Discovery Environment is a platform that combines simple sharing of large data files with a stepwise, modifiable pipeline that allows students to explore parameters within a user-friendly interface. My lab has generated RNA-Seq data that allows comparison of genome-wide gene expression patterns during leaf senescence in *Arabidopsis thaliana* wildtype and histone acetyltransferase mutants. I will report on data sharing, quality filtering of RNA-Seq data, alignment to the genome and identification of differentially expressed genes (DEG) within the CyVerse Discovery Environment. Students can use the analysis pipeline with default and then modified parameters, providing an opportunity for testing of student hypotheses. Students use the Database for Annotation, Visualization and Integrated Discovery (DAVID) to identify Biological Processes (BPs) enriched in their lists of DEGs, and then are challenged to connect the enriched BPs to leaf senescence, thus relating the large-scale transcriptome analysis to a meaningful change in physiology. Students choose a gene based on log2 fold-change and FPKM values, and confirm differential expression on a biological replicate via real-time qPCR, thereby adding a wet lab component to the laboratory experience. Students summarize their findings in a report that conforms to the style of peer-reviewed scientific literature.
We are currently reviewing lab reports from the last two years to assess student comprehension of the following informatics concepts: 1) the power of genome-wide data analysis; 2) how parameter changes affect the catalog of DEGs; 3) BP enrichment; and 4) the relationship between BP and physiological changes. CyVerse allows students to gain a deeper understanding of transcriptome analysis and its physiological impact without the need to master command-line computer language. This course serves as a springboard for a more advanced Bioinformatics course offered in our Department of Biological Sciences.

W911: **Resources and Programs for Undergraduate Education in Genomics**
Student-Driven Characterization of the Microbiomes of Solitary Bees
Katie Burnette, University of California-Riverside, Riverside, CA

W912: **Resources and Programs for Undergraduate Education in Genomics**
Implementing Authentic DNA Barcoding Research in Large Enrollment Freshman Labs
Oliver Hyman, James Madison University, Harrisonburg, VA

W913: **Resources and Programs for Undergraduate Education in Genomics**
Biochemical and Bioinformatics Infrastructure to Support Metabarcoding Cures
Dave Micklos, DNA Learning Center, Cold Spring Harbor, NY

W914: **Resources and Programs for Undergraduate Education in Genomics**
The Genomics Education Alliance: Working Together to Facilitate Undergraduate Research in Genomics
Sarah C. R. Elgin[1], Douglas L. Chalker[2], Wilson Leung[1], Anne G. Rosenwald[3], Jason Williams[4] and The Genomics Education Alliance, (1)Washington University in St. Louis, St. Louis, MO, (2)Washington University in St Louis, St Louis, MO, (3)Georgetown University, Washington, DC, (4)CyVerse, Cold Spring Harbor, NY
The Genomics Education Alliance (GEA), a recently funded NSF Research Coordination Network in Undergraduate Biology Education, aims to bring together existing genomics education networks and interested faculty to make genomics accessible to undergraduate students from any community college, college or university. GEA will identify and curate common analysis tools, associated curricular and assessment materials,

and faculty training strategies to enable Course-based Undergraduate Research Experiences (CUREs) in genomics / bioinformatics. Vetted resources and training materials will be made available online through the QUBES platform (https://qubeshub.org/community/groups/gea), enabling a cost-effective system for maintaining up-to-date curriculum and training materials for a field that is constantly changing. Current members of GEA include the Genomics Education Partnership (GEP; http://gep.wustl.edu), GCAT-SEEK (https://gcat-seek.weebly.com), the Dolan DNA Learning Center (CSHL) (https://www.dnalc.org), Genome Solver (https://qubeshub.org/community/groups/genomesolver), and others. Biology faculty will be able to utilize the resources provided by the GEA to ameliorate the technical barriers and reduce the time and effort required to develop and maintain courses in genomics and bioinformatics for their students. Undergraduate students can gain a basic introduction to genomics, improve their understanding of the research process, and develop their research skills in analyzing large datasets as part of a CURE. Furthermore, massively parallel undergraduates can an partner with researchers to undertake large-scale projects (such as human-curated genome annotation) that cannot be accomplished in any other way. The GEA is currently looking for additional members to join our efforts, and we welcome your input on the most important genome analysis tools and curriculum materials to support. GEA is funded by NSF RCN-UBE grant #1827130.

W915: **Resources and Programs for Undergraduate Education in Genomics**
Thinking about Publishing Your Bioinformatics Education Work?
Anne Rosenwald, Georgetown University, Washington, DC
You've done the hard work of developing thoughtful lessons and curriculum in genomics and bioinformatics. You may have also assessed student outcomes stemming from your lessons and curriculum. Now you're thinking about publishing your ideas and results. Where should you turn? In this interactive session, we'll discuss opportunities and gaps in publications available for this sort of work.

W916: **Rhinoceros Genomics Studies: tools for conservation**
Sequencing the Black Rhinoceros to Inform Conservation Management using a Multiplex PCR Approach
Sergio A Redondo, Stanford University, Stanford, CA

W917: **Rhinoceros Genomics Studies: tools for conservation**
Deleterious Genetic Variation in Northern and Southern White Rhino Genomes
Aryn Wilder, San Diego Zoo Institute for Conservation Research, Escondido, CA

W918: **Rhinoceros Genomics Studies: tools for conservation**
Saving the Northern White Rhino: The Genetic Toolbox Needed to Save a Species
Marisa L. Korody, San Diego Zoo Institute for Conservation Research, Escondido, CA
Species are disappearing at an alarming rate due to human impact on our environment. Conservationists are attempting to offset these effects through research into genetic rescue and advanced reproductive techniques. With the advent of induced pluripotent stem cell (iPSC) technology in 2006, current and future applications are immeasurable for research and conservation. iPSCs are somatic cells that have been reprogrammed, or induced, into stem cells, can then be maintained indefinitely in an undifferentiated state, or directed to become any cell type in the body. Researchers can utilize this ability to study development in the lab and explore a variety of applications, including genetic rescue. iPSCs may hold the key to save endangered species for which populations are dwindling, but biomaterials have been preserved.
Northern white rhinos (NWR) are functionally extinct, with only two non-reproductive females remaining. The San Diego Zoo has developed an initiative which ties multiple fields of inquiry together in an attempt to save this species. We are pioneering the development of species-specific iPSCs to be utilized in genetic rescue through gamete generation for assisted reproduction, with the goal of maintaining population viability and genetic diversity to avoid extinction. The San Diego Zoo's Frozen Zoo® contains twelve (eight of which are unrelated) fibroblast cell lines from this species which makes them a candidate for genetic rescue. Previous genomic analysis has shown we have captured high genetic variability in these NWR cell lines. We have adapted technology pioneered in mouse and human cell lines to generate iPSC lines from nine of these northern white rhinoceros fibroblast lines to date.
A critical step to applying stem cell technologies to genetic rescue is to characterize the iPSCs we develop. Our interdisciplinary project joins the 'omics world with reproductive sciences and animal care to save a species from extinction. We are utilizing next generation sequencing technologies to develop a high quality, well annotated NWR reference genome. The high quality reference genome along with the cryopreserved biological materials from these rhinos will create the toolbox which we will utilize to save a species. We are analyzing our cells at every step to determine conserved and rhino specific genes involved in pluripotency and differentiation using immunocytochemistry (ICC) and quantitative PCR with rhino specific primers. We have further validated these markers with RNAseq analysis. The generation of reporter lines through the use of CRISPR for primordial germ cell (PGC) specific genes will be key in observing the pathways for germ cell development and maturation. We have successfully and repeatedly reprogrammed NWR fibroblasts to iPSCs and identified markers of pluripotency, differentiation and PGCs. Our goal is to use these stem cells to generate *in vitro* gametes and embryos to save the species from extinction.

W919: **Rhinoceros Genomics Studies: tools for conservation**
Transcriptomic Analysis of Pluripotency in the Northern White Rhinoceros Stem Cells
Iñigo Valiente-Alandi, San Diego Zoo Institute for Conservation Research, Escondido, CA

W920: **Rhinoceros Genomics Studies: tools for conservation**
Genome-Wide Signature of Adaptive Evolution in the Rhinocerotidae
Shanlin Liu, BGI-Shenzhen, Shenzhen, China

W921: **Rice Functional Genomics**

Comparative Genome Analyses of Four Geographically Distant, Rice-Infecting Isolates of *Rhizoctonia solani* Anastomosis Group 1-IA (AG1-IA)

Dayoung Lee[1], Jongbum Jeon[2], Ki-Tae Kim[3], Kyeongchae Cheong[2], Hyeunjeong Song[3], Gobong Choi[2], Jaeho Koh[2], Shimin Zuo[4], Sheshu Madhav[5], Yong-Hwan Lee[2] and Guo-Liang Wang[1], (1)Department of Plant Pathology, The Ohio State University, Columbus, OH, (2)Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul, South Korea, (3)Fungal Bioinformatics Laboratory, Seoul National University, Seoul, South Korea, (4)College of Agriculture, Yangzhou University, Yangzhou, China, (5)Indian Institute of Rice Research (ICAR-IIRR), Hyderabad, India

Rice (*Oryza sativa*) is the staple food crop for more than half of the world's population. However, rice sheath blight disease, caused by the fungus *Rhizoctonia solani* AG1-IA, greatly reduces the quantity and quality of rice production, leading to significant yield losses and threatening global food security. In this study, we sequenced four *R. solani* AG1-IA strains isolated from infected rice plants in four geographically distant locations in three countries (USA, India and China), using PacBio *de novo* sequencing and Illumina re-sequencing approaches. Phylogenetic analyses showed that isolate B2 (USA) is closely related to YN-7 (China), while ADB (India) is more related to RNR (India). Isolates YN-7, ADB and RNR possess similar genome sizes, approximately ranging from 38.9~40 Mb, while isolate B2 has a genome size of 45 Mb. Synteny conservation of 74.9%~82% was observed upon pairwise genome alignments of the four genomes. The core genome of the rice-infecting *R. solani* AG1-IA isolates is predicted to be 5,777 genes and 15~56 genes were identified to be isolate-specific. Furthermore, the *R. solani* AG1-IA genomes are enriched with carbohydrate-active enzymes (CAZymes), where extensive gene family expansion and diversification were observed in specific CAZyme families, such as the polysaccharide lyases (PLs). Our study has not only identified core genomic features shared by different isolates but also unique sequences in the rice-infecting *R. solani* AG1-IA group. The high-quality genome sequences from this study has provided rich genomics resources for better understanding the pathogenicity of this important fungal pathogen of rice.

W9**22: Rice Functional Genomics**

Using CRISPR/Cas9 to Target Resistance to the Rice Blast Pathogen

Yulin Jia[1], Haijun Zhao[2], Xueyan Wang[2], Bastian Minkenberg[3], Matthew Wheatly[3], Jiangbo Fan[4], Melissa Jia[5], Adam Famoso[6], Jeremy D. Edwards[7], Yeshi Wamishe[8], Barbara Valent[9], Guo-Liang Wang[10] and Yinong Yang[11], (1)USDA-ARS, Stuttgart, AR, (2)USDA ARS, 130, AR, (3)Pennsylvania state university, University Park, PA, (4)Ohio State University, Columbus, OH, (5)DB NRRC, Stuttgart, AR, (6)LSU AgCenter, Rayne, LA, (7)USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR, (8)University of Arkansas, 130, AR, (9)Department of Plant Pathology, Kansas State University, Manhattan, KS, (10)Department of Plant Pathology, The Ohio State University, Columbus, OH, (11)Pennsylvania State University, University Park, PA

Plant genes that confer resistance to pathogens such as the rice blast pathogen, *Magnaporthe oryzae,* typically encode nucleotide binding sites - leucine rich repeat (NLR)-type receptor proteins. Rice blast resistance (R) genes *Pi-ta* and *Pi-ta2* located near the centromere of chromosome 12 have been effectively deployed to prevent infections by *M. oryzae* for years. The *Ptr* gene was mapped within a 63 kilobase region on chromosome 12 using 11,618 segregating progenies derived from the cross of the Pita containing rice variety Katy with a susceptible rice variety Amane. The *Ptr* gene is located 211 kilobase away from the NLR *R* gene *Pi-ta* and another uncharacterized gene, *Pi-ta2*. Genetic analysis showed that *Ptr* confers broad spectrum disease resistance independent of *Pi-ta*. The *Ptr* gene encodes two predicted proteins that were mainly localized in the cytoplasm of plant cells. A two-base pair deletion within the *Ptr* coding region found in the fast neutron-created mutant line, M2354, produced a truncated protein, resulting in susceptibility to *M. oryzae*. Targeted mutation of *Ptr* in a resistant cultivar using CRISPR/Cas9 led to blast susceptibility, further validating its resistance function. The *Ptr* gene was predicted to encode a protein with four Armadillo repeats. DNA sequence analysis of 3000 rice accessions identified regions of high variability in Ptr suggesting that these regions may be involved in protein- protein interaction in triggering effective resistance. A new haplotype of *Ptr* was identified from a black hull awned weedy red rice (Ptr^{BHA}), and genetic analysis showed that Ptr^{BHA} is responsible for preventing the infections by a highly virulent blast race IB33. Weedy red rice is known to be compete with commercial rice, and functional validation of Ptr^{BHA} will shed more insights into adaptive immunity in rice. This knowledge is important to develop effective strategies to manage blast disease.

W9**23: Rice Functional Genomics**

An Actin Depolymerizing Factor Plays a Positive Role in Abiotic Stress Response in Rice through the Modulation of Cytoskeleton Architecture

Niranjan Baisakh, School of Plant, Environmental, and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA

W9**24: Rice Functional Genomics**

A Less-Biased Approach to Characterize Carbon-Ion Beam-Induced Mutations By Whole Exome Sequencing of Unselected M_2 Populations in Rice

Hiroyuki Ichida, Ryouhei Morita, Yuki Shirakawa, Yoriko Hayashi and Tomoko Abe, RIKEN Nishina Center for Accelerator-Based Science, Wako, Japan

Mutagenesis is a fundamental tool with which to investigate gene functions and create new cultivars in plant breeding. Heavy-ion beams are one of the physical mutagens that are classified in high-linear energy transfer (LET) radiations and known to induce double strand breaks of DNA in a cell along with its track. It have been widely utilized as an effective physical mutagen for mutation breeding in diverse plant species, but the induced mutation spectrum is not fully understood at the genome scale. Previous studies with morphological mutants revealed accelerated heavy-ion beams cause deletions, insertions, inversions and base substitutions into genome, however, such 'select and analyze' approach may reflect the effects of mutant selections.

We developed a cost-efficient whole-exome sequencing procedure in rice, and applied this technique to determine the spectrum of carbon-ion beam-induced mutations in unselected M_2 populations of rice. We sequenced a total of 165 individual M_2 lines derived from six irradiation conditions as well as eight pools from non-irradiated Nipponbare controls. The analysis indicated the induced mutations were mostly consisted

by single nucleotide substitutions (57.5% of all mutations events) and deletions (37.3%), and other types of mutations, such as insertions, replacements and inversions, were also found but less frequent (below 5%). The average number of mutations within the target exon regions was 9.06±0.37 induced by 150 Gy irradiation of dry seeds. We observed a good correlation between the total number of mutations detected by sequencing of unselected M_2 lines and the conventional mutation rate determined by the occurrence of morphological mutants. Therefore, the mutation frequency may be a good indicator for sequencing-based determination of the optimal irradiation condition for plant mutagenesis.

W925: **Rice Functional Genomics**
Unraveling the Retrotranspositional Landscape of Rice at Species Level Using 3000 Genomes
Marie-Christine Carpentier, University of Perpignan, Perpignan, France

W926: **Rice Functional Genomics**
TBD
Xuewei Chen, Sichuan Agricultural University, Chengdu, China

W927: **Root Genomics**
Barley Root Mutants to Uncover the Secrets of Hidden Parts of Plants
Miriam Szurman-Zubrzycka, Wydzial Biologii i Ochrony Strodowiska, Uniwesyet Slaski, Katowice, Poland

W928: **Root Genomics**
Molecular Dissection of the Legume Autoregulation of Nodulation Pathway
Brett Ferguson, Centre for Integrative Legume Research/School of Agriculture & Food Sciences, Brisbane,, Australia

W929: **Root Genomics**
Regulation of Defence and Metal Uptake Genes Roots by the Diazotrophic Endophyte *Herbaspirillum seropedicae*
Emanuel Souza, UFPR, Curitiba, Brazil

W930: **Root Genomics**
Linking Genomic and Phenomic of Rice Root to Discover Aluminum and Drought Tolerance Genes in Rice
M Miftahudin, Bogor Agricultural University, BOGOR, Indonesia

W931: **Root Genomics**
High Quality Assembly of White Lupin's Genome, a Model to Study Root Developmental Adaptations
Bárbara Hufnagel and Benjamin Peret, CNRS, Montpellier, France
White lupin (*Lupinus albus*; 2n=50) stands out as a model legume species since it is the only annual crop producing cluster roots, one of the most outstanding developmental adaptations to nutrient-scarce environments. We report a high-quality chromosome-scale assembly of white lupin genome, together with an extensive transcriptome data from ten different organs of that species. We took advantage of single-molecule real-time technology, in combination with short-reads sequencing and optical and genetic maps in order to have a successful assembly. The final assembly size is 443 Mb with a N50 of 17 Mb. About 98% (434 Mb) of the assembled genome is included on the 25 pseudo-chromosomes. The structural annotation identified 38 258 coding genes and 3129 ncRNA, being 97.3% genes anchored on the pseudo-chromosomes. A majority (94.6%) of the 1440 genes in the Plantae BUSCO dataset were identified in the annotation, which is suggestive of a complete assembly and annotation. White lupin genome revealed to be laden with gene duplications and repetitive elements. It is estimated to comprise around 55% of repetitive sequences, being the LTR retrotransposons the most abundant (~50%), followed by DNA transposons (3%) and tandem repeats (2%). In total, we have identified six families of highly enriched tandem repeats which were mostly localized at the (peri)centromeric regions of chromosomes. A comparative evolutionary analysis of white lupin genome with other legumes species revealed that it experienced a whole genome triplication (with a potential 72-chromosomes intermediate) in about 10 million years ago. It presents extensive duplication blocks inside its own genome and also a high degree of synteny with the close legumes species *Lupinus angustifolius* and *Medicago truncatula*.
We re-sequenced other 15 white lupin accessions, including a landrace and a non-domesticated variety. This has shown a highly polymorphic genome that has been impacted by domestication in different ways. Some transposons families present on the non-domesticated variety have disappeared in modern accessions, as well as the protein content of the seed has changed. Interestingly, the domestication has also modified cluster root formation. The cluster roots are formed earlier on time and closer to topsoil in the cultivated varieties. Altogether, this genome is a valuable resource and represents a keystone for legume genomics research.

W932: **Root Genomics**
Identification of Evolutionarily Conserved Genes Responsive to Root-Knot Nematodes from Root Transcriptome of Four Plant Species
Ana Zotta Mota[1], Fabricio B.M. Arraes[2], Erika V S Albuquerque[3], Anne-Sophie Petitot[4], Diana Fernandez[4], Patricia M. Guimarães[5], Ana C. M. Brasileiro[6], Maria F. Grossi de Sa[7] and Etienne Danchin[8], (1)EMBRAPA, BRASILIA, Brazil, (2)UFRGS, Porto Alegre, Brazil, (3)Embrapa, Brasília, Brazil, (4)Institut de Recherche pour le Développement, Montpellier, France, (5)EMBRAPA, Brasilia, Brazil, (6)EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil, (7)Embrapa, Brasilia, Brazil, (8)INRA, Antibes, France
Root-Knot nematodes (RKN), belonging to the *Meloidogyne* genus, can parasitize a range of plants worldwide, being one of the most important causes of losses in agriculture. Advances in plant genomics now allows uncovering the molecular events underlining incompatible interactions

between RKN and several different resistant plants. These developments allow a better understanding of genetic determinants of plant responses and resistance to these parasites.
We used OrthoFinder to identify groups of orthologs and in-paralogs across 22 plant species from different angiosperm families. We included available predicted proteome of 21 species and enriched this dataset with a new reference transcriptome for *Arachis stenosperma*, that we constructed.
The transcriptomes of four resistant plant species (*Arachis stenosperma*, *Coffea arabica*, *Glycine max* and *Oryza glaberrima*) were studied using RNA-seq, to identify genes differentially expressed during the early stages of RKN infection in comparison to non-infected control conditions.
For the *A. stenosperma* transcriptome assembly, several approaches to obtain the best balance between high completeness and low redundancy were tested, including data from 16 cDNA libraries of roots and leaves submitted to different stresses sequenced with HiSeq-2000 technology. After the raw data cleaning and trimming, a genome-guided strategy was applied, using the closest available genome (*Arachis duranensis*) as reference to reconstruct the transcriptome. We then used EviGene software to eliminate redundancy and CEGMA and BUSCO to assess the completeness of the transcriptome assembly. Finally, TransDecoder was used to predict proteins from the reconstructed transcripts.
The above *A. stenosperma* transcriptome was then used with the other 21 proteomes as input to OrthoFinder, which clusters the proteins in orthogroups based on reciprocal best blast hit relations.
The RNA-Seq data from the four resistant plant species infected with RKN were then mapped to the respective genome/transcriptome and produced read counts tables with HTSeq-Count. To identify genes differentially expressed upon nematode infection, we used both DESeq and EdgeR with only the genes with FDR < 0.05 and Log2FC > 2 and < -2.
OrthoFinder analysis yielded 35,238 orthogroups, of which 6,132 were evolutionarily conserved with all the 22 species. When comparing the four species, a total of 9,323 orthogroups were found in common among them. We found 778, 2757, 743 and 677 genes differentially expressed upon RKN infection for *A. stenosperma*, *O. glaberrima*, *C. arabica* and for *G. max*, respectively. We identified 18 orthogroups, evolutionarily conserved in the four species that contained at least 1 nematode-responsive genes from each of the four species of interest. One gene of each species for 15 out of the 18 groups were validated by qRT-PCR and their expression compared to the RNA-Seq data.
Despite the phylogenetic distance between the resistant species studied, the fact that the resistant plants have 18 orthologs groups in common, suggests that the plant response to the RKN infection might bear core common pathways.

W933: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops
Overview of the Nitrogen-Fixing Clade - the Discovery of a Single Origin of Predisposition
Douglas E. Soltis, University of Florida, Gainesville, FL
In traditional classifications of angiosperms, the 10 families known to have nitrogen-fixing symbioses in root nodules were considered distantly related, implying multiple origins of the underlying genetic machinery in diverse lineages and suggesting that transfer of the symbiotic ability across clades (for example, from legumes to grasses) would be feasible. It has been now over 20 years since early molecular phylogenetic analyses first suggested that in fact those 10 families were closely related and part of a single clade of more than 30,000 species, referred to in 1995 as the nitrogen-fixing clade. The authors noted that, “These findings, furthermore, suggest a single evolutionary origin of the underlying capacity for symbiotic nodular nitrogen fixation” and that, “Future efforts to unravel the process and evolution of nitrogen-fixing symbioses, and the transfer of this capacity to nonnodulating species, should first focus on related taxa from the nitrogen-fixing clade that possess and lack symbiotic nitrogen-fixing ability. Concomitantly, nodulating and nonnodulating members of this nitrogen-fixing clade should be examined to ascertain whether recurrent losses or recurrent gains of nitrogen-fixing ability have occurred.” The simple discovery of a nitrogen-fixing clade with a single predisposition for the origin of the symbiotic machinery, has enormous economic implications, with considerable interest in moving nodulation capabilities to nonnodulating crops. Renewed interest in the origin and evolution of nitrogen-fixing symbiosis is shedding new light on the process. It is exciting to see the recent investigations into the underlying genetic mechanism controlling nodulation and the rigorous testing of hypotheses of repeated loss vs. gain of nodulation within the nitrogen-fixing clade.

W934: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops
Global-Scale Phylogenomics of the Nitrogen-Fixing Clade
Heather Rose Kates, University of Florida, Gainesville, FL
Robust phylogenetic inferences on the origin of the predisposition to nodulation, and how and if nodulation has been gained and lost, are key to understanding the evolutionary lability and thus likelihood of successful transferability of N-fixing symbioses among lineages of angiosperms. A well-resolved and well-sampled N-fixing clade phylogeny is therefore a prerequisite to the discovery of genes that determine nodule development. Multiple phylogenetic analyses have been conducted on the N-fixing clade with the aim of elucidating the origins of N-fixing symbioses; however, these analyses have relied on trees estimated using a few genes and in which species sampling in the N-fixing clade was limited. We present our first steps toward a revised phylogeny based on deliberate and extensive sampling, phylogenomic data, and rigorous statistical analysis. These will allow more accurate inference of precursors of N-fixing symbioses, gain and/or loss events, and potential transferability of the capability to crop plants not in the N-fixing clade. Novel elements of our strategy that enable geographically and taxonomically comprehensive sampling on an ambitious scale include a protocol for rapid tissue sampling of 15,000 historical specimens, a high-throughput, high-yield DNA extraction protocol specifically suited to degraded DNA, a capture kit that works across phylogenetic scales and includes key functional genes, and scaleable information management and processing using a project database. Our preliminary phylogenetic results are based on 100 nuclear loci sequenced for over 2,000 species from the N-fixing clade and inform our interpretation of patterns of presence/absence of ~100 nodulation genes inferred from genome screening of all species in our phylogeny. This result represents an initial analysis comprising 13% of our genomic and taxonomic sampling effort. We will leverage the total phylogenetic and comparative genomic results to discover gene candidates that potentially underlie nodule development so that these genes can be tested for function in nodulating and non-nodulating model systems.

W935: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops
Pathway Discovery in Deep Convergence: Big Data in Phylogenomics for Nitrogen-Fixing Root Nodule Symbiosis

Shifeng Cheng, AGIS-CAAS & BGI-Shenzhen, Shenzhen, China, Jeffrey J. Doyle, Plant Biology Department, Cornell University, Ithaca, NY, Douglas E. Soltis, University of Florida, Gainesville, FL, Pierre-Marc Delaux, LRSV, France, Toulouse, France and Gane Ka-Shu Wong, Department of Biological Sciences, University of Alberta, Edmonton, AB, AB, Canada
Plants play a crucial role in ecosystem, food, nutrition, and medicine; while genome, the basic source code of life, is the fundamental connection in context of biodiversity that helps us better understand and exploit innovative traits in evolution through big data mining, resulting in a huge potential in science as well as in application. Many plant specialized genetic pathways (like C4 photosynthesis, natural medicinal bioactive products) – part of which form biosynthetic gene clusters or gene expression - are not distributed random, but exhibit a pattern enriched in a certain monophyletic clade or family along the tree of life. One of these is the Root Nodule Symbiosis that is exclusively presented in a scattered distribution in N-fixing Nodulation (NFN) clade. However, only ~0.1% of genomes from the 10 N-fixing families and their 18 closely-related non-nodulating relatives have been available, resulting in an elusive molecular clue on the origin and evolution of the RNS in terms of the proposed predisposition event or the recent gains or losses hypothesis. Much data and efforts are still required to elucidate the full-spectrum genetic innovations and genomic modifications that toward the establishment of the ancient RNS and the subsequent dynamic diversification. In the past several years, 1KP (1000 plant transcriptomes) and the newly initiated 10KP (10,000 plant genomes) project has and will continue to demonstrate the power of big data in plant genomics through a broad non-model species sampling and phylogenomics (or phylomics) analysis. In this talk, I will explain why we need big data in biodiverse sequencing and call for new international effort to dedicate large-scale genome sequencing and phylogenomics analysis, basically through the combination of "Data-driven" and "Hypothesis-driven" strategies to help us unravel the genetic toolkit assembly and evolution. All of these will set up a new mode for pathway discovery that is deep convergence, like root nodule symbiosis and many found elsewhere.

W936: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops
Evolution by Gene Loss in Plant-Microbe Symbioses
Jean-Michel Ane, University of Wisconsin, Madison, WI and Pierre-Marc Delaux, LRSV, France, Toulouse, France
More than 70% of land plants can associate with arbuscular mycorrhizal fungi. Fossil evidence suggests that this symbiosis was present more than 400 million years ago in the first land plants. Genes identified in legumes as required for the arbuscular mycorrhizal symbiosis such as *CCaMK* and *IPD3* are functionally conserved across all mycorrhizal lineages of land plants including the earliest diverging ones. Altogether, this suggests that arbuscular mycorrhizal associations appeared in a common ancestor of land plants and that a core set of genes has been used since then for these associations. However, many lineages of land plants are not able to establish such mycorrhizal associations suggesting that they have lost this ability. By comparing genomes of plants able to develop this symbiosis or not, we have shown that a set of genes including *CCaMK* and *IPD3* have been lost repeatedly in non-mycorrhizal lineages.
In contrast to arbuscular mycorrhizal associations, the ability to develop root nodules with nitrogen-fixing bacteria is restricted to land plants belonging to four angiosperm orders (Fabales, Fagales, Cucurbitales, and Rosales) that together form the monophyletic "nitrogen-fixing" clade. However, only ten out of the twenty-eight plant families within this clade contain species able to develop root nodules. By comparing genomes of plants able to form nodules or not within the "nitrogen-fixing" clade, we found that genes such as *NIN* and *RPG* have been lost repeatedly in non-nodulating lineages. These losses of genes required for nodule development and bacterial colonization suggest that root nodulation has been lost several times independently and that again these losses of symbiosis were accompanied by the loss of the same set of genes.
Our studies indicate that comparative phylogenomics is a powerful approach to study the evolution of plant-microbe symbioses. The availability of whole genome sequences allows the correlation of gene gains and losses to the symbiosis status which can enable the identification of genes involved in such symbiotic associations. The repeated losses of symbiotic associations also raise interesting questions on the costs associated with these associations, the benefits of losing them, and the subsequent adaptations of plants to these losses.

W938: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops
Nitrogen Fixation in a Landrace of Maize is supported by a Mucilage-Associated Diazotrophic Microbiota
Allen Van Deynze[1], Cris Heitmann[1], Pablo Zamora[1], Javier Lopez[2], Jonathan Eisen[1], Bart C. Weimer[1], Howard-Yana Shapiro[3], Jean-Michel Ané[4] and Alan Bennett[1], (1)University of California, Davis, CA, (2)Instituto TecnoloÂgico del Valle de Oaxaca, Oaxaca, Mexico, (3)University of California, Davis, Davis, CA, (4)Department of Agronomy - University of Wisconsin Madison, Madison, WI
Plants are associated with a complex microbiota that contributes to nutrient acquisition, plant growth, and plant defense. Nitrogen-fixing microbial associations are efficient and well characterized in legumes but are limited in cereals, including maize. We studied an indigenous landrace of maize grown in nitrogen-depleted soils in the Sierra Mixe region of Oaxaca, Mexico. This landrace is characterized by the extensive development of aerial roots that secrete a carbohydrate-rich mucilage. Analysis of the mucilage microbiota indicated that it was enriched in taxa for which many known species are diazotrophic, was enriched for homologs of genes encoding nitrogenase subunits, and harbored active nitrogenase activity as assessed by acetylene reduction and 15N2 incorporation assays. Field experiments in Sierra Mixe using 15N natural abundance or 15N-enrichment assessments over 5 years indicated that atmospheric nitrogen fixation contributed 29%±82% of the nitrogen nutrition of Sierra Mixe maize.

W939: Root Nodule Symbiosis: Genetics, Evolution, and Engineering for Future Crops
Engineering Nitrogen-Fixing Symbiosis into Poplar
Matias Kirst[1], Jean-Michel Ane[2], Douglas E. Soltis[1,3], Sushmita Roy[2], Pamela S. Soltis[1,3], Robert P. Guralnick[1,3], Daniel Conde[1], Folk Ryan[1], Thomas Irving[2], Lucas Maia[2], Heather Rose Kates[1], Kelly Balmant[1], Sara Knaack[2], Matthew Crook[4] and Christopher Dervinis[1], (1)University of Florida, Gainesville, FL, (2)University of Wisconsin, Madison, WI, (3)Florida Museum of Natural History, Gainesville, FL, (4)Weber State University, Ogden, UT
Nitrogen (N) availability is critical for plant productivity, and most species can only acquire it by absorbing N available from the soil as nitrates or ammonium. In contrast, some angiosperm species in the fabid clade have the capability to obtain N through a mutualistic relationship with N-fixing bacteria, which involves the development of root nodules that host these bacteria. Lateral roots and nodules are both root lateral organs and

share many structural and developmental similarities. Identifying and engineering the genome novelties that led to the mutualistic symbiotic relationship between plants and N-fixing bacteria is a logical path towards introducing that capability into non-nodular species. We are currently identifying these genome novelties using a comparative phylogenomic framework focused on contrasting genomes and transcriptomes derived from species in the fabid clade and outgroups (see the previous talk by H.R. Kates). In parallel, we are manipulating known genetic components involved in nodule development in a cell-specific manner in poplar (*Populus* spp.), a key woody bioenergy crop, to induce the formation of nodules. The latter approach has involved developing root cortex and epidermis-specific promoters to modify the expression of genes involved in nodule development, such as *Nodule INception* (*NIN*) and cytokinin receptors. This targeted approach has resulted in a significant increase in the number of root lateral organs. We are currently characterizing these lateral organs in poplar and assessing their capability to support the establishment of a symbiotic relationship with N-fixing bacteria.

W940: Seed Genomics

The Genetic Puzzle of the Canola Seed: Epigenetic and Transcriptomic Profiling of *Brassica napus* Seed Development

Deirdre Khan, University of Manitoba, Winnipeg, MB, Canada

Canola (*Brassica napus*) is one of Canada's most economically valuable crops. The two most valuable canola products – oil and meal – are both derived from the seed. We still know very little about the regulatory mechanisms that control valuable seed traits such as oil and protein content. We must have a more thorough understanding of the networks of gene expression that underlie seed development to take a more targeted and efficient approach to seed improvement.

We examine the complex architecture of canola seed genetics at the level of DNA methylation, small RNA targeting, and gene expression. Using bisulfite, nanopore, and high-throughput RNA sequencing techniques, we profile the canola seed at the ovule, globular, heart, mature green, and dry seed stages of development. Throughout seed development, the *B. oleracea* subgenome is more heavily methylated than the *B. rapa* subgenome, which correlates with the density and lower expression level of transposable elements in the *B. oleracea* subgenome. Late stage morphogenesis is associated with an increase in small RNAs expressed from the A genomes, concurrent with a small genome-wide decrease in TE expression. Furthermore, several bZIP transcription factors are identified as potential regulators of energy metabolism and development in canola seeds. We used RNA interference to knock down *BZIP11* expression in canola and found that reduced *BZIP11* expression is associated with lower target gene expression and seed lethality. By understanding the patterns of DNA methylation and gene expression, we are finding new tools to profile seed development and investigate the quality of canola seeds.

W941: Seed Genomics

Transcription Factor Regulatory Networks in Soybean Seed Development

Leonardo Jo, University of California Davis, Davis, CA

Soybean (*Glycine Max*) is the most widely cultivated and consumed oilseed crop in the world. Understanding the regulatory mechanisms that govern soybean seed development can provide opportunities to improve the quality of this important crop. Seed development is divided into two main phases, morphogenesis and maturation. The morphogenesis phase is characterized by a series of cell division and differentiation events that establish the basic body plan of the plant. Following morphogenesis, the seeds enter the maturation phase which is characterized by the accumulation of storage compounds and the embryo's acquisition of desiccation tolerance. Because seed development is a complex yet highly coordinated period of the plant life cycle, the temporal and spatial control of the biological events that occur are crucial to ensure the proper development of the plant. Therefore, the transcriptional control of genes involved in seed development needs to be highly coordinated. To define the transcription factors involved in soybean seed development we analyzed whole-genome transcriptome datasets to identify a group of co-expressed genes whose spatial and temporal expression patterns correlated with processes that occur in the embryo during the maturation phase. Several transcription factors that regulate seed development were identified in the clusters, including LEAFY COTYLEDON1 (LEC1), ABSCISIC ACID INSENSITIVE3 (ABI3), BASIC LEUCINE ZIPPER67 (bZIP67) and ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3). To identify target genes that are transcriptionally regulated by LEC1, ABI3, bZIP67 and AREB3, we performed chromatin immunoprecipitation and differential gene expression analyses during the early maturation stage. Detailed analysis of target genes showed a complex transcription factor regulatory network in which different combination of transcription factors are involved in controlling distinct biological programs in soybean embryos. Analysis on transcription factors binding regions revealed cis-regulatory modules that are important to determine the expression of genes involved with distinct biological programs that occurs during seed development. DNA sequence motif enrichment analyses suggested that the formation of distinct transcriptional complexes is determined by the presence of unique combinations of DNA motifs in the cis-regulatory modules. Transient analysis in protoplasts isolated from soybean embryos revealed that these elements are crucial to determine the functionality of specific cis-regulatory modules. We also observed that distinct sets of transcription factor complexes formed due their ability to physically interact to each other. These results are helping to elucidate the complex transcriptional regulatory networks that control distinct processes that occur during soybean seed development.

W942: Seed Genomics

Maternal Small RNAs Mediate Spatial and Temporal Regulation of Gene Expression and Seed Development in Arabidopsis

Ryan Kirkbride[1], Jie Lu[1], Changqing Zhang[2], Rebecca A. Mosher[3], David C Baulcombe[4] and Z. Jeffrey Chen[1], (1)The University of Texas at Austin, Austin, TX, (2)North Carolina State University, Kannapolis, NC, (3)University of Arizona, Tucson, AZ, (4)University of Cambridge, Cambridge, United Kingdom

Arabidopsis seed development involves maternal small interfering RNAs (siRNAs) that induce RNA-directed DNA methylation (RdDM) through the *NRPD1* pathway. To investigate their biological functions, we characterized siRNAs in the endosperm and seed coat that were separated by laser-capture microdissection (LCM) in reciprocal genetic crosses with an *nrpd1* mutant. We identified distinct groups of siRNA loci that were dependent on or independent of the maternal *NRPD1* allele in the endosperm or seed coat. A group of maternally expressed *NRPD1*-siRNA loci targets endosperm-preferred genes, including those encoding AGAMOUS-LIKE (AGL) transcription factors. Using *AGL40:GUS* and *AGL91:GUS* constructs as sensors, we demonstrate that spatial and temporal expression patterns of these genes in the endosperm were regulated by the *NRPD1*-mediated pathway, and altered expression of these siRNA-targeted genes affects seed size; we propose that the corresponding maternal siRNAs could account for parent-of-origin effects on the endosperm in interploidy and hybrid crosses.

W943: **Seed Genomics**

Maternal and Zygotic Effects on Kernel Size in Maize

Shawn Kaeppler[1], Xia Zhang[2], Kathryn Michel[3], Nathan D. Miller[3], Edgar P. Spalding[3] and Natalia de Leon[4], (1)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI, (2)Biotechnology Research Institute, Chinese Academy of Agricultural Sciences (BRI,CAAS), Beijing, China, (3)University of Wisconsin-Madison, Madison, WI, (4)Department of Agronomy UW, Madison, WI, WA

Maize yield is a result of the number of plants per hectare, the number of kernels per plant, and the mass of each kernel. Increases in yield in the hybrid era have been attributed primarily to increasing plant density, improving density tolerance and reducing barrenness, and maximizing the number of ovules filled under stress. Despite dramatic phenotypic variation for kernel size in maize, increased yield has not been accompanied by increased kernel size although there is evidence for increased duration of the grain-filling period. Maize is an ideal system to study genes expressed in the endosperm, and this has facilitated characterization of important pathways such as starch synthesis and storage protein accumulation. Characterization of maize genes that affect all kernels on an ear (maternal effects) have been substantially less-studied as mutations affecting these processes require a second generation to identify and are hampered by plant-to-plant variability relative to assessing kernels varying on the same ear. In an analysis of a long-term selection program for kernel size in the Krug population, we determined that the rate of kernel maturation is much faster in the small-kernel versus the large-kernel types. Evaluation of diverse inbred panel, and multi- and bi-parental inbred populations support that the largest proportion of variation in open-pollinated trials is due to maternal genetics with little influence of pollen sources in standard trials. Our ongoing efforts to use novel phenotyping methodology coupled with meta-analysis across multiple populations to dissect developmental and GxE effects on the yield component of kernel mass and size will be discussed.

W944: **Seed Genomics**

Dynamics of Tissue-Specific Genome Regulatory Programs in the Germinating Barley Seed

Lim Chee Liew[1], Reena Narsai[1], Oliver Berkowitz[2], Yan Wang[1], James Whelan[1] and **Mathew Lewsey**[1], (1)La Trobe University, Bundoora, Australia, (2)University of Western Australia, Crawley, Australia

We are studying tissue-specific genome regulatory programs in germinating barley seeds so as to better understand how they are controlled by transcription factors. The roles of transcription factors in genome-wide regulation of gene expression are subject to much attention currently, but determing these roles is complex. Regulatory events occur dynamically over short time scales (minutes to hours), with many transcription factors interacting to regulate the output of any given gene and cascades of interacting factors operating across the wider program. Studies have historically depended on analysis of bulk tissue samples, due largely to the capabilites of available techniques. However, individual cell-types must be specified by distinct regulatory programs. To investigate cell-type specific genome regulation during germination we have applied laser-capture microdissection RNA sequencing to three tissues of barley seeds that have distinct functions (plumules, radicles and scutellum). We analysed complete transcriptomes from samples of 200 cells over a 36 h time series, enabling us to observe the dynamic changes in gene expression. The data allow us to identify modules of gene expression that are unique to and conserved between these three tissues and to identify associated groups of potential cis-regulatory motifs. By extrapolation from the extensive Arabidopsis transcription factor target datasets we are also able to infer likely families of regulatory factors. The outcome is an improved understanding of the direct regulation of gene expression and how this varies between tissues in a crop species with a relatively complex genome.

W945: **Sequencing Complex Genomes**

Prospects of Pan-Genomics in Barley

Martin Mascher, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany, Gatersleben, Germany

The concept of a pan-genome refers to intraspecific diversity in genome content and structure, encompassing both genes and intergenic space. Pan-genomic studies employ a combination of de novo sequence assembly and reference-based alignment to discover and genotype structural variants. The large size and complex structure of Triticeae genomes were for a long time an obstacle for genomic research in barley and its relatives. Now that a reference genome is available, computational pipelines for high-quality sequence assembly are in place, and sequence costs continue to drop, investigations into the structural diversity of the barley genome seem within reach. We propose the following strategy: (1) the construction of high-quality de novo sequence assemblies for a small core set of representative genotypes, (2) short-read sequencing of a large diversity panel of genebank accessions to medium coverage and (3) the use of complementary methods such as chromosome-conformation capture sequencing and k-mer-based association genetics. The selection of barley genotypes for pan-genome analysis will be based on the genome-wide genotypic data for the entire genebank collection of the German Federal ex situ genebank comprising more than 20,000 accessions representative of global barley diversity.

W946: **Sequencing Complex Genomes**

***De novo* Genome Sequencing and Hybrid Assembly of *Punica granatum*, a Complex Fruit Genome**

Nripendra Vikram Singh[1], Roopa Sowjanya P.[1], Shilpa Parashuram[1], Prakash Patil[1], Babu K. Dhinesh[1], Jyotsana Sharma[1], Sathish Kumer Pedamma[2], Harish Kothandaraman[2], Tapan Baral[2], Sailu Yellaboina[2], Lalitha Bommu[2], Abhishek Tripathi[2], Bipin Kumar[2], Umesh K. Reddy[3] and **Dushyant Singh Baghel**[2], (1)ICAR-National Research Centre on Pomegranate, Solapur, India, (2)Nucleome Informatics Private Limited, Hyderabad, India, (3)Department of Biology, West Virginia State University, Institute, WV

Pomegranate (Punica granatum L.) belongs to the family Lythraceae and is considered a powerhouse of nutrition and medicinal properties. It is a strategic crop for ensuring nutritional and livelihood security in water-scarce regions having suboptimal edaphic conditions. India is the global leader in pomegranate production, and acreage and ‘Bhagwa’ is the ruling variety having medium-sized red fruits yielding up to 30 kg per tree. We present a draft genome assembly for pomegranate cv. Bhagwa using Illumina, PacBio Sequel, 10X Genomics Chromium datasets. The genome was initially assembled using PacBio Trimmed reads using Canu, and 10X Chromium Linked Reads using SuperNova 2.0. Canu Assembly represented 322Mb in 1637 scaffolds and SuperNova Diploid Assembly yielded an assembly of 298Mb in 2256 scaffolds which were further reduced to 1395 scaffolds using trimmed PacBio reads. Finally, the Canu assembly was taken as a base due to a better genome breadth,

BUSCO and other assembly metrics, to reconcile the PacBio assembly with the Super-Scaffolded SuperNova assembly to cover 309Mb of the genome via High Confident Overlap using QuickMerge. Scaffolds/Contigs not taking part in the Reconciliation procedure were analysed for duplicates using BBMAP package and added to the QuickMerge assembly. The hybrid assembly comprised of 341Mb (N50:4.8Mb, 97.6% BUSCO) in 6627 scaffolds, on the estimated genome size of 350-356Mb. Modelling and masking the repeats suggested that over 40% of the genome harboured repeats. At every step during the assembly super-scaffolding, reconciliation and addition of data back to the genome, we validated the genome using multiple metrics. Paired-End short insert datasets were mapped to the genomes to correct and gap-fill the artefacts at each step. Transcriptome mapping and BUSCO validation of the assembled datasets showed minute but significant improvements at each step validating our Assembly Protocol.
We are using optical mapping, and Hi-C approaches to build chromosome level assembly. The high-quality genome sequence information will act as a valuable resource to dissect economically essential traits and propel the pomegranate breeding program.

W947: Sequencing Complex Genomes
Sequencing Reveals the Origins of Domesticated Barley in Tibet
Yu Guo, BGI, Shenzhen, China
Barley has been used as the staple food in Tibet for at least 3,500 years but the history of barley's origin in Tibet still remains unclear. We investigated the origin and domestication history of Tibetan barley - qingke, based on a newly generated deep-coverage whole genome as well as published exome capture re-sequencing data for a total of 437 wild and domesticated barley accessions. This revealed that contemporary qingke was derived from eastern domesticated barley providing genomic evidence that the earliest barley was introduced to southern Tibet most likely via north Pakistan, India, and Nepal between 4,500 and 3,500 calendar years before the present (cal yr B.P.). The low genetic diversity of qingke showed that Tibet can be excluded as a center of origin or domestication for barley. The rapid decrease in genetic diversity from eastern domesticated barley to qingke can be explained by a founder effect from ~4,500 to ~2,000 cal yr B.P. The haplotypes of the five key domestication genes of barley supported a feral or hybridization origin for Tibetan weedy barley rejecting the hypothesis of native Tibetan wild barley.

W948: Sequencing Complex Genomes
Draft Genome Sequence of *Oryza coarctata*, a Halophytic Species of Genus *Oryza*
Tapan Kumar Mondal, ICAR-National Research Center on Plant Biotechnology, New Delhi -110012, India
Salinity is one of the major abiotic stresses that limits rice cultivation worldwide. Rice is canonically a glycophyte except few genotypes. Those genotypes have been extensively used for salinity breeding of rice. However, finding the alternative salt tolerant novel allele from its wild species will be an alternative approach. *Oryza coarctata* (KKLL; 2n = 4x = 48, 665 Mb) also known as *Porteresia coarctata* is an extreme halophyte species of genus *Oryza* which can set seeds up to the salinity level of 40 EC. Being the member of the same genus of rice, it can be useful reservoir of salinity tolerant genes for rice. However to harness the genomic resources that are present in this species, whole genome sequencing is a prerequisite. Using Illumina and Oxford Nanopore reads, we achieved the assembled genome size of 569.9 Mb, accounting 85.69% of the estimated genome size with N50 of 1.85 Mb and 19.89% repetitive region. We also found 230,968 simple sequence repeats (SSRs) and 5,512 non-coding RNAs (ncRNAs). The functional annotation of predicted 33,627 protein-coding genes and 4,916 transcription factors revealed that high salinity adaptation of this species is due to the exclusive or excessive presence of stress-specific genes as compared to rice. We have identified 8 homologs to salt-tolerant *SOS1* genes, one of the three main components of salt overly sensitive (*SOS*) signal pathway. On the other hand, the phylogenetic analysis of the assembled chloroplast (134.75 kb) and mitochondrial genome (491.06 kb) favours the conservative nature of these organelle genomes within *Oryza* taxon. The detail of the salinity specific pathway genes found in this species will be discussed.

W949: Sequencing Complex Genomes
Sequencing the Genome of the Mushroom Coral (*Heliofungia actiniformis*) with Multiple Long Read Technologies
Ira R Cooke, James Cook University, Townsville, Australia, Shana L. McDevitt, University of California Berkeley - QB3, Berkeley, CA, Lutz Froenicke, Genome Center, DNA Technologies Core, University of California, Davis, CA, David Hayward, Research School of Biology, Canberra, Australia, Aurelie Moya, ARC Centre of Excellence in Coral Reef Studies, Townsville, QLD, Australia and David Miller, Department of Molecular and Cell Biology, Townsville, QLD, Australia
Heliofungia actiniformis is a mushroom coral popular amongst marine aquarists, but that also shows promise as a model species for coral biology. Although it belongs to the same taxonomic group as do the colonial corals that are the primary architects of coral reefs, the physical characteristics of this solitary coral present some unique advantages as a model species. In particular, the large size (>50 cm diameter) and permanently extended tentacles of *Heliofungia* enable the tissues involved in calcification or symbiosis to be dissected out, whereas this would be extremely difficult to do in colonial corals due to the small size of polyps. As it provides a means of addressing previously intractable questions about coral biology, to underpin molecular studies we are generating a high quality reference genome assembly for this species. Sequencing coral genomes is challenging because they are repetitive and often highly heterozygous. In addition, their abundant dinoflagellate endosymbionts complicate the extraction of DNA of appropriate size and purity for genome sequencing. By taking advantage of the one night each year when this species spawns, the contamination issue was overcome by collecting symbiont-free coral sperm and using this as the starting material for DNA extraction. We then sought to overcome problematic repeats and heterozygosity by sequencing with a range of long and linked read technologies including high coverage PacBio, Oxford Nanopore and 10x chromium.
In this talk we describe lessons learned from our sequencing and assembly experiences. We reveal relative strengths of these three technologies by comparing assemblies obtained from each in isolation and assembly strategies that combine them for generating contigs, scaffolds and phased haplotypes.

W950: Sequencing Complex Genomes
From Pond to Genome: Challenges Faced in the Black Tiger Shrimp (*Penaeus monodon*) Genome Assembly
Roger Huerlimann, James Cook University, Townsville, Australia

W951: **Sex Chromosomes and sex determination**

Sequencing and Annotation of the Genome and Sex Chromosomes in *Rumex hastatulus*

Joanna Rifkin, University of Toronto, Toronto, ON, Canada

Non-recombining sex chromosomes have evolved repeatedly in eukaryotes. Their evolution is believed to proceed in a stepwise fashion through the addition of "strata" of suppressed recombination that capture loci experiencing sexually antagonistic selection. Improvements in the quality of genome assemblies in plants offer powerful tools to study this process because plants frequently have evolutionarily young sex chromosomes and close relatives without separate sexes.

The genus Rumex (Polygonaceae) includes hermaphroditic, gynodioecious, and dioecious species. The genus contains between two and four transitions from autosomes to sex chromosomes. Rumex hastatulus provides a key opportunity to characterize the speed and timing of degenerative evolution; in one geographic region, males have four autosomes and an XY pair, while in another, an X-autosome fusion event generated a composite X which pairs with an old and a neo-Y chromosome. Using a new genome phased male genome assembly combined with comparative linkage maps for both karyotype varieties we address the following questions:

1) What are the positions and contents (genes, pseudogenes, and repetitive sequence) of strata of recombination suppression on the old and new Y chromosomes?

2) Do loci in the XY pseudoautosomal region and captured autosome show evidence of sexually antagonistic selection?

3) What are the population genomic causes and consequences of the X-autosome fusion?

W952: **Sex Chromosomes and sex determination**

DNA Methylation and Genetic Degeneration of the Y Chromosome in the Dioecious Plant *Silene latifolia*

José Luis Rodríguez Lorenzo, Biofyzikální ústav Akademie věd České republiky, v.v.i., Brno, Czech Republic

Silene latifolia is a model organism for the study of sex chromosome evolution in plants. Its sex chromosomes include large regions in which recombination became gradually suppressed. The regions tend to expand over time resulting in the formation of evolutionary strata. Non-recombination and later accumulation of repetitive sequences is a putative cause of the size increase in the Y chromosome. Gene decay and accumulation of repetitive DNA are identified as key evolutionary events. Transposons in the X and Y chromosomes are distributed differently and there is a regulation of transposon insertion by DNA methylation of the target sequences, this points to an important role of DNA methylation during sex chromosome evolution in *S. latifolia*. The aim of this study was to elucidate whether the reduced expression of the Y allele in *S. latifolia* is caused by genetic degeneration or if the cause is methylation triggered by transposons and repetitive sequences. Several genes belonging to stratum I (older) or stratum II (younger) were examined. Gene expression analysis in *S. latifolia* males has shown expression bias in both X and Y alleles. To determine whether these differences are caused by genetic degeneration or methylation spread by transposons and repetitive sequences, we selected several sex-linked genes with varying degrees of degeneration and from different evolutionary strata. Immunoprecipitation of methylated DNA (MeDIP) from promoter, exon and intron regions was used and validated through bisulfite sequencing. We found DNA methylation in males, and only in the promoter of genes of stratum I. The Y alleles in genes of stratum I were methylation enriched compared to X alleles. There was also abundant and high percentage methylation in the CHH context in most sequences, indicating de novo methylation through the RdDM pathway. These two main conclusions suggest that the age of the stratum in the chromosome affects methylation of the genes in the stratum. The position in the chromosome also has an influence on methylation as well as the proximity of TEs and repetitive sequences. We speculate that TE accumulation and not gene decay is the cause of DNA methylation in the *S. latifolia* Y sex chromosome with influence on the process of heterochromatinization. Preliminary data from the analysis of different histone post-translational modifications indicates, up to date, that histone H4 acetylated in Lys 5, 8, 12 and 16 is neither limited to one stratum nor to X or Y allele.

W953: **Sex Chromosomes and sex determination**

Silkworm Sexual Regulation through W Chromosome-Linked, Targeted Gene Integration

Yongping Huang, Shanghai Institute of Plant Physiology and Ecology, Shanghai, China

Male silkworms produce better filament than their conspecific females. Therefore, it has been a long time approach for silkworm breeders and researchers. Here, we report a method to construct an inheritable male only silkworm strain through integration of sexual regulation cassette into W-chromosome and obtained an expected strain. The chromosome composition of Silkworm is ZW. The heterozygous ZW is female, whereas the homozygous ZZ is male. In the study of the key gene function of sex determination pathway, we found that the mutation of *Bmtra2* resulted in the embryo lethal. Therefore, we targeted gene integration of fluorescent marker expression cassettes onto a RAPD (randomly amplified polymorphic DNAs) marker region in the W chromosome of the lepidopteran model insect, *Bombyx mori*, using TALENs (transcriptional activator-like effector nucleases)-mediated genome editing. This silkworm strain shows female-specific red or green fluorescence ubiquitously from embryonic to adult stage. Furthermore, we developed a binary, female-specific, embryonic lethality system combining TALEN and the cluster, regularly interspaced, short palindromic repeats (CRISPR)-CRISPR-associated protein (CRISPR/Cas9) system, which includes one strain with TALEN-mediated, W-specific Cas9 expression driven by the silkworm *nanos* (*nos*) promoter and another strain with U6-derived sgRNA expression targeting *transformer 2* (*tra 2*), an essential gene for silkworm embryonic development. F_1 hybrids exhibit complete female-specific lethality during embryonic stages. Our study thus provides a promising approach for *B. mori* genetic sexing and sheds light on developing SIT in other insect species especially in lepidopteran pests with WZ/ZZ sex chromosome systems.

W954: **Sex Chromosomes and sex determination**

The Fire Ant Social Supergene Pair Differs by Multiple Large Inversions

John Wang, Academia Sinica, Taipei, Taiwan

Supergenes consist of co-adapted loci that segregate together and are associated with adaptive traits. In the fire ant *Solenopsis invicta*, two 'social' supergene variants regulate differences in colony queen number and other traits. This supergene pair shares many features with sex chromosomes, including at least one large inversion. Supergenes in other systems and sex chromosomes are often composed of more than one large inversion. To determine if this is also the case for the fire ant, we first cloned one extreme breakpoint in the fire ant supergene. In doing so, we found a second large rearrangement. Because the fire ant supergene lacks obvious evolutionary strata, our finding of multiple inversions may

support an introgression model of the supergene. Additionally, we found that one of the inversions swapped the promoter of a breakpoint-adjacent gene, which might have conferred a selective advantage relative to the non-inverted allele. These findings advance our understanding of the evolution of the fire ant supergene.

W955: Sex Chromosomes and sex determination
The Enrichment of Testis-Specific Genes on the Mammalian X and Y Chromosome
Wansheng Liu, Department of Animal Science, Penn State University, University Park, PA
Each chromosome in a genome should, theoretically, contain randomly mixed collections of genes with extremely heterogeneous patterns of developmentally regulated expression in different tissues. This is true for all autosomes, but not for the sex chromosomes in mammalian genomes. The gene content of the sex chromosomes is strikingly different from that of the autosomes. Both sex chromosomes appear to be enriched for genes related to sexual differentiation, brain development, and reproduction. In this presentation, I will focus on the enrichment of testis-specific genes on the mammalian sex chromosome based on gene annotation data from primate (human), rodent (mouse) and ruminant (cattle) lineages. Throughout evolution, mammalian sex chromosomes have accumulated large amounts of X- and/or Y-specific repetitive sequences originated mainly from the “autosome-to-sex chromosome” or “sex chromosome-to-autosome” transposition/retroposition events. Through the transposition/retroposition system, sex chromosomes frequently exchange genetic materials with autosomes in a lineage-specific manner and accumulate a disproportional amount of reproduction related genes, expressed predominantly in the testis. These testis genes are usually amplified in the X- and/or Y-specific repetitive sequences and are difficult to be sequenced and to be properly annotated. Furthermore, the nearly identical multicopy nature of the testis gene families on the sex chromosomes and their broad functions in immunity (such as cancer/testis antigens in cancer development) and brain function complicated research work on their molecular mechanisms. The future research directions will also be discussed.

W956: Sex Chromosomes and sex determination
Structural Variation between the U and V Sex Chromosomes in the Moss *Ceratodon purpureus*
Stuart McDaniel, University of Florida, Gainesville, FL
The possession of heteromorphic sex chromosomes is widely believed to promote the evolution of sexual dimorphism. Nevertheless, rigorous tests of this hypothesis are lacking. Sexual dimorphism results from differential gene expression between males and females, which can derive from either genes with sex-limited transmission (i.e., sex chromosomes) or from sex-specific regulation of autosomal loci. To evaluate the relative roles of sex linkage and sex-biased expression of autosomal loci, we examined the transcriptomes of multiple developmental stages from a novel model system, the dioecious moss *Ceratodon purpurues*. Sex in dioecious bryophytes is determined by a UV sex chromosomal system. In this system, each sex has a non-recombining chromosome (U for females and V for males). The U and V pair at meiosis in the monomorphic diploid sporophyte and segregate to the haploid male and female gametophytes. We generated RNAseq data from two developmental stages from female-male sibling pairs from multiple, geographically disparate populations of *C. purpureus*. We found that autosomal loci showed consistent sex-biased expression differences, but the majority of the transcriptional differences between the sexes derived from sex-linked loci. Moreover, the patterns of sex-linked expression differences, and even sex linkage, varied among populations, suggesting that such differences may evolve quickly. Collectively these data demonstrate that gene traffic onto the sex chromosomes can promote rapid changes in sexual dimorphism, and may occur frequently, in spite of the deleterious consequences of the suppressed recombination associated with sex-limited transmission.

W957: SGN and RTB Databases: Genomics and Breeder Tools.
Introduction and Update on the Tomato Genome
Lukas A. Mueller, Boyce Thompson Institute, Ithaca, NY

W958: SGN and RTB Databases: Genomics and Breeder Tools.
Update on Tomato Expression Data
Jocelyn KC Rose, Plant Biology Department, Cornell University, Ithaca, NY

W959: SGN and RTB Databases: Genomics and Breeder Tools.
Breeding Informatics Platform for Root, Tuber, Banana and Solanaceae Crops
Alex C. Ogbonna[1,2], Isaak Y. Tecle[1], Naama Menda[1], Guillaume J. Bauchet[1], Bryan Ellerbrock[1], Nicolas Morales[1], David A. Lyon[1] and Lukas A. Mueller[1], (1)Boyce Thompson Institute, Ithaca, NY, (2)Cornell University, Ithaca, NY
Solgenomics (SGN) serves as platform for breeders and breeding programs to routinely carry out their breeding activities. From germplasm curation, designing of experiments, phenotyping, genotyping and analytical pipelines for decision making as well as visualization tools. Solgenomics provides Solanaceae (SGN), Root, Tuber and Banana (RTB) crops with seamless integrated informatics database workflow, aimed at increasing breeders' efficiency and reducing errors associated with data handling. In addressing different community needs, dedicated and dynamic web portals (solgenomics.net, cassavabase.org, sweetpotatobase.org, yambase.org, musabase.org) tailored to the specific breeding needs of a crop with versatile tools that efficiently share and retrieve data along with a gwas and genomic selection module components. This presentation will highlight the breeding informatics tools available on SGN platform.

W960: SGN and RTB Databases: Genomics and Breeder Tools.
Breeding Data Analysis and Visualization Tools for the Solanaceae and RTB Crops
Adrian Powell, Boyce Thompson Institute, Ithaca, NY
With the advent of high-throughput sequencing and phenotyping techniques, it is now possible to generate large quantities of phenotype, genotype and expression data. The databases of the Sol Genomics Network (SGN) enable users to access this data. However, interpretation and synthesis of information from such datasets can still be challenging. SGN has developed several tools to assist in analyses of these datasets. SolGWAS is an implementation of genome-wide association methods to assess association of markers with selected phenotypes, using data from

the SGN databases. In addition to histograms and PCA plots, SolGWAS provides the user with Manhattan plots and Q-Q plots as results of the association analyses. For analysis of gene expression data, the SGN expression atlases also provide tools that enable visualization and synthesis. The first expression atlas developed was the Tomato Expression Atlas; several additional expression atlases have been developed or are currently in development. The source code for the SGN expression atlases is available at https://github.com/solgenomics/Tea.

W961: **SGN and RTB Databases: Genomics and Breeder Tools.**
Visualization Tools for Breeding Data Exploration
David A. Lyon, Boyce Thompson Institute, Ithaca, NY

W962: **SGN and RTB Databases: Genomics and Breeder Tools.**
Application of Nicotiana 'Omics Resources for Molecular Farming
Jennifer Bromley, British American Tobacco, Cambridge, United Kingdom
From its biochemistry and the underlying genetic architecture that regulates it, the tobacco plant is highly complex and by comparison with other model species, poorly understood. Traditional incremental improvements have in the past provided successes to adapting tobacco but these techniques alone may no longer be sufficient to sustain the crop for future applications, particularly in the arena of molecular farming..
Technological developments in areas such as sequencing and mass spectrometry, as well as data storage and analytical capabilities, have led to an era in which 'big data' plays a major role in biological research. This is pertinent to tobacco, where increasing amounts of genomic and transcriptomic data are becoming publicly available, including assemblies of the tobacco genome. This data is providing fresh insight into the workings of the tobacco plant, and the tens of thousands of genes that its genome contains.
We have recently published (and continue to improve) an assembly of the tobacco genome, (Edwards *et al*., 2017 DOI:10.1186/s12864-017-3791-6) which has facilitated much of our recent progress into understanding the plant's complexities. The genome serves as a scaffold from which we are able to hang various 'omics data sets. Because the tobacco genome is annotated to contain 69,500 genes, generating and positioning this data is only the starting point, decoding it is where the real challenge lies.
A number of resources are available at Sol Genomics Network (SGN, https://solgenomics.net/) including Basic Local Alignment Search Tool (BLAST), JBrowse genome browser, a Virus Induced Gene Silencing (VIGS) construct designer, a metabolic pathway database and a genetic map viewer. In addition to these tools, we have overlaid the Apollo annotation curation plugin to the JBrowse genome browser which has allowed manual curation to improve the quality of gene annotation. This is now available as a community based manual curation tool for tobacco annotations.
Transcript evidence-based manual curation has been performed for a number of genes associated with secondary metabolic products which are undesirable for co-production with a tobacco "farmed" protein or metabolite, most notably the nicotinic alkaloids. Various alterations have been made to gene models, the most common of which being the extension of 5′ and 3′ untranslated regions, though other curations include modification of splice boundaries to non-canonical splice sites, insertion or deletion of introns, and addition of exons overlooked during automated annotation due to the presence of large (>15kb) introns. These improved gene models were fed into the SGN hosted virus induced gene silencing (VIGS) construct design tool which has allowed greater specificity of targeting of individual genes in multi-gene families.

W963: **SGN and RTB Databases: Genomics and Breeder Tools.**
Ricebase: A Database Integrating Genetic Discovery and Molecular Breeding Tools for Rice
Jeremy D. Edwards, USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR

W964: **Simulation of Genetic and Genomic Systems**
The Breeding Scheme Language: An R Package to Simulate Breeding Schemes from Simple to Complicated
Jean-Luc Jannink, USDA-ARS / Cornell University, Ithaca, NY and Shiori Yabe, Institute of Crop Science, NARO, Ibaraki, Japan
Many plant breeding schemes work in the sense of generating gain from selection, but that does not mean that they are optimal. Optimizing schemes can be done on the basis of theory or empirically. In the former case, changes in the genetic architecture resulting from selection (e.g., the Bulmer effect) are difficult to account for. In the latter case, time, expense, and error often overwhelm the value of experiments. These problems leave stochastic simulation as an option. Simulation can help breeders change their schemes to take advantage of new technologies or shift their schemes toward more optimal resource allocation. Unless the breeding program has a dedicated computational unit, however, simulation is likely to be too complicated. We created a simulation package in the statistical computing environment R. Users define their target species, trait genetic architectures, and breeding schemes by writing simple, self-explanatory scripts. Each function in the package executes a recognizable breeding task (i.e., "phenotype", "select", or "cross"), defining a "BreedingSchemeLanguage", which is the package name. The package seeks simplicity. The fact that it uses the R environment, however, gives it substantial flexibility. Package functions can be intermixed with standard R code. There are nevertheless tradeoffs along the simplicity to flexibility spectrum. In addition, because the package is written in R, speed and scalability are an issue.

W965: **Simulation of Genetic and Genomic Systems**
Population Genetic Simulations: General Concepts and New Tools
Phillip Messer, Cornell University, Ithaca, NY

W966: **Simulation of Genetic and Genomic Systems**
Optimizing Iterative QTL-Seq with Crossword Simulation Suite
Josh Clevenger, Mars Wrigley Confectionery, Athens, GA

W967: **Simulation of Genetic and Genomic Systems**
Applied Breeding Simulations: Using Genomewide Markers to Predict Genetic Correlations
Jeffrey L. Neyhart, Aaron Lorenz and Kevin P. Smith, University of Minnesota, St. Paul, MN
Plant breeders often select on multiple quantitative traits that may be favorably or unfavorably correlated. To improve multiple-trait genetic gain, breeding crosses could be selected based on the expected genetic correlation among offspring. Predictions of the genetic correlation in a prospective cross may be made using simulated populations and genomewide marker effects; however, empirical validation is needed, and the conditions necessary for accurate predictions are unknown. To validate this prediction method, we used a barley (*Hordeum vulgare* L.) training population (TP) to predict the pairwise genetic correlation for three different quantitative traits in 330,072 possible crosses. From these predictions, we subsequently selected, developed, and phenotyped 27 bi-parental populations. The predictive ability of the genetic correlation between traits was variable (r_{MP} = -0.01 – 0.41), but consistent with expectations given the heritability of the traits. Using breeding simulations, we further explored the conditions driving the prediction accuracy of genetic correlations. We perturbed the heritabilities of two traits, the TP size, genetic correlation architecture (i.e. pleiotropy, tight linkage, or loose linkage), and model type (i.e. BayesC or RR-BLUP). While trait heritability and TP size explained much of the variation in prediction accuracy, genetic correlations were predicted best under tight linkage, followed by loose linkage and then pleiotropy. Interestingly, prediction accuracy under conditions of pleiotropy was markedly improved when using a variable selection model (BayesC). Our study demonstrates the utility of simulations to inform decision-making in an applied breeding program.

W968: **Simulation of Genetic and Genomic Systems**
Stochastic Simulation: A Framework for Designing, Optimising, Teaching about and Playing with Breeding Programs
John M. Hickey, University of Edinburgh, The Roslin Institute, Edinburgh, United Kingdom

W969: **Single Cell Genomics**
Tracking Genome Architecture in Space, in Time, and in Single Cells
Erez Lieberman Aiden, Baylor College of Medicine, Houston, TX

W970: **Single Cell Genomics**
A Multiomic Approach at the Single Cell Level to better understand the Transcriptional Regulation of Plant Genes
Marc Libault, University of Nebraska Lincoln, Lincoln, NE

W971: **Single Cell Genomics**
Single-Cell Transcriptome Profiling of Plant Tissue
Diane Dickel, Lawrence Berkeley National Laboratory, Berkeley, CA
Single-cell transcriptome profiling of heterogeneous tissues can provide high-resolution windows into the spatiotemporal dynamics of developmental processes and environmental responses. Despite the widespread use of such technologies to probe metazoan development, their application to plant tissues has been more limited and technically more challenging. We used the high-throughput Drop-seq approach to profile the transcriptomes of >12,000 individual cells from *Arabidopsis* roots. This identified numerous distinct cell types, covering all major root tissues and developmental stages, and it illuminated cell populations that have not been transcriptionally well characterized previously. We identified marker genes for all cell types, enabling the exploration of the gene expression changes underlying cell type specification. We also demonstrate the utility of this approach to study the impact of environmental conditions on developmental processes. Single-cell analysis of roots grown with or without sucrose supplementation revealed no profound changes in cell type identity, but uncovered specific changes in the relative frequencies of cell types in response to sucrose. Finally, we characterized the transcriptome changes that occur across endodermis development and identified nearly 800 genes with dynamic expression as this tissue matures. Collectively, we demonstrate that single cell RNA-seq can be used to quickly profile thousands of plant cells, providing a detailed view of development and how it can be altered by external conditions. Our results suggest that these techniques should be broadly applicable to any plant tissue or species from which protoplasts can be generated, and their widespread adoption will open exciting new lines of inquiry into plant biology.

W972: **Single Cell Genomics**
Single Cell Genomic Sequencing in *Brassica napus*: Application in Monitoring Recombination Frequency
Venkatesh Bollina[1], Yifang Tan[1], Chu Shin Koh[2], Pankaj Bhowmik[1], Erin Higgins[3], Tryston Orr[1], Alison M.R. Ferrie[1], Isobel Parkin[3], Andrew G. Sharpe[1] and **Sateesh Kagale**[1], (1)National Research Council Canada, Saskatoon, SK, Canada, (2)Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada
Meiotic recombination creates genetic diversity and is the fundamental process underlying most crop breeding programs. Efforts to enhance meiotic recombination frequency are routinely undertaken by plant geneticists and breeders. However, the fast progress of such projects is hindered by the lack of a rapid screening method for monitoring the impact of the modulation of recombination in polyploid crops. We have devised a microspore-based single cell genomic sequencing and genotyping strategy for monitoring the frequency of meiotic recombination. Single cell haploid microspores from an F_1 plant is ideal material to quickly assess homoeologous recombination frequencies as it is relatively easy to isolate thousands of microspores carrying segregating genotypes. Our method involves DNA isolation from single microspores derived from F_1 progenies using either the Fluidigm C1 single cell platform or the fluorescence assisted cell sorting (FACS) based microspore sorting. Subsequent sequencing of DNA and genotyping of multiple segregating microspores facilitate assessment of the frequency of homoeologous recombination. This F_1 microspore-based sequencing approach will provide a cost effective breeding tool for genotyping and assessment of recombination frequency in F_1 microspores, thus negating the need for resources and time to be allocated towards the generation of a large F_2 population. These results along with the broad applications and challenges of single cell genomics in plants will be presented.

W973: **Single Cell Genomics**
TBA
Preyas Shah, 10x Genomics Inc., Pleasanton, CA

W974: **Single Cell Genomics**
iScFlow: Addressing the 4 V's of Big Data Challenge for Single Cell RNA-Seq Harnessing the Power of High Performance Computing
Parwinder Kaur, Univ. of Western AU, Perth, WA, Australia
The last decade has witnessed a paradigm shift in the way transcriptomics data is being generated and analysed. It is now possible to interrogate every cell of an organism in order to decipher the important biological processes that occur within. In this respect, single cell RNA-Sequencing (scRNA-Seq) has emerged as a ground-breaking technology that has greatly enhanced our understanding of the complexity of gene expression at a microscopic resolution. Given the hype, it is anticipated that in the next 5-10 years, the wider research community will be routinely employing this powerful and revolutionary technology as a laboratory staple. This is likely to create an exponential growth of scRNA-Seq data, leading to the 4 V's (volume, velocity, veracity and variety) of the Big Data problem, a preview of which is the Human Cell Atlas (HCA) project. This data deluge is likely to create challenges for storage, transfer, analysis, and visualization. Moreover, existing sequential algorithms for scRNA-Seq data analysis are inadequate for Big Data, and high-performance computing (HPC) solutions are almost non-existent. We therefore introduce iScFlow, which is a HPC-adapted pipeline that incorporates the 3 main aspects of scRNA-Seq data analysis, which are raw data pre-processing, post-processing and downstream analyses, including visualisation. We benchmarked iScFlow using the largest HCA dataset available to data (1.3 Tb) and the HPC facilities of the Pawsey Supercomputing Centre to illustrate the power of HPC at handling Big Data.

W975: **Small RNA**
Role of Small RNAs in Establishing Hybridization Barriers in Plants
Guifeng Wang[1], Hua Jiang[2], **Gerardo Del Toro De León**[2], German Martinez Arias[2] and Claudia Köhler[2], (1)Henan Agricultural University, Zhengzhou, China, (2)Swedish University of Agricultural Sciences (SLU), Uppsala BioCenter, Uppsala, Sweden
Small RNAs have regulatory roles at either the posttranscriptional level by RNA cleavage or translational inhibition of their targets, or at the transcriptional level by triggering DNA methylation at specific loci, particularly transposable elements (TEs). Epigenetic activation of TEs in the male reproductive accessory cell of pollen results in the synthesis of epigenetically activated small RNAs (easiRNAs). Previously, we have shown that the unbalanced dosage of easiRNAs is associated with hybridization barriers after fertilization, resembling the role of piRNAs in *Drosophila*. Depletion of easiRNAs in pollen strongly reduces expression of paternally expressed imprinted genes (PEGs) in the placenta-like endosperm tissue and restores hybrid seed viability, indicating that easiRNAs act upstream in controlling PEG expression. In *Arabidopsis* we have shown that the paternally expressed imprinted gene *PEG2* is causally involved in establishing a reproductive barrier in the endosperm after hybridization of plants that differ in chromosome number. I will present data showing that the *PEG2* transcript acts as a sponge and sequesters the TE-derived small interfering RNA *siRNA854* in the endosperm. Depletion of *siRNA854* as a consequence of increased *PEG2* transcript levels inhibits the targeting of *siRNA854* to its natural target genes like *UBP1b*. *SiRNA854* is present in accessory (vegetative) cell of pollen and transferred to the precursor cell of the endosperm after fertilization where it is captured by *PEG2*. Thus, the balance of a male gamete accumulating TE-derived small RNA and a PEG regulate hybrid seed viability, revealing a transgenerational epigenetic speciation mechanism.

W976: **Small RNA**
Diverse Functions and Multi-Regulation of miR528 in Rice
Xiaofeng Cao, Institute of Genetics and Developmental Biology, CAS, Beijing, China
MicroRNAs (miRNAs) are a class of endogenous small noncoding RNAs that negatively regulate gene expression. In plants, the biological functions of conserved miRNAs on their target mRNAs are well-explored, but the mechanisms that regulate miRNA accumulation are poorly understood.
We previously identified miRNA528 (miR528) as a subset conserved miRNAs uniquely found in monocotyledons and one of the most abundant miRNAs in rice. Compared with most miRNAs in plants, miR528 confers unique feature for targeting different families of genes involved in distinct biological processes. For example, miR528 is repressed and sequestered by AGO18 in plants that are infected by Rice stripe virus (RSV) that causes tremendous rice yield losses. We further found that miR528 negatively regulates the mRNA coding for the L-ascorbate oxidase, which regulates the accumulation of reactive oxygen species (ROS). In addition, miR528 promotes flowering under long-day conditions. Intriguingly, we find that the miR528 accumulation is co-regulated by development stage, light and diurnal rhythms. In this talk, I will discuss the evolutionary signatures and detailed mechanisms for fine-tuned accumulation of mature miR528 in rice.

W977: **Small RNA**
Small RNA Functions in Arabidopsis Embryos
Michael Nodine, Gregor Mendel Institute of Molecular Plant Biology GmbH, Vienna, Austria

W978: **Small RNA**
A Small RNA Pathway Regulates Germline mRNAs only *via* the 3' UTR Regions in *C. elegans*
Weifeng Gu, University of California - Riverside, Riverside, CA
We have recently identified a small RNA pathway which regulates hundreds of germline mRNAs only via the 3' UTR regions in C. elegans. To our knowledge, this is the first evidence that a specific small RNA pathway only targets the 3' UTR. In C. elegans germlines, endogenous small RNAs, 22Gs, bind Argonautes to regulate almost all germline genes. There are two major types of 22Gs: one group binding Argonaute CSR-1 and playing important roles in chromosome segregation and embryonic development; the other binding multiple Argonautes, WAGOs, and playing critical roles in silencing aberrant transcripts including transposons and pseudogenes. These 22Gs are generated by nonprocessive RNA-dependent RNA polymerases (RdRPs) using RNAs templates including both coding regions and UTRs. Here we are reporting that a novel small

RNA pathway specifically targets the 3' UTRs of hundreds of important germline genes. We have examined the biogenesis process and function of this particular pathway and are investigating how this pathway may affect translation.

W979: Small RNA

Uncovering Atypical Biogenesis and Function of Plant microRNAs using Next-Generation Sequencing Data

Ho-Ming Chen, Agricultural Biotechnology Research Center, Academia Sinica, Nankang, Taipei, Taiwan

Canonical plant microRNAs (miRNAs) are 21-nt long and repress the expression of target genes by guiding endonucleolytic cleavage of targeted RNA at the miRNA complementary site. However, plant miRNAs of atypical sizes or possessing the ability to inhibit translation have been widely reported. We demonstrated the use of small RNA sequencing data and RNA degradome data in the study of 20-nt miRNA biogenesis. The sequencing data analyses and experimental validation showed that asymmetric bulges or mismatches at specific positions are crucial for 20-nt miRNA formation. Moreover, our findings of ribosome and exon-junction complex (EJC) protected fragments in the RNA degradome led to the discovery of miRNA non-canonical action. Footprints of stacked ribosomes upstream of non-cleavable miRNA target sites provided evidence supporting ribosome stalling by plant miRNAs. On the other hand, as EJCs deposited to newly synthesized mRNAs are displaced during the pioneer round of translation, EJC footprints occurring downstream of some miRNA-guided cleavage sites suggested the regulation of some plant miRNAs on newly synthesized mRNAs that have not completed the pioneer round of translation. Taken together, our work revealed structural features accounting for shorter plant miRNAs and unique action of some plant miRNAs in ribosome stalling or targeting newly synthesized mRNAs.

W980: Small RNA

Small RNAs and Plant Developmental Patterning

Aman Husbands, Ohio State University, Columbus, OH

Flat leaves highlight mechanisms by which small RNAs mediate developmental patterning

Signaling is the foundation of development. Intercellular communication coordinates the growth of organs, generating highly complex morphologies in a remarkably reproducible manner. For instance, many plant species reiteratively produce leaves with nearly-identical flat, thin shapes that maximize photosynthetic efficiency. This is a difficult problem, as leaves do not start out flat, but rather emerge from the stem cell niche as radially-symmetric bumps that then develop into long and wide, yet shallow, structures. Leaves have solved this problem by using the boundary between their dorsal (top) and ventral (bottom) sides as a guide to orient their growth. Flat leaf morphology thus depends on a properly-patterned dorsoventral axis, which must be carefully regulated to avoid fitness consequences. We have identified properties that are critical for this proper patterning, including mutually antagonistic behavior of dorsal and ventral determinants, and intercellular signaling by mobile small RNAs. Intriguingly, small RNAs have the intrinsic ability to create sharp expression boundaries of their targets, behaving much like morphogens in animal systems. These patterning properties confer robustness to dorsoventral patterning, and in turn the uniform, stable positioning of the dorsoventral boundary. The implications of these findings will be discussed in developmental and evolutionary contexts.

W981: Solanaceae

A Common Genetic Mechanism underlies Morphological Diversity in Fruits and other Plant Organs

Esther van der Knaap, University of Georgia, Institute of Plant Breeding, Genetics & Genomics, Athens, GA

W982: Solanaceae

Genetic Analysis of Fruit Quality Traits in Pepper

Arnaud G. Bovy, Wageningen UR Plant Breeding, Wageningen, Netherlands

W983: Solanaceae

Single Primer Enrichment Genotyping Highlights the Worldwide Population Structure of Tomato and Eggplant Germplasm

Lorenzo Barchi[1], Alberto Acquadro[1], Ezio Portis[1], Sergio Lanteri[1], Davis Alonso[2], Pietro Gramazio[2], Santiago Vilanova[2], Maria Josè Díez[2], Jaime Prohens[2], Jeremy Salinier[3], Veronique Lefebvre[3], Gancho Pasev[4], Stanislava Grozeva[4], Hatice Filiz Boyaci[5], Abdullah Unlu[5], Laura Toppino[6], Laura Bassolino[6], Giuseppe Leonardo Rotino[6], Andreas Boerner[7], Ronny Brandt[7], Nils Stein[7], Richard Finkers[8], Arnaud G. Bovy[8], Roland Schafleitner[9], Davide Scaglione[10], Eleonora Di Centa[10], Sara Pinosio[10], Giuseppe Aprea[11], Paola Ferrante[11] and Giovanni Giuliano[11], (1)DISAFA, Plant Genetics and Breeding, University of Torino, Grugliasco, Italy, (2)COMAV-UPV, Valencia, Spain, (3)INRA, UR1052 GAFL, Montfavet, France, (4)Maritsa Vegetable Crops Research Institute (MVCRI), Plovdiv, Bulgaria, (5)Bati Akdeniz Agricultural Research Institute (BATEM), Antalya, Turkey, (6)CREA-GB, Montanaso Lombardo, Italy, (7)IPK Gatersleben, Stadt Seeland, Germany, (8)Wageningen UR Plant Breeding, Wageningen, Netherlands, (9)World Vegetable Center, Tainan, Taiwan, (10)IGA Technology Services, Udine, Italy, (11)ENEA, Rome, Italy

G2P-SOL (http://www.g2p-sol.eu) is an EU-funded project, bringing together the main European and international genebanks hosting germplasm of the four major Solanaceous crops: potato, tomato, pepper and eggplant. 23,900 tomato and 5,900 eggplant accessions, including wild relatives of both crops, have been inventoried within the project.

To gain information about population structure of the collections, the novel Single Primer Enrichment Technology (SPET, US Patent 9,650,628) developed by NuGEN was used for genotyping. An SNP/indel panel was developed by assaying 14k probes for tomato and 11k for eggplant, evenly distributed in the gene-rich regions, and selecting the 5k best performing probes for each species.

DNA samples were prepared by the genebanks using pre-tested protocols, and genotyping partners (ENEA for tomato and University of Torino for eggplant) performed QC and sample normalization. Genotyping and sequencing was performed by IGA Technology Services. Reads were aligned to the eggplant and tomato reference genomes using BWA-MEM and SNP calling was performed using GATK-4.0.

We report on the assessment of the genetic relationships in a wide set of tomato and eggplant accessions maintained in genebanks as well as the extent of duplications and possible mis-classifications. The results suggest that SPET genotyping is a reliable, high-throughput, low cost technology for genetic fingerprinting of crops, with a high degree of cross-transferability to their wild relatives. SPET-based higher density genotyping is being developed to characterize the core collections developed in G2P-SOL for GWAS analyses.

W9**84: Solanaceae**
Potato Wild Pangenome Placeholder
Maria Kyriakidou, Department of Plant Science - McGill University, Ste. Anne de Bellevue, QC, Canada

W9**85: Solanaceae**
Tomato Traditional Placeholder
Antonio Granell, IBMCP, Valencia, Spain

W9**86: Sorghum/Millet**
Introduction of Sorghum and Millet Workshop
Yinghua Huang, USDA ARS, Stillwater, OK

W9**87: Sorghum/Millet**
The Sorghum QTL Atlas: A Powerful Tool for Trait Dissection, Comparative Genomics and Crop Improvement
Emma Mace, Queensland Alliance for Agriculture and Food Innovation (QAAFI), Hermitage Research Facility, Brisbane, Australia and David R Jordan, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, Australia
The mechanisms governing the genetic control of many quantitative traits are only poorly understood and have yet to be fully exploited. Over the last 2 decades, over a thousand QTL and GWAS studies have been published in the major cereal crops including sorghum, maize and rice. A large body of information has been generated on the genetic basis of quantitative traits, their genomic location, allelic effects and epistatic interactions. However, such QTL information has not been widely applied by cereal improvement programs and genetic researchers world-wide. In part this is due to the heterogeneous nature of QTL studies which leads QTL reliability variation from study to study. Using approaches to adjust the QTL confidence interval, this platform provides access to the most updated sorghum QTL information than any database available, spanning 23 years of research since 1995. The QTL database provides information on the predicted gene models underlying the QTL CI, across all sorghum genome assembly gene sets and maize and rice genome assemblies and also provides information on the diversity of the underlying genes and information on signatures of selection in sorghum. The resulting high resolution, open-access research platform facilitates candidate gene identification across 3 cereal species, sorghum, maize and rice. Using a number of trait examples, we demonstrate the power and resolution of the resource to facilitate comparative genomics approaches

W9**88: Sorghum/Millet**
***Setaria* as a Model System to Explore Domestication in Grasses**
Andrew Doust, Margarita Mauro-Herrera and Hao Hu, Oklahoma State University, Stillwater, OK
The grass family (Poaceae) is economically and ecologically important, and contains multiple domestications that, when analyzed, allow us to study the genetics of morphological change. We are particularly interested in domestication traits such as flowering time, non-shattering, and plant architecture, and our findings to date suggest that the gene networks underlying these traits are only partially conserved, and that both gene identity and network function are labile over evolutionary timescales. Much of our work is on the genus *Setaria*, where our work on the ancient domesticate foxtail millet (*S. italica*) and its wild progenitor green foxtail (*S. viridis*) have shed light on evolution within the panicoid grasses, and has illustrated the importance of building new model systems to study development and evolution. We have also used these studies in our Research Experiences for Teachers (RET) program, coupling research experience with professional development, in order to raise the effectiveness and confidence of high school science teachers and the appreciation of science by their students.

W9**89: Sorghum/Millet**
Genome Assembly and Resequencing of Finger Millet Facilitated by the Advance of Polyploid Genomics
Kentaro K. Shimizu[1,2], Masaomi Hatakeyama[1,3], Lucy Poveda[3], Sirisha Aluri[3], Catharine Aquino[3], Ralph Schlapbach[3], Mathi T. Balachadran[4], Sajeevan Radha Sivarajan[4], Yellodu Yellodu Nanja Reddy[5], Rajashekar Reddy B H[5], Ramapura L. Ravikumar[4] and Sheshshayee M. Sreeman[4], (1)University of Zurich, Zurich, Switzerland, (2)Yokohama City University, Yokohama, Japan, (3)Functional Genomics Center Zurich, Zurich, Switzerland, (4)University of Agricultural Sciences, Bangalore, India, Bangalore, India, (5)Division of Nutrition, St. John's Research Institute, Bangalore, India, Koramangala, Bengaluru, India
Polyploid species are common in wild and crop plants. However, genome assembly, resequencing and RNA-seq of polyploid species has been difficult due to the sequence similarity of duplicated genes or homeologs. To facilitate genome assembly of polyploid species, we showed that the optical molecular data using BioNano is a powerful tool to conduct hybrid assembly of polyploid genomes. To conduct transcriptome and RNA-seq analysis, our validation using the model allopolyploid Arabidopsis kamchatica and wheat showed that tools developed for diploid species such as Kallisto may map often to wrong subgenomes. We developed subgenome-classification approaches by using HomeoRoq and EAGLE-RC for polyploid species.
Finger millet (Eleusine coracana) is an allotetraploid species derived from *E. indica* and another unknown diploid species. It is an important crop in India and Africa owing to its drought tolerance and high levels of nutrition. We reported the genome the assembly of the cultivar PR202 with N50 length >2 MB. Using the PR202 as the reference, genome-wide SNPs were identified by resequencing 24 cultivars. Rapid decay of linkage disequilibrium suggested the feasibility of association mapping of finger millet.

W990: **Sorghum/Millet**

Translational Genomics for Sorghum: Developments and Deployment

Gloria Burow[1], Zhanguo Xin[2], Ratan Chopra[3], John Burke[1], Paxton Payton[1], Yinping Jiao[4] and Doreen Ware[5], (1)USDA-ARS, Lubbock, TX, (2)USDA ARS, Lubbock, TX, (3)Dept. of Agronomy & Plant Genetics/ University of Minnesota, St. Paul, MN, (4)USDA-ARS/Cold Spring Harbor Laboratory, Lubbock, TX, (5)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

The completion and publication of the sorghum genome sequence led to advances in molecular genomics research for the species. However, the translation and leveraging of genomic discoveries in sorghum for practical use as DNA markers for crop improvement is still limited. As in many cereal crops, translational genomics is a critical key towards elevating yield and quality from current plateau values for sorghum. This presentation will focus on development and deployment of marker assisted rapid trait introgression for three sorghum traits to implement translational applications and enable marker assisted selection. The demonstration of the use of causal gene markers for the three traits resulting in effective selection will be discussed.

W991: **Sorghum/Millet**

In-Season and On-Target Prediction for Traits of Complex Plasticity in Diverse Environments

Xianran Li, Xin Li, Tingting Guo, Qi Mu and Jianming Yu, Iowa State University, Ames, IA

The bottleneck of understanding and leveraging phenotypic plasticity, the varied performance of same genotype, is the quantification of relevant environmental stimuli from natural environments. I will present the JGRA framework established from our multi-environment trial with a sorghum RIL population.

The RIL population were grown across 4 years at three fields spanning latitude from 18˚ to 42˚. We recorded a complex flowering time pattern from these seven tested environments. We discovered that photothermal time (PTT = GDD × day length) from a growth period (18-43 days after planting) can be used to quantify external stimuli of each environment as its high correlation with mean flowering time. By leveraging PTT, the complex flowering time can be explained, modelled, and predicted with a simple straightforward model. The power of in-season and on-target prediction from this model was empirically validated with data from next two seasons.

Encouraged by the successful application for bi-parental populations, we are expanding the applications of JGRA framework into in-season and on-target forecasting other traits for diverse and elite crop germplasm.

W992: **Sorghum/Millet**

Leveraging Multiple Genomic Resources to Dissect Nonstructural Sugar Accumulation in *Sorghum bicolor*

Zachary Brenton, Clemson University, Clemson, SC

Despite the enormous importance of nonstructural sugar accumulation in sorghum, very little is understood about the distinct molecular mechanisms leading to an accumulation of these sugars. A multitude of researchers have examined candidate genes related to carbohydrate metabolism and transport, but causality has not been established. By leveraging a newly established association panel and a *de novo*assembled reference genome specifically chosen to identify the causes of nonstructural sugar accumulation, we identify a tandem gene duplication with no known effect of sugar accumulation or metabolism that strongly associated with sugar accumulation. The newly discovered candidate implicated through GWAS seems antithetical to pervious work characterizing genes through *a priori* candidate selection. This works may be held in exemplary role for highlighting the advantages and disadvantages of specific approaches while demonstrating the value of creating multiple contrasting references for trait dissection.

W993: **Sorghum/Millet**

Sorghumbase, an Online Portal for the Sorghum Community

Ivar P. Meijs, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

W994: **Soybean Genomics**

Functional Soybean Genomics: Bridging the Genotype-to-Phenotype Gap in Disease Resistance

Michelle A. Graham, USDA-ARS-MWA-CICGRU, Ames, IA

Our research team has used virus-induced gene silencing to characterize candidate disease resistance genes for the fungal pathogens *Phakopsorapachyrhizi*and *Phytopthorasojae*.*P. pachyrhizi*is the causal agent of Asian soybean rust. Germplasm screening and genetic analyses have led to the identification of seven loci, *Rpp1– Rpp7* (Resistance to *P.pachyrhizi*), that provide varying degrees of resistance.*Rpp4*limits fungal growth and sporulation through the formation of reddish-brown (RB) lesions associated with a hypersensitive response. Silencing of *Rpp4*candidate genes in the resistant parent resulted in tan lesions with fungal sporulation, mirroring the susceptible response. We recently silenced candidate genes for *Rpp1*, which confers immunity, or a lack of visible systems, to *P. pachyrhizi*. Surprisingly, silencing of *Rpp1*candidate genes resulted in the formation of RB lesions without sporulation. Similarly, we recently silenced candidate genes for *Rps2* (*Resistance to Phytophthora sojae 2*). Silencing *Rps2*candidate genes resulted in an extreme hypersensitive response, even in the absence of the pathogen. In the case of the *Rpp1*and *Rps2 silencing experiments*, the candidate genes contained novel integrated domains which could impact resistance gene function and other pathways unrelated to defense. Analyses of whole genome sequencing data across species have revealed the complexity and number of resistance genes with integrated domains within plant genomes. Understanding the impact of integrated domains on resistance gene function is essential to bridge the genotype to phenotype gap in disease resistance.

W995: **Soybean Genomics**

Genetic Association between Reniform Nematode and Soybean Cyst Nematode Resistance in Soybean

Mariola Klepadlo, University of Missouri, Columbia, MO

W996: **Soybean Genomics**

Soybean Resistance to Sudden Death Syndrome

Hao-Xun Chang[1], Ruijuan Tan[1], Zixiang Wen[1], Hyunkyu Sang[1], Leslie Domier[2], Steven Whitham[3], Mitchell Roth[1], Silvia Cianzio[3], David Lightfoot[4], Glen Hartman[5], Dechun Wang[1] and Martin Chilvers[1], (1)Michigan State University, East Lansing, MI, (2)University of Illinois, USDA-ARS, Urbana, IL, (3)Iowa State University, Ames, IA, (4)Southern Illinois University, Carbondale, IL, (5)USDA-ARS, Urbana, IL

W997: **Soybean Genomics**

Symbiotic Incompatibility *via* Effector-Triggered Immunity between Soybean *Rj2*-Genotype and Bradyrhizobial NopP

Masayuki Sugawara[1], Yosuke Umehara[2], Satoko Takahashi[1], Akito Kaga[2], Masaki Hayashi[2], Masao Ishimoto[2], Shusei Sato[1], Hisayuki Mitsui[1] and Kiwamu Minamisawa[1], (1)Graduate School of Life Sciences, Tohoku University, Sendai, Japan, (2)National Agriculture and Food Research Organization, Tsukuba, Japan

Rj2-genotype soybeans restrict nodulation by specific bradyrhizobial strains including *Bradyrhizobium diazoefficiens* USDA 122. It has been revealed that a plant TIR-NBS-LRR resistance protein (Rj2) and rhizobial type III secretion system (T3SS) are responsible for this symbiotic incompatibility[1, 2], suggesting that effector-triggered immunity underlying pathogenic host-bacteria interactions control this genotype-specific host specificity. In this study, we aimed to 1) identify rhizobial T3SS effector inducing *Rj2*-mediated incompatibility, and 2) investigate the distribution of *Rj2*-genotype in Japanese soybean germplasm.

In order to identify the effector, we have obtained spontaneous mutants of *B. diazoefficiens* USDA 122 that overcome *Rj2*-mediated incompatibility. The results of resequencing of the genomes and further genetic analysis revealed that a type III-secretory protein NopP is a causal effector of this incompatibility. The analysis of *nopP* mutations and variants in a culture collection reveal that three amino acid residues (R60, R67, and H173) in NopP are required for *Rj2*-mediated incompatibility. Complementation of *rj2*-soybean by the *Rj2* allele confers the incompatibility induced by USDA 122-type NopP. In response to incompatible strains, *Rj2*-soybean plants activate defense marker gene *PR-2* and suppress infection thread number at two days after inoculation. These results suggest that *Rj2*-soybeans monitor the specific variants of NopP and reject bradyrhizobial infection via effector-triggered immunity mediated by Rj2 protein[3].

We also investigated the distribution of *Rj2*-genotype in Japanese soybean germplasm. Comparison of predicted Rj2 amino acid sequences in a mini-core collection[4] and further complementation analysis revealed that *Rj2*/*rj2*-genotype is determined by a single amino acid substitution (I [Rj2] or R [rj2]) at position 490 in the protein. SNP genotyping for I490 against 1,583 soybean accessions including wild soybeans revealed that *Rj2*-genotype soybeans are distributed in 5.8% of accession in Japanese soybean germplasms. In Japanese cultivated soybeans, *Rj2*-genotype was detected at relatively higher rate than exotic varieties. In addition, the Rj2 alleles in certain wild soybean accession were confirmed by SNP genotyping and cDNA sequencing, suggesting that *Rj2*-genotype has been maintained in cultivated soybeans in the process of soybean domestication.

References:

Tsukui et al. 2013. Appl. Environ. Microbial. 79:1048-1051.
Yang et al. 2010. Proc. Natl. Acad. Sci. USA. 107:18735-18740.
Sugawara et al. 2018. Nat Commun. 9:3139
Kaga et al. 2012. Breed. Sci. 61: 566-592.

W998: **Soybean Genomics**

Genome-Wide Association Study and Meta-Analyses for Identifying QTL associated with Soybean [*Glycine max* (L.) Merr.] Seed Compositions

Kyujung Van[1], Sungwoo Lee[2], Mikyung Sung[3], Randall Nelson[4], Jonathan LaMantia[5], Rouf Mian[6] and Leah K. McHale[1], (1)The Ohio State University, Columbus, OH, (2)Chungnam National University, Dajeon, South Korea, (3)North Carolina State University, Raleigh, NC, (4)USDA ARS, Urbana, IL, (5)USDA-ARS, Wooster, OH, (6)USDA-ARS, NC State University, Raleigh, NC

Significant efforts have previously been made towards identification and validation of QTL associated with soybean seed composition traits because soybean processors desire cultivars with both elevated seed protein and oil contents as well as specific amino acid and fatty acid profiles. Since protein content is generally negatively correlated with oil content and seed yield, development of cultivars with both high protein and high oil contents is difficult. Meta-analysis was our first approach for identifying and validating QTL by mining from literature for seed contents of protein and oil, as well as seed amino acid and fatty acid compositions. A total of 55 meta-QTL for these seed traits were identified on 6 of the 20 chromosomes. Meta-analysis helped narrower confidence intervals than original QTL and candidate genes were identified within each meta-QTL. As the second approach, we conducted a genome-wide association study (GWAS) using phenotypes of 621 soybean accessions collected from five environments and 34,014 SNP markers. Three and five genomic regions significantly associated with seed protein and oil contents, respectively, were identified. Our GWAS approach reconfirmed that QTL on chromosomes 15 and 20 associated with seed protein and oil contents both exhibited a negative relation between the two traits. A multi-trait mixed model allowed to identify “specific effect” loci on chromosome 5 that increased oil with no effect on protein and chromosome 10 that increased protein with little effect on oil, demonstrating the possibility of reducing the negative relationship between protein and oil. Frequencies of positive effect haplotypes from chromosomes 5, 10, 15 and 20 for protein and oil in germplasm across maturity groups and geographic regions can be utilized for improvement of these seed traits in the specific geographic regions.

W999: **Soybean Genomics**

GmBRC1 Is a Candidate Gene for Branching in Soybean

Sangrea Shim, Seoul National University, Seoul, South Korea

The number of branches is one of the important factors affecting the yield of soybean (*Glycine max* (L.)). So far, a dozen of genetic locus associated with branch number has been identified and reported by five quantitative trait locus (QTL) studies. Among these, a recent study has been identified one major QTL referred as *qBR6-1* spanning 450 kb on chromosome 6 and promising candidate gene *BRANCHED1* (*BRC1*)

which regulates axillary bud outgrowth in Arabidopsis. In this study, a genome-wide association analysis combined with linkage analysis were conducted to identify a candidate gene controlling soybean branching. A total five quantitative trait nucleotides (QTNs) were associated with branch numbers in a soybean core collection consisting of 400 soybean accessions. Among these, a linkage disequilibrium (LD) block of *qtnBR6-1* harboring 20 genes was overlapped with a previously identified major QTL *qBR6-1*. To validate and narrow down *qtnBR6-1*, a set of near-isogenic lines (NILs) harboring high-branching (HB) and low-branching (LB) alleles of *qBR6-1* was developed. This NILs showed 99.96% isogenicity and significant difference in branch numbers. A SNP cluster segregating between NIL-HB and NIL-LB was located in the LD block of *qtnBR6-1*. Among the five genes displaying differential expression between NIL-HB and NIL-LB, *BRC1* was down-regulated in shoot apex of NIL-HB and one missense mutation and two upstream SNPs of *BRC1* were tightly associated with branch numbers in additional 59 USDA soybean accessions. Based on these results, we propose the *BRC1* encoding TEOSINTE-BRANCHED1/CYCLODEA/PROLIFERATING CELL FACTOR 1 and 2 (TCP) transcription factor and regulating axillary bud outgrowth in Arabidopsis as a candidate gene for branching in soybean.

W1**000: Soybean Genomics**

Soy GenoMAGIC™, a Novel Solution to Describe and Manage Genomic Variation for High Resolution Genotyping and Gene Discovery

Paul Chomet, NRGene, Ness-Ziona, Israel

Next Generation sequencing technologies have opened the door to multiple genome analyses and an increased understanding of the variations present in populations. To date, most of the germplasm analyses have relied on the comparison of sequence reads to one reference genome assembly, limiting our understanding of genomic variation. NRGene has developed novel analytics and approaches to efficiently perform denovo-assemblies and to describe the relevant variation across germplasm using sequence-based haplotypes. We have applied this system to organize and describe breeding germplasm variation in a number of species including soy. This system efficiently calls genotype data from low pass sequencing reads, integrates legacy data, as well as imputes genotype data with precision and accuracy. This talk will describe the GenoMAGIC™ system using data from soy.

W1**001: Statistical Genomics**

Assessing the Performance of Deep Neural Networks on Genomic Prediction of Body Weight in Broilers

Tiago L. Passafaro[1], Fernando B. Lopes[1], Joao R. R. Dorea[1], Vivian Breen[2], Rachel J. Hawken[2] and **Guilherme J. M. Rosa**[3], (1)UW-Madison, Madison, WI, (2)Cobb Vantress Inc, Siloam Springs, AR, (3)University of Wisconsin-Madison, Madison, WI

The superiority of deep neural network (DNN) in genomic selection (GS) is not totally clear when compared to traditional regression approaches, with results seeming highy dependent on trait and species of application. The relatively small datasets generally used for such comparions could detrimentally affect the performance of DNN in the GS context. Therefore, this study aimed to: 1) Compare the predictive accuracy of DNN with Bayesian Ridge Regression (BRR) and Bayes C_{Π} in predicting genomic breeding values for body weights in broilers, and 2) Assess the effect of dataset size on the performance of each model. Data were provided by Cobb-Vantress Inc., and included phenotypic and genotypic information on 33,015 broilers. Models were compared using the predictive accuracy in 2-fold cross validation, and six sizes of datasets were created to train each model: 5, 20, 35, 50, 70 and 100% of the entire training set size. DNN and Bayesian models were implemented using *TensorFlow* and the R package *BGLR*, respectively. DNN had the highest predictive accuracy using up to 20% of the training set, and all models showed similar performance above this amount of data. DNN resulted in a greater predictive accuracy of 34 and 7% when using 5 and 20% of the training set size, and a lower predictive accuracy of 2, 4, and 3% when considering 50, 70, and 100% of the training set size relatively to the best Bayesian model, Bayes C_{Π}. These results suggest that DNN may perform better than benchmark GS Bayesian regression models in predicting body weight in broilers, but that its superiority vanishes with increased sample sizes.

W1**002: Statistical Genomics**

Probabilistic Egger Regression for Two Sample Mendelian Randomization Analysis in Genome-Wide Association Studies

Xiang Zhou, University of Michigan, Ann Arbor, MI

W1**003: Statistical Genomics**

Good Learners, Faster Learning

Alencar Xavier, Corteva Agrisciences, Johnston, IA and Corteva Biostatistics

New sources of data are being incorporated into breeding pipelines. Intelligent decision-making relies on extracting useful information from data to achieve our goals more efficiently. The breeding success depends on machines able to learn target patterns from data. The search for good learners is based on well stated problems, but evaluation metrics do not always mimic the actual prediction scenario. Another concern is practicality, as the adoption of machines also relies on efficient implementations. Simplifications, such as approximations and conditioning, can lead to fast learners. This presentation provides an insight about evaluation metrics and benefits of simplified algorithm.

W1**004: Statistical Genomics**

Statistical and Computational Methods for Analyzing Chromatin Spatial Organization Data

Wenxiu Ma, UC Riverside, Riverside, CA

W1**005: Statistical Genomics**

A Tutorial of Statistical Power in Genome-Wide Association Studies

Shizhong Xu, Department of Botany & Plant Sciences, University of California, Riverside, CA

Power calculation prior to a genetic experiment can help investigators choose the optimal sample size to detect a quantitative trait locus (QTL). Without the guidance of power analysis, an experiment may be under powered or over powered. Either way will lead to wasted resource in terms of fund and time. QTL mapping and genome-wide association studies (GWAS) are often conducted using a linear mixed model (LMM) with a polygenic background control through a marker inferred kinship matrix. Power analysis for such a mixed model is often conducted via Monte

Carlo simulations. In this study, we derived a non-centrality parameter for the Wald test statistic for association, which allows analytical power analysis. We show that large samples are not necessary to detect a biologically meaningful QTL, say explaining 5% of the phenotypic variance. Several R functions are provided so that users can perform power analysis to determine the minimum sample size required to detect a given QTL with a certain statistical power or calculate the statistical power with a known sample size and other population parameters.

W1**006:** **Strawberry Genomics**

Towards the Strawberry Pan-Genome; Current Status and Future Challenges

Patrick Edger, Michigan State University, EAST LANSING, MI

Due to recent advancements in sequencing platforms and bioinformatic tools, the construction of a pan-genome for diploid and polyploid strawberry species has become feasible. Pan-genome construction involves the comparison of genomes from different genotypes to identify the core genome (genes shared by all genotypes) and dispensable genome (genes found in a subset of genotypes) for each species. Pan-genomes will serve as a powerful platform to the broader research community to dissect the underlying genetic architecture that encodes various traits. The current status and future directions towards this goal will be presented with the goal to initiate conversation and collaboration across the strawberry community.

W1**007:** **Strawberry Genomics**

Genome Analysis of European Wild Strawberries

Timo Hytönen, University of Helsinki, Helsinki, Finland

W1**008:** **Strawberry Genomics**

Strawberry Genome Sequencing

Richard J. Harrison, NIAB EMR, East Malling, United Kingdom

W1**009:** **Strawberry Genomics**

Cross-Talks among Plant Hormone Coordinate Receptacle Fruit Growth and Ripening in Diploid Strawberry *Fragaria vesca*

Chizuko Yamamuro, Fijian Agriculture and Forestry University, Fuzhou, China

Xiong Liao[1, 2], **Mengsi Li**[2], **Miaoling Yan**[2] **Chizuko Yamamuro**[2], [1]College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou, China, [2]FAFU-UCR Joint Center for Horticultural Biology and Metabolomics, Haixia Institute of Science and Technology, Fijian Agriculture and Forestry University, Fuzhou, Fujian, China

Fruit growth and ripening are coordinated to determine the final fruit size and are modulated by multiple phytohormones. How these hormones coordinate and interact with each other to control these processes at the molecular level is not clear. In the early stages of *Fragaria vesca* fruit development, auxin increases both widths and lengths of receptacle fruits, while gibberellin (GA) mainly promotes their longitudinal elongation. We showed that auxin promoted GA biosynthesis and signaling by activating GA biosynthetic and signaling genes, suggesting auxin function is partially dependent on GA function. At the onset of fruit ripening, both auxin and GA levels decreased, leading to a steep increase in the endogenous level of ABA that drives receptacle fruit ripening. I will discuss how we use strawberry receptacle fruits as a model to understand cross-talk among plant hormones.

W1**010:** **Strawberry Genomics**

Mapping Strawberry Flavor Genes

Chris Barbey, University of Florida, Gainesville, FL

W1**011:** **Strawberry Genomics**

Natural Variation in Fruit Color Among *Fragaria* Species Explained by Independent Mutations in a Single Transcriptional Factor: *MYB10*

Cristina Castillejo[1], Julie Caruana[2], Veronica Waurich[3,4], Henning Wagner[3,4], Rubén Ramos[1], José Vallarino[5], Nicolás Oiza[1], Pilar Muñoz-del Rio[1], Juan C. Triviño[6], Sonia Osorio[5], Zhongchi Liu[2], David Posé[5], Tuomas Toivainen[7], Timo Hytönen[7], José F. Sánchez-Sevilla[1], Klaus Olbrich[3] and **Iraida Amaya**[1], (1)IFAPA Centro de Málaga, Málaga, Spain, (2)Dept. of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, (3)Hansabred GmbH & Co, Dresden, Germany, (4)Institut für Botanik, Dresden, Germany, (5)IHSM - University of Málaga - CSIC, Málaga, Spain, (6)Sistemas Genómicos, Valencia, Spain, (7)University of Helsinki, Helsinki, Finland

Anthocyanins are the pigments responsible for the red color of strawberries. Their biosynthesis is controlled at the transcriptional level by a ternary complex consisting of R2R3-MYB and bHLH transcription factors associated with a WD40-repeat protein.

In order to map the genetic factors involved in fruit coloring we generated a mapping population crossing a *F. vesca* accession bearing white fruits (ESP138.596) with the red-fruited 'Reine des Vallées'. DNA from white- or red-fruited F2 individuals was pooled to perform a bulk segregant analysis (BSA) linked with high-throughput genome sequencing. This analysis revealed the presence of a *gypsy*-like retrotransposon inserted in the third exon of *FvMYB10.* The presence of this retroelement in homozygosis co-segregated with white fruits in the complete F2 population. We further extended this analysis to other white-fruited *F. vesca* accessions but none of them harbored this retroelement in *FvMYB10*. Instead we identified two additional polymorphisms affecting *FvMYB10*, (1) a single nucleotide insertion, which generates a truncated protein, and (2) a large deletion of ~100 Kb spanning a genomic region that contains 7 genes, one of them being *FvMYB10.* The three newly identified polymorphisms on *FvMYB10* differ from the previously described single nucleotide mutation, responsible for the lack of anthocyanins in other *F. vesca* white/yellow fruited accessions[1].

We next analyzed QTL for fruit color in a segregating population derived from the red-fruited *F.* x *ananassa* 'Senga Sengana' and a *F. chiloensis* accession with white flesh. A major QTL controlling 45.7 - 54.7% of variance in internal flesh color was detected on LG I-3. The confidence interval spans the orthologous region where *FvMYB10* is located. Furthermore, transient overexpression on *FvMYB10* on different *F. chiloensis* accessions resulted in red sectors both in the epidermis and fruit flesh.
Altogether, these results show that a single R2R3-MYB, *FvMYB10*, regulates fruit color in different *Fragaria* species, indicating convergent evolution for anthocyanin biosynthesis in strawberry fruit.
[1]Hawkins, C., et al. (2016). Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Sci. Rep.* 6, 29017.
This work was supported by grants RFP2015-00011-00-00 (Spanish Ministry of Economy and Competitivity and FEDER), EI.AVA.AVA201601.10 (IFAPA, FEDER funds) and by the European Union's Horizon 2020 research and innovation programme (GoodBerry; grant agreement 714 number 679303). This study is also part of the joint research network SPIRED which was funded by the German Federal Ministry of Education and Research (BMBF, FKZ 031A216A and B).

W1**012:** **Strawberry Genomics**

New Strawberry QTLs for Pleasant Key Aroma Compounds in F1 Population

Amparo Monfort, IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UB-UAB, Barcelona, Spain and Pol Rey, IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Bellaterra, Spain

Strawberry is a plant with vegetative reproduction whose fruits are highly appreciated due to its organoleptic properties and health benefits. Cultivated strawberry (*Fragaria x ananassa*) is an allo-octoploid species with a fairly small genome, mostly diploidized. Wild strawberry (*F. vesca)* is a diploid species with small and soft, but very aromatic, fruits. The aroma of cultivated strawberry has been widely studied: the profile of volatiles obtained by gas chromatography and mass spechtrometry (GC-MS) is above 350 compounds, in particular esters, terpenes, furans, lactones, alcohols, aldehydes and sulphur compounds. Analysis of the volatile compounds and the threshold of aroma detection indicate that only 20 compounds significantly contribute to the aroma of strawberry with esters being the most representative, which give fruity aromas. Other important compounds are Z-3 and E2-hexenal which give an unpleasant smell (green aroma) detected in some cultivars. A recent analysis using GC-MS for the lines *F. vesca* Reina de valles and *F. vesca* Yellow Wonder and a collection of NILs (Near Isogenic Lines) in strawberry established the volatilome and the QTL map of wild strawberry for all the compounds (Urrutia *et al* 2017). To determine the genomic regions responsible for the variation in aromatic compounds of cultivated strawberry, we generated a segregating F1 population of the cross between two lines segregating for aroma in wild strawberry, "Dream x Starlette" (70 individuals). This population was genotyped with the IStraw35K® array giving a map of 28 linkage groups (14,335 SNPs, 3,167 loci) covering >92% of the genome compared with the consensus map. Volatile analyses (GC-MS) of the population mature fruit, measured over a period of three years and at different stages of harvest, detected more than 300 segregating compounds. Analyses of the 19 specific compounds responsible for strawberry aroma revealed 62 QTLs, 14 of them highly significant in all of the analyses. We specifically detected a single QTL which regulates accumulation of the esters methyl hexanoate and ethyl hexanoate, both responsible for a fruity aroma in strawberry, in a region limited by LG4B. New QTLs will be check for candidate genes, which genes are in these regions and select some of these as candidate genes. All these new QTLs should be check in different populations.

W1**013:** **Strawberry Genomics**

Efficient Genome-Editing of Wild Strawberry Genes, Vector Development, and Validation

Junhui Zhou, UNIVERSITY OF MARYLAND-CP, COLLEGE PARK, MD

The CRISPR-Cas9 system is an effective genome editing tool for plant and animal genomes. However, there are still few reports on the successful application of CRISPR–Cas9 to horticultural plants, especially regarding germ-line transmission of targeted mutations. We developed a high-efficiency genome editing system in the wild strawberry *Fragaria vesca* and demonstrated its successful application to mutate the auxin biosynthesis gene TAA1 and auxin response factor 8 (ARF8). In our system, both *Arabidopsis* U6 promoter AtU6-26 and the wild strawberry U6 promoter FveU6-2 were shown to lead to high-efficiency genome editing in strawberry. The progeny of the primary (T0) transgenic plants carrying the CRISPR construct was shown to harbor both germline transmitted mutations as well as new mutations, including large deletions between the two sgRNA-targeted sites. Both UBQ promoter and YAO promoter driven CAS9 show high efficient genome editing. Further, T1 seedlings harboring *arf8* homozygous knockout mutations grew considerably faster than wild-type seedlings. The floral receptacle of *arf8* homologous mutants are also larger than that of wild type. Both seedling and receptacle phenotype of *arf8* suggested a negative role of ARF8 for growth. Our results open up exciting opportunities for engineering strawberry and related horticultural crops to improve traits of economic importance.

W1**014:** **Sugar Beet Workshop**

Sugar Beet Stress Germination Transcriptomes. First Impressions Indicate High Complexity Expression

Mitch McGrath, USDA-ARS Sugarbeet & Bean Research Unit, East Lansing, MI

W1**015:** **Sugar Beet Workshop**

Two Constans-like Genes Jointly Control Flowering Time in Beet

Christian Jung[1], Nadine Dally[1], Nadine Höft[1], Maike Eckel[2] and Alfred Batschauer[2], (1)Christian Albrechts University of Kiel, Kiel, Germany, (2)Department of Plant Physiology and Photobiology, Faculty of Biology, Philipps-University, Marburg, Germany

The biennial species sugar beet makes shoot elongation ('bolting') followed by flowering after a long period of cold temperatures. Flowering is controlled by a regulatory pathway with four major components: *BTC1* (locus *B*) and *BvBBX19* (*B2*) are upstream regulators of *BvFT1* and *BvFT2*, two orthologs of the Arabidopsis gene *flowering LOCUS T* (*FT*). While *BvFT2* is a floral inducer such as *FT* in Arabidopsis, *BvFT1* represses floral transition. We present new data how transcription factors *BTC1* and *BvBBX19* jointly control the expression of their downstream targets. We detected an epistatic interaction between both genes because F_2 plants homozygous for two *B/B2* mutant alleles did not bolt even after vernalization. Fluorescence complementation studies revealed that both proteins form a heterodimer *in vivo*. In non-bolting plants, the

bolting activator *BvFT2* was completely downregulated whereas the repressor *BvFT1* was upregulated which suggests that both genes acquire a *CONSTANS* (*CO*) like function in beet. Like CO, B and B2 proteins house CCT and BBX domains which, in contrast to CO are split between the two beet genes. We propose an alternative regulation of *FT* orthologs in beet that can be exploited to breed winter beets. We also screened a panel of Beta accessions including wild species from different geographical regions for sequence variations within the 4 bolting time genes. We found SNPs correlated with phenological development in response to daylength which sheds light on the evolution of cultivated and wild species of the genus Beta.

W1**016:** **Sugar Beet Workshop**

Discovery of Genetic Diversity of Interest By Comparing a Progeny from (sugar beet elite x exotic) Crosses with a Sugar Beet Elite Panel

Prune Pegot-Espagnet[1,2], Olivier Guillaume[1], Bruno Desprez[2], Brigitte Devaux[2], Pierre J. Devaux[2], Karine Bounan Henry[2], Nicolas Henry[2], Glenda Willems[3], Ellen Goudemand[2] and Brigitte Mangin[1], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France, (2)Florimond Desprez, Cappelle-en-Pévèle, France, (3)SESVANDERHAVE N.V., Tienen, Belgium

The selection of new stable sugar beet varieties producing more extractable sugar per hectare is the main target for sugar beet breeding. However the genetic variability useful for genetic improvement of crop is increasingly narrow due to successive inbreeding crosses. It is therefore necessary to look for new and interesting genetic diversity to enrich sugar beet breeding material.

AKER is a research program funding by the French Government that aims to improve the competitiveness of sugar beet by 2020. This program allowed the identification of 16 exotic accessions that cover 100% of the allelic variability available that is not already present in elite germplasm. Through the comparison of association studies done on an elite panel and an (elite x exotic) progeny, this study shows that introducing exotic germplasm into breeding programs can bring new interesting allelic diversity.

The two populations were phenotyped for three traits: potassium quantity (K meq/100g of fresh material), sodium quantity (Na meq/100g of fresh material) and N-alpha-amino quantity (N meq/100g of fresh material). The less we have impurities in a sugar beet, the more the white sugar can be extracted. Genome Wide association studies (GWAS) were performed in each population for each trait with the multi-locus mixed model approach (MLMM)[1], and the most parsimonious model was selected thanks to the eBIC criterion[2]. SNPs were then merged into quantitative trait loci (QTLs) in each population. A QTL is defined as a single SNP associated with traits or as a group of SNPs associated with traits of interest, located on the same chromosome with a maximum of 5cM between two consecutive SNPs, and with linkage disequilibrium greater than a significante threshold calculated for each population.

Results of association studies showed a common trait genetic architecture between the two populations, e.g. a QTL for Na trait located at the middle of the chromosome 5 found in both populations. However differences were also identified, i.e. some QTLs were only detected in one population. Moreover some QTLs in the progeny have a favorable effect for the exotic allele, e.g. the amount of N impurity decreases in homozygotes for the exotic allele with a QTL located at the end the chromosome 5. These kind of genomic regions can be interesting to add in sugar beet breeding programs to increase the useful genetic variability.

This Research program is funding by the French Government, under the management of the Research National Agency (ANR). AKER is included in the « Investissements d'avenir » as refered as ANR-11-BTBR-0007. This study would not have been possible without the grant of the ANRT.

References

[1]Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., and Nordborg, M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics, 44(7):825–830, 2012.

[2]Chen, J. and Chen, Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika, 95(3):759–771, 2008.

W1**017:** **Sugar Beet Workshop**

Mitochondrial Proteomics and Transcriptional Analysis of Cytoplasmic Male Sterility in Sugar Beet using iTRAQ and qRT-PCR

Dayou Cheng, Harbin Institute of Technology, Harbin, China

W1**018:** **Sugar Beet Workshop**

Comparative Proteomic Analysis of Two Sugar Beet Phenotype with Contrasting Salt Tolerance by iTRAQ

Gui Geng, Heilongjiang University, Harbin, China

W1**019:** **Sugar Beet Workshop**

SNP-Assisted Selection to Reduce Rhizomania Virus-Content in Sugar Beet

Claudia Chiodi, University of Padova, Legnaro (Padova), Italy

W1**020:** **Sugar Beet Workshop**

SNP Alleles Associated with Low Bolting Tendency in Sugar Beet

Samathmika Ravi, University of Padova, Legnaro (PD), Italy

W1**021:** **Sugar Cane (ICSB)**

Introductory Remarks

Nathalie Piperidis, SUGAR RESEARCH AUSTRALIA, MACKAY, Australia

W1**022:** **Sugar Cane (ICSB)**

Allele Defined Genome of the Autopolyploid *Saccharum spontaneum L*.

Jisen Zhang, Center for Genomics and Biotechnology, Fujian Provincial Key laboratory of Haixia applied plant systems biology, Haixia Institute of Science and Technology,Fujian Agriculture and Forestry University, Fujian, China, Xingtan Zhang, FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, China, Haibao Tang, Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, China, Qing Zhang, Center for Genomics and Biotechnology, Fujian Provincial Key laboratory of Haixia applied plant systems biology, Haixia Institute of Science and Technology,Fujian Agriculture and Forestry University, Fuzhou, China, Chifumi Nagai, Hawaii Agriculture Research Center, Kunia, HI and **Ray Ming**, University of Illinois at Urbana-Champaign, Urbana, IL

Sugarcane is one of the most important first generation biofuel crops and contributes to about 70 % sugar production worldwide. However, lack of genome sequences hinder sugarcane genomic research and crop improvement. Sugarcane genome is much more complex than any other crops due to recent polyploidization and large genome size. The PacBio Single Molecular Real-Time sequencing and chromosomal confirmation capture (Hi-C) technologies offer the opportunity to improved contig continuity and generate chromosomal level assembly. To reduce the complexity of genome sequencing and assembly, a haploid clone of *Saccharum spontaneum*, AP85-441 (2n=4x=32), was selected for sequencing. The overlap-based assembly algorithm (Canu program) resulted in a ~ 3.13 Gb genome assembly, covering ~ 98 % of the estimated genome size. Hi-C mapping anchored ~ 93 % of sequences into 32 chromosomes. Comparison of sorghum and AP85-441 genomes exhibited extensive collinearity and revealed chromosomal rearrangements specific in *S. spontaneum*. This genomic resource will be applied to dissect the integrated *S. spontaneum* genomic sequences in leading sugarcane hybrid cultivars.

W1**023:** **Sugar Cane (ICSB)**

Allelic Organization for Sugar Accumulation in *Saccharum* Hybrid

Melina C. Mancini[1], Danilo A. Sforça[1], Fan Zhu[2], Claudio Benicio Cardoso- Silva[3], Ray Ming[4] and Anete P Souza[5], (1)University of Campinas - UNICAMP, Campinas, Brazil, (2)University of Illinois at Urbana-Champaign, Urbana, IL, (3)University of Campinas- UNICAMP, Campinas, Brazil, (4)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, China, (5)University of Campinas, Campinas, Brazil

Keywords: Poliploid, BAC library, Sorghum bicolor

Modern sugarcane cultivars are allopolyploids and result from artificial crosses between sweet *S. officinarum* L. (2n = 8x = 80) and wild *S. spontaneum* L. (2n = 4x–16x = 32–128), followed by a several backcrosses with *S. officinarum*. The *Saccharum* hybrid shows chromosome number ranging from 80 to 135; highly ploidy level; aneuploidy condition; high amount of repetitive sequences; and genome size around 10 Gb. Sugarcane genome studies uses the synteny with sorghum to explain the sugarcane genome. Genes involved in sugar accumulation were chose to discover the allele origin of the sugarcane hybrid concerning the *S. officinarum* and *S. spontaneum* genome to observe the allelic behavior. The genes related to sugar accumulation selected on SP80-3280 BAC clones through the approach "Targeted Sequencing by Gene Synteny" and it was performed a BLASTn against the ancestors genes and phylogenetic trees were build using Neighbor-Joining method. The *Saccharum* hybrid gene sequences clearly separated from the ancestors species, resulting in 70% alleles from *S. officinarum*; 23% alleles from S. spontaneum; and 7% alleles from recombinant between *S. officinarum* and *S. spontaneum*, according with cytogenetic studies in sugarcane. No gene sequence recovered all alleles, but at least one allele per gene was found duplicated. The allele-dosage effects represent a big challenge for sugarcane genetic mapping resulting in several bias reflecting in recombination fraction. Our findings for evolutionary relationships showed the genes present on hybrid is closer to *S. officinarum*, also appointed the main limitation cause for duplicated alleles on linked mapping.

Funding: FAPESP-PD fellowships 2014/11482-9 and FAPESP-BEPE fellowship 2017/05014-0.

W1**024:** **Sugar Cane (ICSB)**

Application of ALLHiC on Scaffolding of an Autopolyploid Sugarcane Genome based on Hi-C Data

Xingtan Zhang, FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, Fujian, China, Ray Ming, University of Illinois at Urbana-Champaign, Urbana, IL and Haibao Tang, Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, China

Construction of chromosome-level assembly is a vitally important process to achieve the goal of 'Platinum' genome, but it remains a great challenge to anchor sequences to chromosomes in polyploid or highly heterozygous genomes. High throughput chromosome conformation capture (Hi-C) technology serves as a robust tool to dramatically advance chromosome scaffolding, however, existing approaches are mostly designed for diploid genomes often with the aim of reconstructing a haploid representation, and have limited power to construct chromosomes for polyploid genomes. To solve the problem of Hi-C scaffolding in complex genomes, we introduce a novel algorithm (ALLHiC) that provides an allele-aware assembly method and integrates genetic algorithm (GA) to optimize the ordering and orientation of contigs using Hi-C paired-end reads. Application on simulated genome data reveals that ALLHiC has significant effect to phase allelic contigs and improves ordering and orientation comparing to existing Hi-C assemblers. We also tested ALLHiC on an auto-tetraploid sugarcane genome with 8 sets of homologous chromosomes, and successfully constructed the phased chromosomal level assemblies. We demonstrate that ALLHiC pipeline is an effective tool for *de novo* scaffolding of polyploid as well as an alternative program to improve the ordering and orientation for diploid genomes with similar performance compared to existing approaches.

W1**025:** **Sugar Cane (ICSB)**

Analysis of Three RGAs from Sugarcane in Recently Published Genomes

Marie-Anne Van Sluys, Universidade de São Paulo, Sao Paulo, Brazil

W1**026:** **Sugar Cane (ICSB)**

Comparative Analysis of Homologous Sequences of *Saccharum officinarum* and *Saccharum spontaneum* Reveals Independent Polyploidization Events

Qingyi Yu, Texas A&M AgriLife Research, Dallas, TX, Anupma Sharma, Texas A&M AgriLife Research Dallas, Dallas, TX and Ray Ming, University of Illinois at Urbana-Champaign, Urbana, IL
Sugarcane is an economically important crop widely grown in tropical and subtropical regions for sugar and energy production. Modern sugarcane is derived from interspecific hybridization between *S. officinarum* and *S. spontaneum*. We constructed BAC libraries of *S. officinarum* variety LA Purple (2n=8x=80) and *S. spontaneum* haploid clone AP85-441 (2n=4x=32), and selected and sequenced 97 BAC clones from the two libraries. A total of 5,847,280 bp sequence from *S. officinarum* and 5,011,570 bp from *S. spontaneum* were assembled and 749 gene models were annotated. A relatively higher gene density and lower repeat content were observed in *S. spontaneum* than in *S. officinarum*. Comparative analysis of syntenic regions revealed a high degree of collinearity in genic regions between *Saccharum* and sorghum and between *S. officinarum* and *S. spontaneum*. *S. spontaneum* showed expansion relative to *S. officinarum*, and both *S. officinarum* and *S. spontaneum* showed expansion relative to sorghum. Among the 75 full-length LTR retrotransposons identified in the *Saccharum* BACs, none of them are older than 2.6 mys and no full-length LTR elements are shared between *S. officinarum* and *S. spontaneum*. In addition, divergence time estimated using a LTR junction marker and a syntenic gene shared by 3 *S. officinarum* and 1 *S. spontaneum* BACs revealed that the *S. spontaneum* intergenic region was distant to those from the 3 homologous regions in *S. officinarum*. Our results suggested that *S. officinarum* and *S. spontaneum* experienced at least two rounds of independent polyploidization in each lineage after their divergence from a common ancestor.

W1**027:** **Sugar Cane (ICSB)**
Morning Intermission
Nathalie Piperidis, SUGAR RESEARCH AUSTRALIA, MACKAY, Australia

W1**028:** **Sugar Cane (ICSB)**
The Impact of Normalization on the Composition of the Transcriptome in Polyploid Sugarcane
Van Nam Hoang, College of Agriculture and Forestry, Hue, Viet Nam, Agnelo Furtado, University of Queensland/QAAFI, Brisbane, Australia, Virginie Perlo, University of Queensland, Brisbane, Australia, Frederik C Botha, Sugar Research Australia Limited, Indooroopilly, Australia and **Robert J. Henry**, University of Queensland/QAAFI, Brisbane, QLD, Australia
cDNA normalization is used to improve the coverage of rare transcripts in analysis of transcriptomes employing next-generation sequencing. Long-read technology has become a powerful tool for sequencing and construction of transcriptomes, especially for complex genomes containing highly similar transcripts and transcript-spliced isoforms. We analyzed the transcriptome of sugarcane, with a highly polyploidy plant genome, by PacBio isoform sequencing (Iso-Seq) using two different cDNA library preparations, with and without a normalization step. The results demonstrated that, while the two libraries included many of the same transcripts, many longer transcripts were removed and many new generally shorter transcripts were detected by normalization. For the same input cDNA and the same data yield, the normalized library recovered more total transcript isoforms, number of predicted gene families and orthologous groups, resulting in a higher representation for the sugarcane transcriptome, compared to the non-normalized library. Functional annotation of the unique transcripts suggested that each library enriched different functional transcript fractions. This demonstrated the complementation of the two approaches in obtaining a complete transcriptome of the sugarcane genome at the sequencing depth used in this study.

W1**029:** **Sugar Cane (ICSB)**
Increasing Genetic Gain in Sugarcane using Genomic Selection
Kai P Voss-Fels, The University of Queensland, St Lucia, QLD, Australia, Phil Jackson, CSIRO, Brisbane, QLD, Australia, Xianming Wei, Sugar Research Australia, Mackay, Australia, Karen Aitken, CSIRO Agriculture and Food, St Lucia, Australia, Elizabeth Ross, QAAFI - University of Queensland, Brisbane, Australia, Mark Cooper, The University of Queensland, Brisbane, QLD, Australia and Ben J. Hayes, Department of Environment and Primary Industries, Bundoora, Australia
Breeding progress in sugarcane is constrained by the long breeding cycle length and the contribution of non-additive genetic components to many agronomically important traits. Genomic selection (GS) has led to significant genetic improvements in many crop species and it holds the potential for increasing genetic gain in sugarcane as well. Reducing the cycle length and improving the accuracy at which genetically superior genotypes can be selected for crossings are two improvements that could be achieved through the implementation of GS. At the same time, practical experience and empirical evidence for the most optimal integration of GS in current sugarcane breeding programs is very limited. Based on high quality phenotype and genotype data from current Australian sugarcane breeding programs we are able to demonstrate the potential of GS for predicting clone performance for several important traits. Furthermore, using computer simulations developed in close cooperation with sugarcane breeders we propose alternative GS-based breeding strategies which could help to significantly increase rates of genetic improvement in the future.

W1**030:** **Sugar Cane (ICSB)**
From Zhongzhe Cultivar Genomic Selection to Trait Studies based on Genetic Population
Wei Yao, Guangxi University,China, Nanning, China

W1**031:** **Sugar Cane (ICSB)**
Differential Gene Expression Analysis of Flower Development in Ravenna Grass (*Tripidium ravennae*)
Nathan A. Maren, North Carolina State University, Raleigh, NC

W1**032:** **Sweetpotato Genomics**
Genome Sequences of Two Diploid Wild Relatives of Cultivated Sweetpotato
Shan Wu[1], Kin H Lau[2], Qinghe Cao[3], John Paul Hamilton[4], Honghe Sun[5], Chenxi Zhou[6], Lauren Eserman[7], Dorcus C. Gemenet[8], Bode Olukolu[9], Haiyan Wang[4], Emily Crisovan[2], Grant Godden[4], Chen Jiao[1], Xin Wang[1], Mercy N. Kitavi[10], Norma C. Manrique

Carpintero[4], Brieanne Vaillancourt[11], Krystle Wiegert-Rininger[4], Xinsun Yang[12], Kan Bao[1], Jenn E. Schaff[13], Jan Kreuze[14], Wolfgang Gruneberg[15], Awais Khan[15], Marc Ghislain[10], Daifu Ma[16], Jiming Jiang[4], Robert O.M. Mwanga[17], Jim Leebens-Mack[7], Lachlan J Coin[18], Craig Yencho[13], C. Robin Buell[2] and Zhangjun Fei[1], (1)Boyce Thompson Institute, Cornell University, Ithaca, NY, (2)Department of Plant Biology, Michigan State University, East Lansing, MI, (3)Sweetpotato Research Institute, CAAS, Xuzhou, China, (4)Michigan State University, East Lansing, MI, (5)National Engineering Research Center for Vegetables, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, (6)Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia, (7)University of Georgia, Athens, GA, (8)International Potato Center (CIP), Lima 12, Peru, (9)North Carolina State University, Knoxville, TN, (10)International Potato Center (CIP), Nairobi, Kenya, (11)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, (12)Hubei Academy of Agricultural Sciences, Wuhan, China, (13)North Carolina State University, Raleigh, NC, (14)International Potato Center, Lima, Peru, (15)International Potato Center (CIP), Lima, Peru, (16)Jiangsu Xuzhou Sweetpotato Research Center, Xuzhou, China, (17)International Potato Center (CIP), Kampala, Uganda, (18)Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia

Sweetpotato [*Ipomoea batatas* (L.) Lam.] is a globally important staple food crop, especially for sub-Saharan Africa. Agronomic improvement of sweetpotato has lagged behind other major food crops due to a lack of genomic and genetic resources and inherent challenges in breeding a heterozygous, clonally propagated polyploid. Here, we report the genome sequences of its two diploid relatives, *I. trifida* and *I. triloba*, and show that these high-quality genome assemblies are robust references for hexaploid sweetpotato. While *I. trifida* is the more closely related diploid to hexaploid sweetpotato, and is a better reference for sweetpotato, the *I. triloba* genome can be a complementary reference. Comparative analyses revealed insights into the history of ancient whole genome triplication (WGT) of *Ipomoea*, which has contributed to additional copies of genes involved in storage root development. We then identified gene families specifically expanded in *I. trifida* and/or *I. triloba*, significantly enriched with those associated with adaptation and stress resistance, as well as storage root development in the hexaploid sweetpotato such as sporamins. Using resequencing data from 16 genotypes widely used in African breeding programs, genes and alleles associated with carotenoid biosynthesis in storage roots are identified, which may enable efficient breeding of varieties with high provitamin A content. We also discovered aneuploidy in cultivated sweetpotato, which may present an extreme form of structural variation that clearly would affect transcript dosage and consequently, phenotypic variation. These resources will facilitate genome-enabled breeding in this important food security crop.

W1**033:** **Sweetpotato Genomics**

A High-Fidelity Genotype and Allele Dosage Calling Platform, GBSpoly and GBSapp, Elucidates the Genomic Structure of Hexaploid Sweetpotato

Bode Adebowale Olukolu, University of Tennessee, Knoxville, TN

W1**034:** **Sweetpotato Genomics**

Sweetpotato Genome Sequencing Efforts by the TRAS Community

Sachiko Isobe, Kazusa DNA Research Institute, Kisarazu, Japan

W1**035:** **Sweetpotato Genomics**

Transcriptomic Responses of the Sweetpotato Clones Beauregard and Tanzania under Polyethylene Glycol-Simulated Drought Stress

Kin H Lau[1], Maria del Roasrio Herrera[2], Emily Crisovan[1], Shan Wu[3], Zhangjun Fei[3], Awais Khan[4], C. Robin Buell[1] and Dorcus C. Gemenet[2], (1)Department of Plant Biology, Michigan State University, East Lansing, MI, (2)International Potato Center (CIP), Lima, Peru, (3)Boyce Thompson Institute, Cornell University, Ithaca, NY, (4)Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, NY

Cultivation of sweetpotato, like many crops, can be severely impacted by drought. Fortunately, wide variation in drought tolerance among sweetpotato accessions suggests that breeding for sweetpotato with improved drought tolerance is possible. Understanding how drought tolerance is genetically controlled and identifying the genetic loci involved would greatly aid this endeavor. Towards these goals, we profiled the transcriptional response of a US cultivar and a Ugandan landrace, Beauregard and Tanzania respectively, under drought simulated by polyethylene glycol (PEG) treatment. At the two time-points of 24 and 48 hours, between 4,000 to 6,000 genes in leaf tissue were differentially expressed in each cultivar. About half of these genes were shared between the two genotypes and these were significantly enriched for drought-related Gene Ontology (GO) terms. Orthologs of drought response genes studied in model species were identified for the *Ipomoea trifida* reference genome, revealing that less than half of these were drought-responsive under our experimental conditions. Finally, we identified genes differentially regulated between Beauregard and Tanzania, and found co-regulated gene clusters. Some of these clusters may be involved in the two genotypes' differential decreases in chlorophyll content in response to drought, which was reported in work by other researchers. These data have been published and are available for use by breeders and geneticists (doi.org/10.1002/pld3.92).

W1**036:** **Sweetpotato Genomics**

Identification of Quantitative Trait Loci of Storage Root Micronutrients (Iron and Zinc) in Cultivated Sweetpotato

Benard Yada[1], Bonny Michael Oloka[2], Bode Adebowale Olukolu[3], Milton O. Anyanga[1], Doreen Chelangat[1], Paul Musana[1], Agnes Alajo[1], Guilherme Da Silva Pereira[4], Marcelo Mollinari[4], Zhao-Bang Zeng[4] and G. Craig Yencho[2], (1)National Agricultural Research Organization (NARO), Kampala, Uganda, (2)North Carolina State University, RALEIGH, NC, (3)University of Tennessee, Knoxville, TN, (4)North Carolina State University, Raleigh, NC

Development of biofortified sweetpotato [*Ipomoea batatas* (L.) Lam.] is desired for the improvement of the nutritional status and livelihoods of millions of households in sub-Saharan Africa who depend on the crop as a staple. However, sweetpotato improvement for important traits has

been slow over the years largely due to the crop's complex hexaploid (2n = 6x = 90) genetics, its large genome size and a lack of genomic tools to decipher the architecture of its complex traits. We generated a mapping population of 287-F1 progeny (NKB population) derived from a cross between "New Kawogo", an African landrace and "Beauregard", a popular cultivar in USA for QTL analysis of iron and zinc in sweetpotato. The population was phenotyped in the laboratory using a standard near infrared spectroscopy (NIRS) protocol and genotype predicted means were obtained. We also genotyped the population using modified genotyping by sequencing (GBS) protocol for hexaploid sweetpotato (GBSpoly). The genotype means for storage root Fe and Zn content ranged from 1.6–2.1 and 0.9–1.2 µg/100g dry matter, respectively. There were transgressive segregants that performed better than the means of the higher parents for these traits thus presenting a huge potential for genetic gain in future hybridization schemes. We identified quantitative trait loci (QTL) for iron (Fe) and Zinc (Zn) using random effects QTL mapping model implemented in the R-based software '*Polyqtl*', utilizing SNP marker dosage information and the NKB linkage map. A total of three QTL for storage root Zn content were identified on linkage group 6, 13 and 15 explaining 15%, 11% and 5% of the variation, respectively, accounting for a total variation of 32%. One QTL was identified for iron content on linkage group 6 explaining 12% of the variation. Additive effects of the QTL alleles from both parents were estimated and important alleles that impacted the population mean of Fe and Zn content were identified. This work represents a significant step forward in our understanding of the genetic architecture of the micronutrients, Fe and Zn in sweetpotato. This sets the stage for our long-term goal of developing and using marker assisted breeding tools in applied sweetpotato breeding programs.

W1**037:** Sweetpotato Genomics

Genome-Wide Association Study for Continuous Storage Root Formation and Bulking Revealed SNPs Markers Highly Associated with High Yielding Ability in Sweetpotato (*Ipomoea batatas* (L.) Lam)

Robert Mwanga[1], Astere Bararyenya[2], Bode Olukolu[3], Phinehas Tukamuhabwa[4], Wolfgang Gruneberg[5], Wellington Ekaya[6], Jan Low Wayland[7], Mildred Ochwo Ssemakula[4], Nasser Yao[8], Dorcus C. Gemenet[5] and Mercy N. Kitavi[9], (1)International Potato Center (CIP), Kampala, Uganda, (2)Makerere University, Kampala, Uganda, (3)University of Tennessee, Knoxville, TN, (4)Makerere university, Kampala, Uganda, (5)International Potato Center (CIP), Lima, Peru, (6)ILRI, Nairobi, Kenya, (7)International Potato Center, Nairobi, Kenya, (8)BecA-ILRI Hub, Nairobi, Kenya, (9)International Potato Center (CIP), Nairobi, Kenya

Continuous storage root formation and bulking (CSRFAB) in sweetpotato is undisputedly an important trait from agronomic and biological perspectives. Information about the molecular mechanisms underlying storage root formation and development overtime is lacking. Here, as a first step toward understanding the genetic basis for CSRFAB in sweetpotato, we performed a genome-wide association using phenotypic data from four distinct developmental stages and 33,073 SNP and Indel markers. Continuous storage root formation and bulking scores (CSRFABs) at different harvesting times, the area under growth progress curve and the slope of growth curve were used to characterize 358 sweetpotato cultivars comprised of 130 parental genotypes and 228 progenies derived from a cross of 20 of the 130 parental materials. These cultivars were phenotyped for the CSRFAB traits at four different harvesting times over two locations in order to identify their quantitative trait nucleotides (QTNs) using a compressed mixed linear model (CMLM). As a result, 11 and 13 QTNs were found to be associated with CSRFAB and Discontinuous storage root formation and bulking (DCSRFAB), respectively. Associated QTNs were completely different between CSRFAB and DCSRFAB and commonly found at 150DAP and AUGPC suggesting a contrasted gene expression pattern. Candidate genes proximal to or that are co-localized with these SNPs were identified using *I. trifida* reference genome, an ancestral progenitor. Their functional annotations support potential roles underpinning processes such as growth and regulation of growth hormones, leaf senescence and root development. These findings will provide valuable insights in understanding regulatory networks and associated functions to develop strategies for sweetpotato yield improvement

Keywords**:** *sweetpotato, genome-wide association study, quantitative trait nucleotide, continuous storage root formation and bulking, Area under growth progress curve, dynamic development*

W1**038:** Swine

Characterization of CRISPR Knockouts of Pathogen Receptors for Porcine Reproductive and Respiratory Syndrome Virus and Transmissible Gastroenteritis Coronavirus

Raymond R. Rowland, Kansas State University, Manhattan, KS and Randall S Prather, University of Missouri, Columbia, Columbia, MO

Infections caused by porcine reproductive and respiratory syndrome virus (PRRSV) incur significant costs and other hardships to the swine industry. Current live attenuated vaccines offer some economic relief, but suffer from significant drawbacks, including vaccine virus shedding, reversion to virulence, and other side effects. There are no new vaccines on the horizon. One of the functions of CD163 is to serve as a receptor for the PRRS virus on macrophages. CD163 receptor KO pigs were constructed using CRISPR/Cas9 to create a frameshift mutation in exon 7 of *CD163*. The absence of CD163 expression was confirmed using flow cytometry and anti-CD163 antibodies. The CD163 KO pigs are resistant to infection by both PRRSV-1 and PRRSV-2 isolates. Differences in how the two PRRSV genotypes recognize CD163 are found in pigs possessing a domain swap between scavenger receptor cysteine-rich (SRCR) domain 5 (exon 7) and a synthesized exon encoding a homolog of human CD163-like SRCR domain 8. Pigs possessing the domain swap are resistant to PRRSV-1 but not PRRSV-2 isolates. Current research is directed at finding the amino acids in SRCR domain 5 responsible for recognition by PRRSV-1 and PRRSV-2 with the ultimate goal of finding the smallest mutation in CD163 that confers resistance to both PRRSV genotypes. A second target for genetic modification is aminopeptidase N (APN), the receptor for transmissible gastroenteritis virus (TGEV). Along with porcine epidemic diarrhea virus (PEDV), TGEV is placed within the alphacoronavirus group. Both coronaviruses are sources of morbidity and mortality in neonatal pigs, a consequence of infecting enterocytes. The biological relevance of amino peptidase N (APN, ANPEP, CD13) as a receptor for TGEV and PEDV in pigs was tested using CRISPR/Cas9 to edit exon 2 of *ANPEP*. Porcine APN is a 963 amino acid type II membrane metallopeptidase responsible for removing N-terminal amino acids from protein substrates during digestion in the gut. *ANPEP* KO pigs possessing a null APN phenotype did not support infection with TGEV but retained susceptibility to infection with PEDV. Even though TGEV and PEDV belong to the same taxonomic group, target the same cell, and

produce similar clinical signs, the pig infection results confirm the different receptor requirements. Taken all together, the genetic modification of virus receptors creates the opportunity to better understand the virus-host relationship and create pigs that are resistant to disease.

W1**039:** **Swine**
Sequence Enabled Genomic Prediction in Commercial Pig Populations
John Hickey, Roslin Institute, Midlothian, Scotland, United Kingdom

W1**040:** **Swine**
Loss of Function Mutations in Essential Genes cause Embryonic Lethality in Pigs
Martijn F.L. Derks[1], Arne B Gjuvsland[2], Mirte Bosse[1], Marcos S. Lopes[3], Maren van Son[2], Barbara Harlizius[3], Eli Grindflek[2], Martien A.M. Groenen[1] and Hendrik-Jan Megens[1], (1)Wageningen University, Wageningen, Netherlands, (2)Topigs Norsvin, Hamar, Norway, (3)Topigs Norsvin Research Center, Beuningen, Netherlands
Lethal recessive alleles cause pre- or postnatal death in homozygous affected individuals, reducing population fertility. Especially in small size domestic and wild populations, those alleles might be exposed by inbreeding, caused by matings between related parents that inherited the same recessive lethal allele from a common ancestor. In this talk I will discuss the methodology to screen for recessive lethals and report six relatively common (up to 13.4% carrier frequency) recessive lethal mutations in commercial pig populations. While these lethal mutations have a large effect on carrier-by-carrier matings and decrease litter sizes by 15.1 to 21.6%, the mutations are maintained at stable frequencies within the population. The causal mutations are of different type including two splice-site variants (affecting *POLR1B* and *TADA2A* genes), one large deletion (*BBS9*), one frameshift (*URB1*), and one missense (*PNKP*) variant, resulting in a complete loss-of-function of these essential genes. Moreover, we show that the deletion within the *BBS9* gene is maintained in the population because of a strong positive effect on growth in heterozygotes (i.e. balancing selection)*, while the other lethals are likely the* result of genetic drift. Together, the recessive lethal alleles affect up to 2.9% of the litters within a single population and are responsible for the death of 0.52% of the total population embryos. Moreover, we provide compelling evidence that the identified embryonic lethal alleles contribute to the observed heterosis effect for fertility (i.e. larger litters in crossbred offspring). None of the mutations are found across populations. Therefore, no embryonic losses are expected in the crossbred offspring used for pork production. Together, this work marks specific recessive lethal variation describing its functional consequences at the molecular, phenotypic, and population level, providing a unique model to better understand fertility and heterosis in livestock.

W1**041:** **Swine**
WGS Analysis of Teats- Number of Vertebrae QTL
Barbara Harlizius[1], Maren van Son[2], Marcos S. Lopes[1], Henry Martell[3], Martijn F.L. Derks[4], Jørgen Kongsro[2] and Eli Grindflek[2], (1)Topigs Norsvin Research Center, Beuningen, Netherlands, (2)Topigs Norsvin, Hamar, Norway, (3)School of Biosciences, University of Kent, Canterbury, United Kingdom, (4)Wageningen University, Wageningen, Netherlands
In this study, we combine large genotype and phenotype data sets from two different pig breeds (Duroc and Landrace) to analyze a QTL region for number of teats (NTE) on chromosome 7 at the quantitative and the molecular level. Refining trait definition by counting vertebrae (NVE) and ribs (RIB) from CT scan data increases heritability from 0.28 for NTE up to 0.62 for NVE and 0.78 for RIB in Duroc. In Landrace, heritability for RIB is much lower (0.24) compared to NVE (0.59). At the molecular level, haplotypes derived from 660K SNP data identify a common haplotype of seven SNPs in Duroc. Sequence analysis of 16 Duroc animals shows that two functional mutations increasing expression of Vertnin (*VRTN*) known to increase number of thoracic vertebrae (ribs) reside on this haplotype. In Landrace, the linkage disequilibrium extends over a region of more than 3 Mb also containing both *VRTN* mutations. In all sequenced Landrace animals, additional variants are found on the wild type haplotypes surrounding the *VRTN* region. Also, variants at other modifying loci are expected across the genome explaining the breed-specific effect. Together, the integration of large-scale genotype, phenotype, and sequence data shows exemplarily how population parameters are influenced by underlying variation at the molecular level i.e. differences between lines in allele frequencies, in LD with functional mutations, in genetic background and in accuracy of rib counting.

W1**042:** **Swine**
An Atlas of *Sus scrofa* RNA Editome in Developmental Skeletal Muscle
Zhonglin Tang, Agricultural genomics Institute, CAAS, Shenzhen, China

W1**043:** **Swine**
Error Checking and Improving Annotations of the Current *Sus scrofa* Build
Harry D. Dawson, ARS/USDA, Beltsville, MD

W1**044:** **Swine**
Weaning Induces Rapid DNA Methylation and Transcriptional Changes in Piglet PBMCs Associated with Impaired Stress Response
Ryan J. Corbett[1], Kaitlin E. Wurtz[2], Janice M. Siegford[2], Nancy E. Raney[2], Laura M. Ford[2] and Catherine W. Ernst[2], (1)Genetics Graduate Program, Michigan State University, East Lansing, MI, (2)Department of Animal Science, Michigan State University, East Lansing, MI
Prolonged immune system activation can impair numerous aspects of livestock physiology, including the ability to respond to subsequent stressors via hypothalamic-pituitary-adrenal (HPA) axis dysregulation. Piglet weaning is associated with nutritional and social stress, both of which elicit an immune response. DNA methylation is proposed as a mechanism by which stress induces long-term physiological changes. Previous animal studies have observed hypermethylation and downregulation of the glucocorticoid receptor (*NR3C1*)—a key regulator of the HPA axis—in blood of stressed versus non-stressed individuals. However, the impact of weaning on gene methylation and associated expression

in pig tissues has not been assessed. We quantified changes in CpG methylation and transcript abundance in peripheral blood mononuclear cells (PBMCs) in piglets following weaning. Blood was collected from 10 females (5x2 litters) 48h before and 24h after weaning, and littermates with the highest and lowest cortisol concentrations and lesion scores (n=4) were selected for whole-genome bisulfite sequencing and RNA-sequencing. We identified 15,568 differentially methylated regions (DMRs, 8,913 hypermethylated and 6,655 hypomethylated post-weaning), which were enriched within promoters of genes associated with transcriptional regulation and immune cell activation. DMRs were enriched among 533 identified differentially expressed genes, 440 of which were upregulated and enriched for immune and inflammatory processes. We observed post-weaning hypermethylation at the *NR3C1* promoter, and significant decrease in *NR3C1* expression that was validated in a larger set of PBMC samples via RT-qPCR (n=9, $p=6.1x10^{-3}$). Our results indicate weaning-associated stress elicits genome-wide methylation changes associated with differential gene expression and a reduced HPA axis response.

W1**045: Swine**

Jorgensen Award 2: Deciphering the dynamics of Porcine Circovirus Associated Diseases using Gene Silencing and Editing

Lianna Walker[1], Taylor Engle[2], Hiep Vu[2], Emily Tosky[2], Dan Nonneman[3], Timothy P.L. Smith[3], Tudor Borza[4], Thomas Burkey[2], Graham S Plastow[5], Stephen D. Kachman[2] and Daniel C. Ciobanu[2], (1)UNL, Lincoln, NE, (2)University of Nebraska - Lincoln, Lincoln, NE, (3)USDA, ARS, USMARC, Clay Center, NE, (4)Dalhousie University, Truro, NS, Canada, (5)University of Alberta, Edmonton, AB, Canada

Porcine Circovirus 2 (PCV2) is the smallest known DNA virus capable of infecting mammals and the primary agent required for the development of a set of symptoms collectively known as Porcine Circovirus Associated Diseases. However, infection with PCV2 does not guarantee clinical disease with variation in severity observed between breeds and individuals. Experimental infection of ~1,000 pigs with PCV2b and genome-wide association analyses revealed two QTL for PCV2b viral load, located on SSC7 and SSC12. Dissection of the SSC12 QTL using *ab initio* gene prediction, RNAseq, and genomic sequencing identified 66 novel polymorphisms within and upstream of five positional candidate genes. Single marker association analysis of a subset of pigs with extreme high and low viral loads, identified two novel polymorphisms in high LD accounting for 21-23% of the phenotypic variation. One polymorphism is a missense mutation (*p.Arg63Cys*) within the second exon and critical domain of the *SYNGR2* gene. *In vitro* silencing of *SYNGR2* in PK15 cells via siRNA transfection resulted in a one-log reduction in PCV2b titer ($P<0.05$) compared to scramble siRNA and non-transfected control cells, indicating a direct role of this gene in PCV2b infection. Additionally, gene editing using CRISPR-Cas9 ribonucleoprotein complexes targeting *SYNGR2* generated a PK15 edited clone homozygous for a 106bp deletion within the second exon that exhibited a two-log reduction in PCV2b titer following infection compared to wild-type PK15 cells ($P<0.05$). Given these findings and that *SYNGR2 p.Arg63Cys* is the only missense mutation within this gene, *SYNGR2 p.Arg63Cys* is a plausible QTN for PCV2b susceptibility.

W1**046: Swine**

Cataloguing Genetic Variants in Commercial Swine

Brittney N Keel, Dan Nonneman and **Gary A. Rohrer**, USDA, ARS, USMARC, Clay Center, NE

To aid in the development of a comprehensive list of functional variants in the swine genome, single nucleotide polymorphisms (SNP) and copy number variations (CNV) were identified from whole genome sequence of 240 pigs. In this work, we utilized whole genome sequence from 240 members of a heavily phenotyped swine herd at the U.S. Meat Animal Research Center (USMARC). These animals included all 24 of the founding boars (12 Duroc and 12 Landrace), 48 of the founding Yorkshire-Landrace composite sows, 109 composite animals from generations 4 through 9, 29 composite animals from generation 15, and 30 purebred industry boars (15 Landrace and 15 Yorkshire) used as sires in generations 10 through 15. Sequence reads were mapped to Sscrofa 11.1 reference genome. A total of 26,850,263 high confidence SNP and 3,716 copy number variable regions (CNVR) were identified. Variation was detected in the coding sequence or untranslated regions (UTR) of 78% of the genes in the porcine genome: 1,729 loss-of-function variants were predicted in 1,162 genes, 12,686 genes contained 64,232 nonsynonymous variants, 250,403 variants were present in UTR of 15,739 genes, and 15,284 genes contained 90,939 synonymous variants. CNVR covered 0.94% of the porcine genome and overlapped 1,442 genes. The Gene ontology analysis identified that CNV genes were enriched for functions related to neurophysiological processing of environmental stimuli, including sensory perception, signal transduction, and olfactory receptor activity, as well as many functions related to organism development. Many of these genetic variants (SNP and CNV) are expected to alter protein production and likely contribute to phenotypic variation in economically important traits.

W1**047: Swine**

MSU Station Report: Behavioral Phenotyping and Modeling of Social Genetic Effects

Juan P. Steibel[1], Belcy K. Angarita Barajas[2], Junjie Han[3], Sarah Chewning[4], Kaitlin Wurtz[4], Carly O'Malley[4], Rodolfo Cantet[5], Janice M. Siegford[6] and Catherine W. Ernst[6], (1)MSU, East Lansing,, MI, (2)Michigan State University, East Lansing`, MI, (3)MSU, East Lansing, MI, (4)MSU, East Lansing`, MI, (5)UBA, CABA, Argentina, (6)Department of Animal Science, Michigan State University, East Lansing, MI

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. 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 for 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. The proposed approach recovered 40% to 80% more genetic variance compared to a model without social genetic effects, while a model of traditional social genetics effects, assuming uniform interactions between all group mates, was not able to recover any additional variance. Automatic behavioral phenotyping unlocks the use of better social genetic effects modeling as well as precision livestock management applications.

W1**048:** **Swine**

Concerted Analysis of the Expression of 230 Genes Distinguishes Fetal and Placental Responses to Congenital Infection with Porcine Respiratory and Reproductive Syndrome Virus and Affirms Intra-Litter Diversity in Immune Responses

Joan K. Lunney, APDL, BARC, 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 first time pregnant pigs, in their third-trimester, and assessed maternal and fetal factors that could be predictive of PRRS severity and resilience in fetal pigs. To better understand whether fetal mortality is the result of viral disruption of placental function or aberrant fetal immune response to infection, we evaluated transcriptomic responses in fetal and placental tissues following maternal PRRSV challenge.

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 placenta (PLC) samples were assigned to 3 experimental groups: ND (none detected) (PLC=0, N=12), LOW (PLC >1 & <3.9, N=30) and HI (PLC >3.9, N=34). For fetal thymus (THY) equivalent groups were selected based on a combination of THY and Serum (SER) viral load: ND (THY & SER virus=0, N=22), LOW (THY=0, serum >1 & < 5, N=22) and HI (THY>4 and SER>5, N=11). The resulting data was normalized using a combination of NanoString positive control probes and 9 housekeeping genes, and a univariate analysis conducted using a Generalized Least Squares (GLS) model with false discovery rate (FDR) correction.

Overall, gene selection for the NanoString array was effective in distinguishing PLC and THY immune responses to PRRSV infection. Gene expression changes were clearly evident especially in the fetuses with the high viral burden. This is particularly evident in the Antigen Presentation and Interferon Response pathway genes with fetal thymus showing higher expression than placenta for many genes. Disease progression in fetuses was accompanied by changes in inflammatory and early protective immune responses. 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.

W1**049:** **Swine**

The Microbiability of Meat Quality Traits in Swine

Christian Maltecca, NCSU, Cary, NC; NCSU, Raleigh, NC, Clint Schwab, The Maschhoffs LLC, Carlyle, IL, Piush Khanal, North Carolina State University, Raleigh, NC and Francesco Tiezzi, NCSU, raleigh, NC

To date studies on effect of host microbiota on meat quality (**MQ**) traits of swine remain limited. The objective of this study is to estimate the microbiability of meat quality traits and its impact on heritability estimates in commercial swine. The study population consisted of 1,123 crossbred individuals genotyped and phenotyped for meat quality traits. Fecal 16S microbial sequences for all individuals were obtained at three different stages: off test (**OT:** 196.4 ± 7.80 days); week 15 (**W_15:** 118.2 ± 1.18 days); and weaning (**WEAN:** 18.64 ± 1.09 days)). Data were analyzed using single-trait models, which included the fixed effects of dam line, contemporary group (**CG**), and gender as well as random effects of pen animal and the microbiome. The last two were modeled with the use of similarity matrices among individuals obtained through the use of genomic and microbial information. Analyses were conducted in ASREML v.4. Microbial composition contributed significantly to the overall variation at OT and W_15. Microbial contribution was significant for most traits, but it varied with time with estimates of microbiability increasing from WEAN to OT for almost all traits and ranging from 0.03 ± 0.02 for intramuscular fat to 0.13 ± 0.04 for firmness score. In most cases inclusion of microbial composition did not affect estimates of genomic heritability suggesting that for MQ traits gut microbial composition is in large part an environmental effect.

W1**050:** **Swine**

UNL Station Report

Hiruni R Wijesena and Daniel Ciobanu, University of Nebraska - Lincoln, Lincoln, NE

W1**051:** **Swine**

NIFA Functional Annotation of Animal Genomes Pig Resource Project and the SCID Pig Project in the Tuggle Group

Christopher K. Tuggle, Animal Science, Iowa State University, Ames, IA

This talk will summarize the swine research work in the Tuggle lab in 2018. 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 resource project and discuss its relevance to swine genomics. In (2), I will provide a summary of recent results including a demonstration that pig phagocyte cells are much more tolerant of human cells than are mouse phagocytes, that SCID pigs can be raised in biocontainment facilities with novel Standard Operating Procedures (SOPs), and that the SCID pig is an excellent model for xenotransplantation.

W1**052:** **Swine**

NRSP8 Bioinformatics Presentation

James M. Reecy, Iowa State University, Ames, IA

W1**053:** **Synthetic Biology**

Homology Directed Repair of CRISPR/Cas9 Induced DNA Breaks in Sugarcane

Tufan Mehmet Oz, Agronomy Department, Plant Molecular and Cellular Biology Program, Genetics Institute University of Florida - IFAS, Gainesville, FL; DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Florida - IFAS, Gainesville, FL

Genome editing tools such as CRISPR/Cas9 have been employed in several crop genomes. They enable precise targeting and introduction of double stranded DNA breaks in vivo. Subsequent cellular repair mechanisms, predominantly non-homologous end joining (NHEJ), act as critical steps to endogenous gene editing or correction. However, there is very limited control over these mechanisms, which generate an abundance of random insertions and deletions (indels). Frameshift mutations associated with these indels of unspecified size and sequence might result in loss of function phenotypes of agronomic importance. Gain of function mutations, on the other hand, generally require precise nucleotide substitutions in the target locus. This can be accomplished with the aid of a homologous repair template and involves the cellular homology directed repair (HDR) mechanism. We will present data that supported efficient HDR mediated precision editing of multiple alleles of the acetolactate synthase (ALS) gene in the highly polyploid sugarcane and conferred herbicide resistance.

W1**054:** **Synthetic Biology**

Chloroplast and Nuclear Gene Editing Platform in Chickpea with CRISPR/Cas9

Yurdagul Ferhaoglu, U of Saskatchewan, Saskatoon, SK, Canada

Yurdagul Ferhatoglu[1], Pankaj Bhowmik[2], Bunyamin Tar'an[1]. 1.Plant Science, University of Saskchewan, 2. NRC Saskatoon, CANADA

Genome editing has challenged molecular biologist, biochemists with new discoveries in recent years, now it became a new field of science within genome engineering and carries the potential for germplasm development. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) editing system has enabled editing of plant genes in a wide range of plant species protoplasts. CRISPR/Cas9 target online predictor (CTOP) was initially validated in the non-plant cell, later used in plant Cas9 target predictions. Phytoene desaturase catalyzing the first step of carotenoid biosynthesis, whose inhibition leads to albino phenotypes has become the validating model for editing of nuclear-encoded genes. Here we demonstrate CCTOP predicts Cas9 targets efficiently in an in-vitro psbA cleavage and mesophyll protoplast PDS assays from chickpea leaf cells. The CRISPR /Cas9 activity assays with 20-22 bp gRNA seed regions generated indels and point mutation and validated NHEJ system by editing the fifth exon of PDS in chickpea protoplasts, it's assaying was performed with 35S-Cas9 and U6-gRNAs plasmids concurrent transformation. Assay results were confirmed with agarose gel and Sanger sequencing. The established platform is now being used in generating a base for plastidic proof of concept experiments for psbA edits in chickpea protoplasts.

W1**055:** **Synthetic Biology**

A Systems Biology Approach to Explore Abiotic Stress Response in Sorghum

Indrajit Kumar, Donald Danforth Plant Science Center, Saint Louis, MO

Sorghum is a major cereal crop worldwide, and is an emerging bioenergy feedstock because of its tolerance to drought and low nutrient inputs. There is also a rich genetic diversity that represents an untapped resource for improving productivity on marginal lands. A comprehensive understanding of the molecular and physiological responses of bioenergy sorghum varieties under abiotic stresses will aid in creating stress-tolerant and high yielding cultivars. Here, we use a systems biology approach to compare responses to drought and low nitrogen stress among three diverse sorghum genotypes on multiple scales; i) high-throughput phenotyping and image-based analysis, ii) genome-wide transcriptional changes using RNA-seq and chromatin accessibility using ATAC-seq, and iii) physiological responses including photosynthetic rate, transpiration rate and water use efficiency. Preliminary results indicate genotype-specific responses at all scales. Our research targets identification of regulatory regions and mechanisms that confer tolerance to drought and low nitrogen in bioenergy sorghum.

W1**056:** **Synthetic Biology**

Knock-out of SEED Fatty Acid Reducers by CRISPR-Cas9 Increases the Seed Oil Content in Rapeseed (*Brassica napus*)

Nirosha L Karunarathna, Christian Albrechts University of Kiel, Kiel, Germany

W1**057:** **Synthetic Biology**

Gene Editing for Improving Yield and Yield Related Traits in Winter Wheat

Ragupathi Nagarajan, Oklahoma State University, Stillwater, OK

W1**058:** **Synthetic Biology**

Genome-Wide Evaluation CRISPR Off-Target Effect through Cas9 and Base-Editing System in Cotton

Jianying Li, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Jianying Li, Lei Qin, Xianlong Zhang and Shuangxia Jin*

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

*Correspondence: jsx@mail.hzau.edu.cn

The CRISPR/Cas9 system has been extensively applied for crop improvement. However, our understanding of Cas9 specificity is very limited in Cas9-edited plants. Recently, we established the CRISPR/Cas9 (*SpCas9*) system in cotton with high efficiency, which no off-target mutations were detected by targeted deep sequencing in the top 26 potential off-target sites of four sgRNAs. To further investigate the specificity of CRISPR/Cas9 system on the genome-wide scale in cotton genome editing, we described whole genome sequencing (WGS) of 14 Cas9-edited cotton plants targeted to three genes (6 sgRNAs), and 3 negative (Ne) control and 3 wild-type (WT) plants. In total, 4,188 - 6,404 unique single nucleotide polymorphisms (SNPs) and 312 - 745 insertions/deletions (indels) were detected in 14 Cas9-edited plants compared with WT, negative and cotton reference genome sequences. Since the majority of these variations lack a protospacer-adjacent motif (PAM), we demonstrated that the most variations following Cas9-edited are due either to somaclonal variation or/and pre-existing/natural variation from

maternal plants, but not off-target effects. Of a total of 4,413 potential off-target sites (allowing ≤ 5 mismatches within the 20-bp sgRNA and 3-bp PAM sequences), the WGS data revealed that only 4 are *bona fide* off-target indel mutations, validated by Sanger sequencing. Moreover, inherent genetic variation of WT can generate novel off-target sites and destroy PAMs, which suggested great care should be taken to design sgRNA for the minimizing of off-target effect. These findings suggested that CRISPR/Cas9 system is highly specific for cotton plants.
At the same time, we also developed a base editing system by fusing cytosine deaminase (APOBEC), nCas9 and uracil glycosylase inhibitor (UGI) and then this unit was inserted into our previous CRISPR/Cas9 plasmid vector to create single base mutations in the allotetraploid cotton genome, which designated as *G. hirsutum*-Base Editor 3 (GhBE3). Two target genes *GhCLA* and *GhPEBP* were selected to test the efficiency and accuracy of GhBE3. Firstly, we performed Sanger sequencing on three target sites with the editing efficiency ranging from 26.67% to 57.78%. In addition, targeted deep sequencing revealed that the C→T substitution efficiency within the "editing window" (the position 3 to 9 of the 5'-end distal to the PAM sequence) was up to 18.63% from the total sequences. Secondly, the 27 most potential off-target sites were sequenced by targeted deep sequencing, and the results showed that rare C→T substitution (average < 0.01%) was detected in editing window of these sites. Furthermore, two GhCLA-nCas9 edited and wild-type plants were applied for whole genome sequencing with about 100× depth, no any *bona fide* off-target mutation was detected from 1500 predicted potential off-target sites on genome-wide scale. Taken together, these results demonstrated that the *GhBE3* has high specificity and accuracy to generate targeted point mutation in allotetraploid cotton.

W1**059:** **Systems Biology and Ontologies**
Epigenetic Regulation of Genome Dynamics and Evolution in the Model Aquatic Monocot *Spirodela polyrhiza*
Todd Michael, J. Craig Venter Institute, La Jolla, CA
Duckweeds are the fastest growing angiosperms and have potential to be a source of biomass for sustainable agriculture and renewable biofuel. Here we generate a chromosome-level genome sequence map of *Spirodela polyrhiza* utilizing short-read *de novo* assembly coupled with a high-depth, motif-specific physical map enabled by a high-throughput technology. In addition, the physical maps resolved the 18S-5.8S-26S rDNA cluster revealing less than 100 copies in Spirodela. Using RNA-seq, small RNA-seq and Bisulfite-seq data to empirically annotate features of the genome, we confirmed the lowest gene count of any angiosperm to date at 18,507, resolved repeat content upwards of 25%, identified a cryptic 119 basepair centromere repeat, and discovered 29 miRNAs specific to Spirodela. Compared to other plants, Spirodela has the lowest global DNA methylation levels at 9%, and even lower levels (3%) in syntenous regions consistent with these regions being protected from gene loss. In addition, Spirodela has a high Solo to Intact (S:I) Long Terminal Repeat ratio of 8.52, the highest of any plant genome tested to date. These results are consistent with a purging event in Spirodela evolution that resulted in a low total gene count while important genes can be selectively retained at higher copies via hypomethylation.

W1**060:** **Systems Biology and Ontologies**
Using Gene Networks to Understand Natural Variation in Leaf Form
Neelima R. Sinha, University of California, Davis, Davis, CA

W1**061:** **Systems Biology and Ontologies**
Hops Genome to be Added
David Hendrix, Dept of Biochemistry and Biophysics, Oregon State University, Corvallis, OR

W1**062:** **Systems Biology and Ontologies**
Evolution of the Natural Genetic Engineer *Agrobacterium*
Alexandra J. Weisberg, Oregon State University, Corvallis, OR

W1**063:** **Systems Biology and Ontologies**
SuRE: An Unbiased, Genome-Wide Survey of Regulatory Elements that provides Targets for CRISPR Gene Editing
Ferdinand Los, Hudson River Biotechnology B.V., Wageningen, Netherlands and **Lotte Westerhof**, Hudson River Biotechnology, Wageningen, Netherlands
SuRE, 'survey of regulatory elements', is a novel platform technology that allows unbiased identification of gene regulatory elements. These elements can serve as unique and sophisticated targets for molecular plant breeding through e.g. CRISPR and TILLING, and as novel, strong endogenous promoters to drive transgene expression.
With SuRE, a plasmid library is constructed, consisting of random 500bp genomic fragments inserted upstream of unique 20-bp barcodes, and then characterized. The SuRE library is transfected into protoplasts, and barcode expression is quantified by high-throughput sequencing. The genomic sequences associated with the expressed barcodes are then mapped onto a reference genome. Over 50-fold genome coverage can be reached, allowing mapping of autonomous promoter and enhancer activity in the form of a peak pattern to a reference genome.
Here, we report on the first phase of translating SuRE from validated biomedical applications to new agricultural applications. We demonstrate that we could scale several technical barriers and that the SuRE platform technology can be successfully applied to plant genomes, starting with tomato.

W1**064:** **Systems Biology and Ontologies**
NASA GeneLab: Enabling Open Science for Life in Space
Matthew Geniza, USRA/NASA, Mountain View, CA

W1**065:** **Systems Genomics**
Discovery of Expressible Gene Sets of Sorghum and Maize by Machine Learning with Omics Data
Laura de Boer, University of California San Diego, La Jolla, CA

Determining which predicted gene models produce functional products is an exciting challenge in genome-wide biology. While the number of predicted gene models from plant genomes can exceed 60,000, the plant research community has detected transcript products from only a subset of these genes, and an even smaller subset have detected protein products. Gene homology and expression evidence is often used to curate a high confidence group of genes from full gene model sets, e.g. the filtered versus working gene sets of maize. An open question is whether genes outside of these curated high confidence sets are expressible at the protein level. A random-forest based approach utilizing gene DNA methylation patterns as model features was used to identify the expressible gene sets of two staple food crops. These expressible gene sets were defined with high accuracy for sorghum (*Sorghum bicolor*) and two diverse inbred lines of maize (*Zea mays*), at both the transcript and protein level. CG and CHG methylation levels of the gene body, specifically for exons, were the most important features for model accuracy. While sorghum and the maize inbreds have similar gene content, the expectation is that phenotypic differences between species or inbreds is driven by differences in the proteome. Synteny between grasses was leveraged to identify gene models which are predicted to be uniquely expressible between species and which may explain a portion of the phenotypic diversity between species.

W1**066:** **Systems Genomics**

Statistical Learning on Heterogeneous Omics Data Highlights the Molecular Basis of Heterosis of Drought Response in Sunflower

Harold Duruflé[1], Thierry Balliau[2], Nicolas Blanchet[1], Mélisande Blein-Nicolas[2], Adeline Chaubet[1], Brigitte Mangin[1], Pierre Maury[3], Michel Zivy[2] and Nicolas Langlade[1], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France, (2)Génétique Quantitative et Evolution - le Moulon, PAPPSO, Gif-sur-Yvette, France, (3)University Toulouse INPT ENSAT UMR1248AGIR, Castanet Tolosan, France

Climate change is a major concern because of its potential effects on biodiversity and the agricultural sector. The domesticated sunflower, *Helianthus annuus* L., is the fourth most important oilseed crop in the world [1], is cultivated as hybrids and is promising for agriculture adaptation because it can maintain stable yields across a wide variety of environmental conditions, especially during drought stress [2].
Here, we studied the responses of eight parental lines (four males and four females) and their 16 hybrids cultivated on the outdoor high-throughput automated phenotyping platform Heliaphen [3] where ecophysiological traits were measured. Leaf samples from 144 samples were collected for multi-omics analyses including label-free shotgun analysis and protein identification and quantification.
Statistical learning methods without a priori allowed us to reduce this complex molecular system [4]. Statistical modelling of relationships combining proteomics and phenomics between these datasets aimed at establishing molecular players of heterotic behaviours and its role during drought stress. We demonstrate strong behavioral differences between hybrids and parental lines with solid interaction with water deficit that are exemplified by some highly interesting protein candidates. Further analysis of these genes or proteins highlights the impact of genetic variation on regulatory networks that could be selected to improve heterotic groups and breed for drought stress tolerance.
Current integration of eco-physiological, proteomics, transcriptomics and metabolomics levels in order to achieve a more holistic view will be presented and discussed.

Keywords: Helianthus, Systems biology, Quantitative genetics, Candidate proteins
Acknowledgements: These data were produced with the funding of the French National Research Agency (ANR SUNRISE ANR-11-BTBR-0005).

References

[1] FAS. USDA, Oilseeds: World Markets and Trade, 2018.
[2] H. Badouin, *et al.,* The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution, Nature. 546 (2017) 148–152. doi:10.1038/nature22380.
[3] F. Gosseau *et al.,* Heliaphen, an outdoor high-throughput phenotyping platform designed to integrate genetics and crop modeling, BioRxiv. (2018) 362715. doi:10.1101/362715.
[4] H. Duruflé, *et al.,* A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis, BioRxiv. (2018) 357921. doi:10.1101/357921.

W1**067:** **Systems Genomics**

Tea Tree Genome Sequencing reveals why many Health-Beneficial Natural Products are Concentrated in One Leaf

Jian Zhao, Anhui Agricultural University, Hefei, China
Jian Zhao[1], Guangbiao She[1], Shuwei Yu[1], Penghui Li[1], Yingying Li[1], Chaoling Wei[1], Hua Yang[1], Songbo Wang[2], Linlin Liu[1], Shancen Zhao[2], Xiaochun Wan[1].
[1]*State Key Laboratory of Tea Plant Biology and Utilization, School of Tea & Food Science and Technology, Anhui Agricultural University, Hefei, 230036, China;*
[2] *BGI–Shenzhen, Shenzhen 518083, China.*

As the largest non-alcohol beverage in the world, only next to water, teas own numerous health benefits, which now are one of the hottest research areas. How tea plant buds and leaves can synthesize and accumulate such high levels of flavonoid catachins (~12-24% of dry mass), free amino acid theanine (~1-4% of dry mass), purine alkaloid caffeine (~1-3% of dry mass), and triterpenoid saponins (~10-20% of dry mass), remains mysteries. Until recently, the genome sequences of *Camellia sinensis* var. sinensis (CSS), and *C. sinensis* var. assamic (CAS) were revealed, which enable us to understand not only the evolution of tea plants, but also its characteristic production of all major types of plant secondary metabolites in one leaf. Theanine contributes to tea infusions with unique umami and sweet tastes and health-promoting functions, such as sleep improvement, mental relaxation, cognition enhancement, brain protection, and liver detoxification, however, the mechanism and molecular identity for its biosynthesis remain obscure. A higher quality 3.1-Gb-size genome of CSS tea plants has been annotated with more than 33,000 high-confidence genes in a ~92 % complete assembly of the tea plant genome with high heterozygosity (~ 0.8 %). Two rounds of WGDs occurred ~30-40 mya and ~90-100 mya triggered the tea plant evolution and eventual speciation. But subsequent paralogous duplications more significantly contributed to the formation of the characteristic secondary metabolites, such as catechins, theanine, and caffeine, which were dissected on their biosynthetic genes for divergent or convergent evolutions. By using genomic, genetic and biochemical approaches, we have clarified the molecular identity of theanine synthetase (TS). CsTSs are the bifunctional enzymes that can synthesize both glutamine and theanine

from condensation of glutamic acid with ammonium and ethylamine, respectively. *CsTS*s co-expression with theanine accumulation patterns, subcellular location, in situ cell type expression, overexpression in Arabidopsis and soybean, and knockdown of *CsTS*s in tea plants, and enzyme-substrate affinity assays, we demonstrated the mechanism of CsTSs for theanine production in tea plants, which may facilitate the breeding of high-theanine-content tea plant varieties for improvement of human health.

References:
Wei C*, Yang H*, Wang S*, Zhao J*, Liu C*, Gao L*, Xia E, Lu Y, Tai Y, She G,.... Bennetzena JL(*), Zhao S(*), Wan X(*).(2018) Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. Proc. Natl Acad Sci USA. 2018; 115(18):E4151-E4158

W1068: Systems Genomics

Quantification of Gene Expression while taking into account RNA Alternative Splicing

Meiping Zhang[1], Yun-Hua Liu[1], Chin-Sheng Chang[1], Hui Zhi[1], Shichen Wang[2], Wenwei Xu[3], C. Wayne Smith[1] and **Hong-Bin Zhang**[1], (1)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, (2)Genomics and Bioinformatics Service, Texas A&M AgriLife, College Station, TX, (3)Texas A&M AgriLife Research, Lubbock, TX
M.Z. & Y.-H.L. equally contributed to this work.

Gene expression has been widely used for functional genomics research, especially gene cloning, functional analysis, regulation, pathway inference and network construction. However, the gene expressions quantified with different methods have been frequently inconsistent, thus challenging the results and conclusions from such research. Here we have addressed this issue by taking into account RNA alternative splicing. We confirmed that RNA alternative splicing that may lead to one to hundreds of transcripts that may be translated into different proteins having different biological functions is prevalent in plants and varies dramatically among tissues and within a bi-parental population. RNA alternative splicing substantially influenced the accurate qualification of gene expression. When a gene was subjected to RNA alternative splicing, it was impossible or difficult to properly quantify the expression of a transcript of the gene or its overall expression using quantitative real-time PCR (qPCR), Northern hybridization, microarray, or serial analysis of gene expression. The results showed that of the genes analyzed, the expressions of at least 53% could not be properly quantified by qPCR. Shot-gun RNA-seq was the most proper to quantify the expression of a transcript or a gene in such cases. Moreover, the expressions of individual transcripts quantified by shot-gun RNA-seq were highly reproducible ($r = 0.90 – 0.98$) between biological replicates. Therefore, shot-gun or full-length RNA-seq should be the method of choice to properly quantify the expression of a transcript or a gene. These findings also strongly indicate that it is necessary to investigate the functions of a gene using its transcript(s). For detail, see Genomics, 2018 (https://doi.org/10.1016/j.ygeno.2018.10.009).

W1069: Systems Genomics

Systematic Detection of Orthologous Gene Groups Sharing an Evolutionary History, using Network Topology Search

Dong-Ha Oh, Louisiana State University, Baton Rouge, LA

tentative abstract - I will revise it before Dec. 1st

The CLfinder-OrthNet pipeline (1) detects co-linearity among multiple closely-related genomes, (2) finds orthologous gene groups, and (3) encodes the evolutionary history of each ortholog group into an Ortholog Network (OrthNet). OrthNets connect orthologs with edges representing either the presence or absence of co-linearity between them. Each OrthNet encodes in its network topology the evolutionary history of an orthologous locus, including different modes of gene duplication, deletion, transposition, and combinations of them, occurred in a lineage or multiple lineages. Orthologous gene groups with the same evolutionary history can be retrieved by searching OrthNets with a network topology query.

As a proof-of-concept, we applied CLfinder-OrthNet to characterize gene transposition-duplication (*tr-d*) events among six Brassicaceae genomes, including those of *Arabidopsis thaliana* and two extremophytes, *Eutrema salsugineum* and *Schrenkiella parvula*. We identified subsets of lineage-specific *tr-d* events with signatures of selective retention and sub-functionalization in all six genomes. These included lineage-specific *tr-d* of genes that may be critical for the local adaptation of extremophytes, such as orthologs of *SALT TOLERANCE 32* and *ZINC TRANSPORTER 3*.

CLfinder-OrthNet offers a flexible toolset for systematic comparative studies of closely-related genomes. Beside the detection of all orthologs showing the same evolutionary history, the application includes but not limited to: (1) improving orthology inference assisted by co-linearity, (2) identification of gene duplication or transposition events co-occurring with certain phenotypic traits among closely-related genomes, and (3) detection of truncated, split, and chimeric gene models based on co-linearity.

CLfinder-OrthNet is available at https://github.com/ohdongha/CL_finder

tentative abstract - I will revise it before Dec. 1st

W1070: Systems Genomics

Deciphering the Regulatory Code of Metabolism in *Arabidopsis thaliana*

Kangmei Zhao, Michael Banf, Pascal Schlapfer and Sue Y. Rhee, Carnegie Institution for Science, Palo Alto, CA

Coordination of metabolic genes is crucial for plant development and adaptation to various environments, but little is known about how metabolic pathways are transcriptionally regulated. Here, we sought to discover general rules of metabolic regulation in *Arabidopsis thaliana* by exploring omics data, regression models, and experimental validation. The goal of the first part of the study was to understand how metabolic genes are controlled by epigenetic marks. We determined the predictability of epigenetic marks on expression by constructing multiple linear regression models using sixteen genome-wide maps of epigenetic marks. The best model explained 61% of expression, which was similar to the human model. Different sets of epigenetic marks were enriched within various types of metabolic genes. Genes involved in energy metabolism were enriched with activation markers, such as histone 3 lysine 4 trimethylation (H3K4me3) and H3K36me3, and depleted of a repression mark, H3K27me3, which might be important in maintaining active expression of genes involved in energy metabolism. In contrast, specialized metabolic genes were significantly enriched with the repression mark H3K27me3 and an activation mark H3K18ac. We confirmed the colocalization of these two marks on the same genes using sequential-ChIP PCR on a representative pathway in specialized metabolism. To understand how these two marks control gene expression, we analyzed transcriptional kinetics of specialized metabolic genes under flagellin22

treatment. Among all induced genes, members with dual marks were turned on more rapidly than those with single modification. This suggested a potential novel mechanism controlling the expression of specialized metabolic genes, which we are currently following up genetically. In a second part of the study, we aimed to understand patterns of metabolic network regulation by transcription factors by constructing a condition- and tissue-specific regulatory network named MERIT. Based on previously characterized 584 interactions between transcription factors and metabolic genes, MERIT performed well compared to other currently existing regulatory networks. We identified 16 predominant transcription factor families that may directly regulate metabolism. We then computed the regulatory hierarchy and found that feed-forward loop was the major structure of metabolic regulatory network with five layers on average. Our next step is to validate the predicted transcription factors on a large scale. This study is the first systematic examination of epigenomic and transcriptional regulatory principles of metabolism in plants, which sheds light on our understanding of metabolism coordination and enabling pathway engineering in plants.

W1**071:** **Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

Public Repositories using Cloud Computing in Bioinformatics Teaching

Subhashini Srinivasan, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Pushpinder Singh Bawa and Subha Srinivasan Institute of Bioinformatics and Applied Biotechnology Bangalore 560100, India Impressive advances in sequencing technologies has made bioinformatics an indispensable discipline under biology. Data generation, which was a huge bottleneck only a decade ago, is now proliferating public repositories with petabytes of data from diverse NGS applications. Most of these these datasets are still underutilized and remain rich with yet undiscovered answers. Today, biological research can begin without ever entering a laboratory. The volume and velocity with which data are pouring into the public repositories is leaving no room for the research communities to equip themselves with skills or computing infrastructure that is needed to make biological sense of the data. There is a dire need for trained bioinformatician in biology departments across the globe. Yet, there is lack of trainers, facilities and infrastructure to meet this demand. Also, the speed with which scientists have embraced this technology leaves little room for the equally diverse analysis tools to mature. Virtual classrooms, public repositories and cloud computing infrastructure can be harnessed to take full advantage to scale training programs in bioinformatics. Since 2010, IBAB has extensively used public sequence repositories to provide hands on training in bioinformatics. Here, we will be presenting case studies where we have demonstrated the usefulness of public databases in biomarker discovery and validation. The future of biology will, for the most part, be driven by mining data stored somewhere in the cloud by someone in the distal past/part of the world.

W1**072:** **Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

Research Assisted Teaching Biology at Fayetteville State University

Jiazheng (John) Yuan, Fayetteville State University, Fayetteville, NC

W1**073:** **Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

Discovering Bioinformatics Analytic Tools Leveraging the Bioscience Query Tool Discovery System (Bio-TDS)

Carol Lushbough, Biomedical Engineering, University of South Dakota, Vermillion, SD and Etienne Gnimpieba, University of South Dakota, Sioux Falls, SD

Biological data management and analysis can become a daunting task due to the amount and diversity of data that experiments produce. Researchers might spend a great deal of time attempting to locate the right bioinformatics tool for their specific problem. Bio-TDS (Bioscience Query Tool Discovery System, http://biotds.org/) has been developed to assist in retrieving the most relevant bioinformatics analytic tools by allowing researchers to formulate their questions as free text. For example, a researcher could pose the question, "what are the best analytic tools to help me assign reliable mapping quality scores to mappings of reads returned by an alignment tool?" and be presented with a list of suggested tools. Bio-TDS leverages ontology based annotations and natural language processing to provide a robust system for tool discovery. The system includes functionality to gather individual community tool definitions and integrate these descriptions into a centralized retrieval system. By integrating analytic tool definitions from multiple repositories and providing flexible querying capabilities, the end-user is assisted in locating the most relevant computational toolsets required for their data analysis.

W1**074:** **Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

Undergraduates Contain Multitudes: Course-Based Metagenomics Analysis using the DNA Subway Purple Line

Ray A. 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 novice levels of experience in genomics analysis. Downloadable classroom modules and lesson plans for these activities are publicly available to educators via my personal James Madison University SelectedWorks faculty website (https://works.bepress.com/raymond_enke/).

W1**075:** **Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

Strategies for Incorporation of Bioinformatics into Life Science Curricula

Anne G. Rosenwald, Georgetown University, Washington, DC

The Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE) was initiated in 2014 with the overall goal of establishing bioinformatics as a crucial part of undergraduate life sciences education. In order to do so, we have created a network of educators and researchers sharing this vision. Specific objectives include

- how best to prepare students for bioinformatics instruction
- how best to integrate bioinformatics at all levels of the life sciences curriculum

- how to assess the outcomes
- how to prepare faculty trained in the life sciences to teach bioinformatics

To accomplish these goals, NIBLSE first conducted a survey of more than 1200 life science educators to understand the extent to which faculty are teaching bioinformatics currently and what aspects of bioinformatics they believe to be important to convey to their students. A paper outlining the community-refined core competencies in bioinformatics was recently published (Sayres et al., 2018 PLoS One). Efforts are currently underway to establish assessment tools that align with these core competencies.

As part of the survey, faculty were also asked about the barriers they encounter in teaching bioinformatics to their students; the most frequent response was faculty lack of training in the discipline (Williams et al., in preparation). In response, NIBLSE is sponsoring a faculty mentoring network (FMN) that will provide training and support to faculty interested in integrating a bioinformatics exercise into their introductory biology course. In addition, faculty also commented that it was difficult to find appropriate curriculum to use with their students. To that end, a set of resources is being gathered (the NIBLSE Learning Resource Collection). This resource acts as a clearinghouse of vetted materials, tagged with the appropriate level for instruction, including materials that were first published through a NIBLSE peer-review process called an incubator. The exercise used in the NIBLSE FMN is an example of an incubated resource. The Resource Collection, FMN, and other information can be found on the NIBLSE website (https://qubeshub.org/community/groups/niblse).

We encourage the education community to join NIBLSE to help us further our goals. We are in the process of establishing an implementation group, to pilot integration of bioinformatics into curriculum and we seek interested faculty to join. NIBLSE is supported by an RCN-UBE grant from the National Science Foundation (DBI 1539900).

W1**076:** **Teaching Genetics, Genomics, Biotechnology, and Bioinformatics**

Teaching Statistical Genomics to Biology Students

Zhenyu Jia, University of California Riverside, Riverside, CA

It is challenging to teach statistics to biology students with little background in quantitative genetics. My experience in teaching both statistics and genetics helped me find an efficient way to demonstrate critical statistical concepts and how they work in testing genetic hypotheses using intuitive examples or analogies, which has been well received by the graduate students in my Statistical Genomics class.

W1**077:** **The Analysis and Role of the Microbiome**

The Power of the Plant Microbiome

Sharon L. Doty, Andrew W. Sher, Jenny L. Knoth, Zareen Khan, Matthew R. Joseph, Pierre M. Joubert and Andrea Firrincieli, University of Washington, Seattle, WA

Just as the human microbiome has proven to be an important factor for our health, the plant microbiome may be an essential aspect of health for plants living in natural environments. Unable to move or to genetically adapt as environmental conditions change, long-lived plant species can utilize symbiotic partnerships with various microorganisms in order to meet their basic needs. The early successional pioneer tree species, poplar (Populus) and willow (Salix) colonize newly available primary substrates. Lacking the nutrients found in soils, these plants associate with N-fixing (diazotrophic) endophytes that can supply this essential macronutrient. Many endophyte strains also solubilize phosphate, potentially making this macronutrient more bioavailable. A consortium of the endophyte strains was added to hybrid poplar, increasing growth and N-fixation under greenhouse conditions. Under drought conditions, the endophytes promoted host plant survival and reduced host stress responses. Not only do the microbes improve growth of this important bioenergy plant species, they also increased growth, health, and yields of an exceptionally broad range of plant species, including rice, tomato, pepper, strawberries, ryegrasses, and Douglas-fir. Water use efficiency and drought survival were shown to increase in rice and conifer species. Some endophyte strains also have strong antimicrobial activities, inhibiting the growth of several major agricultural pathogens. With the need to substantially improve production to meet the needs of a growing human population, and with the increased stress of climate change, the implications of plant-microbe symbioses for agriculture, forestry, and bioenergy production are profound.

W1**078:** **The Analysis and Role of the Microbiome**

Microbiomes Associated with the Infective Stage of Reniform Nematode

Venkateswara R. Sripathi, Alabama A&M University, Normal, AL

W1**079:** **The Analysis and Role of the Microbiome**

The Porcine Gut Microbiome and its Relation with Host Phenotypes

Jordi Estellé i Fabrellas, GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

W1**080:** **The Analysis and Role of the Microbiome**

Filling the Void: Assembly of High-Quality Microbial Genomes from the Rumen Microbiome using Short- and Long-Read Sequencing

Amanda Warr[1], Robert Stewart[1], Marc Auffret[2], Alan Walker[3], Rainer Roehe[2] and Mick Watson[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)Scotland's Rural College, Edinburgh, United Kingdom, (3)The Rowett Institute, Aberdeen, United Kingdom

Ruminants such as cows and sheep are important livestock species. They convert low nutritional value plant matter into high-quality meat and dairy products. Within a specialised stomach called the rumen, microbes ferment the plant matter producing short-chain fatty acids from difficult to digest plant matter. The composition of the rumen microbial community can affect the animal's health, feed efficiency and level of methane production. The contents of the rumen are also of interest for discovering enzymes of use in biofuels. Species in the rumen are typically difficult to culture and despite its importance, it remains an underexplored environment. DNA sequencing of the contents of the rumen offers the potential to identify microbial species without culture techniques, however due to their poor culturability the species in the rumen are underrepresented in databases and reads often remain unclassified. Here we sequence fluid from the rumen of a single cow using long-read Nanopore sequencing. We

show that despite these data coming from a highly complex microbial sample we can assemble high-quality, single-contig whole genomes of known and novel species, including several circular contigs. Additionally, we use short-read Illumina sequencing of 282 cow rumen samples and binning methods to assemble 4,941 metagenome assembled genomes. These genomes will drastically improve classification rates for data from the rumen microbiome and allow for better analyses of the microbial community and discovery of enzymes involved in the process of digesting complex carbohydrates.

W1**081:** **The Analysis and Role of the Microbiome**
Microbiota Profiling with Long Amplicons using Nanopore Sequencing: Full-Length 16S rRNA Gene and 16S-ITS-23S from the *rrn* Operon
Anna Cusco, Vetgenomics, Bellaterra, Spain

W1**082:** **The Analysis and Role of the Microbiome**
Rapid, Culture-Free Microbiome Deconvolution and Mobile Element Tracking using Proximity Ligation Technology
Ivan Liachko, Phase Genomics, Seattle, WA

W1**083:** **The Analysis and Role of the Microbiome**
Effect of Sugarcane Polyphenols on Oral Microbiota
Pranav Chhaliyil, MSAE, Fairfield, IA

W1**084:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Global Drivers for Plant Research--a Big Picture Synthesis
Robert Bertram, US Agency for International Development, Washington, DC

W1**085:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Science Breakthroughs to Advance Food and Agricultural Research By 2030
Patrick S. Schnable, Department of Agronomy, Iowa State University, Ames, IA

W1**086:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Data Science, Computation, and Artificial Intelligence
Daniel Jacobson, Oak Ridge National Laboratory, Oak Ridge, TN

W1**087:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Sensing Technologies and Field-Based Phenotyping
Mitchell Tuinstra, Purdue University, West Lafayette, IN

W1**088:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Reproductive Biology and Apomixis
Venkatesan Sundaresan, University of California-Davis, Davis, CA

W1**089:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Plant Microbiome
Devin Coleman-Derr, Plant Gene Expression Center, USDA-ARS, Albany, CA

W1**090:** **The National Plant Genome Initiative: Catalyzing transdisciplinary frontiers in genomics for breakthroughs, innovation, and community development**
Genomics and Precision Breeding
Shawn Kaeppler, Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI
The vast resource of sequence information across species, including multiple individuals within species, provides a broad catalog of gene content and allelic variation. This catalog is increasing in depth and accuracy, but is currently poor in functional annotation. The next era of research will utilize novel measurements of plants and environments across time, coupled with big data analysis, to characterize the complex network of queues and responses that lead to productivity and quality phenotypes in our crop plants. Crop improvement approaches will include upgrades by targeted genome modification versus relying on meiosis and segregation to bring together combinations of alleles into improved individuals. Diversity within crops will be created with genome-enabled information either to deliver specific outcomes or to generate novel phenotypic variation that will be assessed for value. Interesting potential new avenues of research include rapid domestication of new crop species, and

matching crop genetics with human and animal genetics to optimize dietary needs or preferences. Investment in genome to phenome research to identify gene and phene function will lead us to the next era of crop improvement and crop creation.

W1091: Tools, Resources, and Funding for Basic and Applied Research: What Early Career Scientists Need to Know

Scale Your Science with Community Cyberinfrastructure

Nirav Merchant, University of Arizona, Tucson, AZ

Movements such as open data and data management plans mandated by funding agencies are liberating data and democratizing access to massive data sets. Complementing this data democratization are advances in computational and data handling capabilities, fueled by cloud computing and significant investment by NSF for establishing national scale cyberinfrastructure (CyVerse, XSEDE etc.). These national resources provide no cost access to some of the largest computational platforms in the world. Similarly, the software tools and analytics capabilities are catching up to this rapidly expanding array of computational platforms with containers (Docker, Singularity), computational notebooks (Jupyter, Zepplin) that are facilitating reproducibility and portability of complex workflows.

Leveraging these amazing resources to manage your own "big data" challenges for asking bold research questions necessitates interdisciplinary collaborations, hands-on training and technology orientation. This talk will provide a roadmap with a broad overview of exemplar communities and projects that have successfully established their own cyberinfrastructure utilizing these national resources, and strategies for unleashing the "data scientist" embodied in every researcher.

W1092: Tools, Resources, and Funding for Basic and Applied Research: What Early Career Scientists Need to Know

GETBIO-PGR: The Gateway for Education, Training, Broader Impacts and Outreach in Plant Genome Research

Carol Lushbough[1], Michael D. Gonzales[1] and Etienne Gnimpieba[2], (1)University of South Dakota, Vermillion, SD, (2)University of South Dakota, Sioux Falls, SD

GETBIO-PGR (https://getbiopgr.org) is a platform developed to enable NSF funded researchers to share the full range of their broader impact activities within a centralized, integrated infrastructure. GETBIO-PGR works towards creating an integrated Gateway for Education, Training, Broader Impacts and Outreach that provides infrastructure for researchers, educators, students, project managers, and the general public to access, create, and share information about broader impact activities and resources. This platform makes it possible to provide more rapid dissemination and more effective use of resources such as best practices for mentoring, workshop management, and data analysis along with an increase in peer-to-peer interactions at all levels, from undergraduate students to faculty. Additionally, GETBIO provides the ability to advertise broader impact activities and opportunities (e.g. workshops/training, internships, Post Doc opportunities) and project information through customizable web pages that can include project descriptions, images, videos, links, social networks, maps, and other useful tools/widgets.

W1093: Translational Genomics: Redesigning genetic improvement projects

The Role of Genomic Information in Quantifying the Trade-Offs between Genetic Gain and Genetic Diversity

William D. Beavis, Iowa State University, Ames, IA

W1094: Translational Genomics: Redesigning genetic improvement projects

Managing Homozygosity and Diversity in Livestock

Christian Maltecca, NC State University, Raleigh, NC, Christine F. Baes, Centre for the Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada and Francesco Tiezzi, NCSU, raleigh, NC

The implementation of routine genotyping in livestock breeding systems has become a common practice as a tool to make more effective selection decisions leading to increased genetic progress. The use of genomic information exploits the variation in relationships due to the stochastic nature of recombination and can be used to obtain measures connected to the homozygous proportion of an individual's genome. An increasing number of animals are now genotyped in breeding programs, and as such, genomic methods that allow for greater control of the diversity and inbreeding should be investigated. Different genome-wide metrics of inbreeding have been proposed to manage genetic diversity. For the most part their properties and the long-term consequences of their use are not clearly understood. One of the significant limitations of genome-wide metrics is that they do not account for the fact that the effective population size and regions that impact inbreeding depression are heterogeneous across the genome. The use of region-specific inbreeding metrics should allow to more effectively manage the risks associated with choosing a given mating pair and evaluating the trade-off between the genetic value of the progeny and undesirable side effects associated with inbreeding. Here we present results from both simulated data and from swine and cattle programs aimed at partially elucidating the architecture of inbreeding depression for economically relevant traits, as well as the consequences of employing different measures to curtail inbreeding accumulation.

In most cases measures related to recent inbreeding (such as ROH metrics) are best as striking a balance between genetic gains and the accumulation of recessive load both for lethal as well as sublethal recessives, thus representing a viable solution for short-term genetic response and long-term fitness management. Identification of detrimental recessives remains a challenging task due to the complex structure of livestock populations and the small marginal effects or partial recessives alleles. Hybrid strategies aimed at curtailing the overall accumulation of homozygosity while placing additional emphasis on functional inbreeding could be devised. Besides, in silico prediction of detrimental variants and the use of sequence information could further guide the identification of specific variants to penalize.

W1095: Translational Genomics: Redesigning genetic improvement projects

Efficient Breeding with Rapid Genomic Optimal Contribution Selection

Gregor Gorjanc, R. Chris Gaynor and John M. Hickey, University of Edinburgh, The Roslin Institute, Edinburgh, United Kingdom

In this presentation we demonstrate the use of optimal contribution selection in a plant breeding program with rapid recurrent genomic selection. Genomic selection has a large potential for plant breeding. However, the implementation of genomic selection in plant breeding requires a change of breeding programs, in particular faster cycling of selected individuals. The most extreme and impactful change is rapid recurrent

genomic selection. While this change increases the rate of genetic gain, it increases reduction of genetic variation further still compared to standard breeding programs. We use optimal contribution selection to manage this issue.
The presentation will show the principles of optimal contribution selection, present a practical implementation in the program AlphaMate, and show its use within a simulation of a full-scale wheat breeding program. We show that in the context of a rapid recurrent genomic selection optimizing contributions increases long-term genetic gain by optimizing efficiency of converting genetic diversity into genetic gain through reducing the loss of genetic diversity and reducing the drop of genomic prediction accuracy with rapid cycling.

W1**096:** **Translational Genomics: Redesigning genetic improvement projects**
Accounting for Prediction Error Var-Covariances in Selection to Reduce Risk and the Loss of Genetic Diversity
Dorian Garrick, Massey University, Ruakura, New Zealand

W1**097:** **Translational Genomics: Redesigning genetic improvement projects**
The Return of a Genetic Algorithm for Trade-Offs in Genomic Selection
Vishnu Ramasubramanian, Iowa State University, Ames, IA

W1**098:** **Transposable Elements**
Genome-Wide Hypermethylation of Transposable Elements following Low Temperature Exposure in Maize
Clémentine Vitte, Zeineb Achour and Johann Joets, GQE - Le Moulon, Gif sur Yvette, France
Transposable elements (TEs) are major players in shaping genome structure. TE sequences are transcriptionally silenced by epigenomic modifications to limit the mutagenic potential of their transpositional activity. In particular, several DNA methylation pathways are responsible for TE silencing in the various chromosomal locations where TE reside. While DNA methylation is known to be modified by abiotic constraints, the extent to which it can be remodeled remains to be fully elucidated.
We show that low temperature triggers genome-wide hypermethylation in maize, mainly at transposable elements and centromeres. This hypermethylation is mediated by the parallel activation of multiple methylation pathways across chromosomes, to actively hypermethylate TEs in the various chromatin locations where they reside. This likely reflects the importance of taming transposable elements following an abiotic stress in maize, a species for which over 85% of the genome is constituted of transposable elements.

W1**099:** **Transposable Elements**
Inhibition of RNA Polymerase II Leads to the Stress-Dependent Mobilization of Retrotransposons in Plants
Michael Thieme, University of Zurich, Zurich, Switzerland
Retrotransposons (retroTEs) can increase in DNA copy numbers by the occasional stable integration of reverse transcribed RNA- intermediates into the genome. This copy and paste mechanism that is often triggered by (a)biotic stresses is known to induce epi/genetic diversity and is therefore one of the major drivers of genome evolution and stress-adaptation in plants. However, as transposition can also lead to detrimental mutations and genome instability, retroTE mobility in plants is normally controlled by sophisticated silencing pathways. DNA methylation and small RNA pathways have important functions in the defense against retroTEs. Even though retroTEs fully depend on transcriptional activity of the host RNA polymerases for their mobility, it was so far unclear whether RNA polymerase II (Pol II) is directly involved in repressing their activity.
Here we report on the importance of Pol II in the repression of retrotransposition in plants. We show that plants defective in Pol II activity lose methylation at repeat sequences and produce more extrachromosomal retroTE DNA upon stress. Further, we demonstrate that combined drug-mediated inhibition of both DNA methylation and Pol II activity efficiently mobilizes the *ONSEN* retrotransposon in heat-stressed wild type *Arabidopsis* plants. New copies of this retroTE stably integrated into the genome and remained immobile over subsequent generations. Using this approach, we generated plants with up to 75 novel *ONSEN* copies. Notably, the lines that had acquired novel insertions also displayed large environment-dependent phenotypic diversity substantiating the biological importance of Pol II in regulating retroTE-mediated evolution. Finally, we generalized our findings by demonstrating that the simultaneous inhibition of both DNA methylation and Pol II activity also reactivated a retroTE in rice. Overall these findings contribute to the elucidation of de-novo silencing of retroTEs and establish a novel highly conserved player in retroTE regulation in plants. In addition, the ability to mobilize retroTEs in wild type plants unlocks a so far sealed epi/genetic resource that could be harnessed for crop improvement and to study the evolutionary consequences of stress-induced retroTE- bursts in real-time.

W1**100:** **Transposable Elements**
Parental Genetic Distance and Transposable Element Load Influence Transposable Element Dynamics in Young *Nicotiana* Allotetraploids
Corinne Mhiri[1], Christian Parisod[2], Julien Daniel[1], Maud Petit[1], K. Yoong Lim[3], François Dorlhac de Borne[4], Ales Kovarik[5], Andrew Leitch[3] and Marie-Angèle Grandbastien[1], (1)Institut Jean-Pierre Bourgin, INRA Versailles, Versailles, France, (2)Ecological Genomics, Institute of Plant Sciences, Bern, Switzerland, (3)Queen Mary University of London, London, United Kingdom, (4)Imperial Tobacco Group, Bergerac, France, (5)Institute of Biophysics, Brno, Czech Republic
As postulated by Barbara McClintock, allopolyploidy (interspecific hybridization associated with genome duplication) would represent a genomic shock, involving genome changes driven by transposable elements. To explore this hypothesis, we compared the recently formed (< 1MYA) allotetraploids *Nicotiana arentsii*, *N. rustica* and *N. tabacum* with their progenitor diploids, and with synthetic hybrids.
We used SSAP (sequence-specific amplification polymorphism) to compare the dynamics of six transposable elements in these allopolyploids, their diploid progenitors, and in corresponding synthetic hybrids.
We show that element-specific dynamics in young *Nicotiana* allopolyploids reflect their dynamics in diploid progenitors. Transposable element mobilization is not concomitant with immediate genome merger, but occurs within the first generations of allopolyploid formation. In natural allopolyploids, such mobilizations correlate with imbalances in the repeat profile of the parental species, which increases with their genetic

divergence. Other restructuring leading to locus loss is immediate, non-random, and targeted at specific subgenomes, independently of cross orientation.
The observed correlation between transposable element mobilization in *Nicotiana* allopolyploids and quantitative imbalances in parental transposable element loads supports the genome shock hypothesis proposed by Barbara McClintock.

W1**101:** **Transposable Elements**
Exhaustive Analyses of Highly Repetitive Genomes (wheat, barley, powdery mildew)
Thomas Wicker, Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

W1**102:** **Transposable Elements**
Analysis of Polymorphic TE Insertions in Maize Reveals Family Specific Influences on Insertion Site Preference and Spreading of DNA Methylation
Jaclyn Noshay, University of Minnesota, St. Paul, MN
Maize is an important crop species with a complex genome organization. A majority (>64%) of its genome is comprised of transposable elements (TEs), most of which are highly methylated. These TEs are interspersed with unmethylated genic regions. We are interested in understanding the interactions between genes and transposons and the role DNA methylation plays in these interactions. These interactions could include how chromatin influences the insertion site preference of TEs and how TE insertions could influence the spreading of DNA methylation. The analysis of CG and CHG methylation levels flanking TE families reveals three patterns characterized by high, moderate, or decreasing methylation. The proportion of TIR families with the decreasing methylation pattern is significantly higher than that of LTR families. The TE families that display these three patterns also show differences for other chromatin modifications within the TE itself. The differences in these patterns could be due to variable insertion site preference or variability of the influence TEs have on presence of flanking DNA methylation. A comparison of the TE content in four de novo maize genome assemblies identifies >10,000 polymorphic TE insertions. The analysis of DNA methylation at haplotypes without the TE reveals that some families have a strong preference to land in lowly methylated regions while others have a strong preference for highly methylated regions. For insertions that have occurred at unmethylated regions we can assess haplotypes containing the TE to determine whether DNA methylation spreads to previously unmethylated flanking regions. Many of the TE insertions are associated with DNA methylation spreading with a higher frequency of spreading observed for LTRs compared to TIRs. These analyses reveal a complex interplay between pre-existing chromatin state and the influence a TE exerts on nearby chromatin.

W1**103:** **Transposable Elements**
DNA Methylation of the Transposable Elements Is Reconfigured at the Onset of the Reproductive Phase in the Rice Shoot Apical Meristem
Hiroyuki Tsuji, Kihara Inst. Biol. Res., Yokohama City Univ., Yokohama, Japan
In mammals, DNA methylation is reprogrammed in the zygote and the germline. In flowering plants, while DNA methylation is also largely altered in germ cells and embryo, the entire picture of reconfiguration remains elusive because the germline is not clearly defined. Here, we found that DNA methylation is reconfigured in the rice shoot apical meristem (SAM), a pluripotent stem cell tissue. Methylation at CHH sites is kept high at the leaf-forming vegetative SAM and increases particularly at transposable elements in the inflorescence- and flower-developing reproductive SAM, which finally produces germ cells. Analyses of the transcriptome, small RNA transcriptome, and proteome of the SAM revealed that this reconfiguration is shaped via the RNA-dependent DNA methylation pathway. DNA methylation reconfiguration at transposons in the SAM partially overlapped with regions undergoing reprogramming in egg cells. Our results suggest that reconfiguration of DNA methylation beggins in the SAM long before germ cell differentiation.

W1**104:** **Tripal Database Network and Initiatives**
Tripal 3: Ontology-Driven Content, Flexible Data Storage and Data Exchange
Lacey-Anne Sanderson, Dept of Plant Sciences/University of Saskatchewan, Saskatoon, SK, Canada
Tripal is a freely available, open-source construction toolkit for community-focused, biological data web portals. It uses the GMOD Chado database schema, a community-derived relational database schema, and Drupal, a popular content management system (CMS), to provide data-specific content pages, data loaders and simple search tools. In response to the increase in data magnitude, Tripal 3 provides more flexible data storage to allow direct use of flat file formats (e.g. BAM, VCF) and no-SQL. Additionally, the Tripal Importer class improves the efficiency of many built-in loaders, as well as, makes development of efficient loaders easier in the future. Furthermore, the desire to exchange and share data has intensified among biological data web portals with the need to reduce data duplication. In response, Tripal 3 now provides intuitive ontology-driven content creation and RESTful web services with built-in exchange capabilities between Tripal web portals. To facilitate collaboration on development, Tripal 3 now provides modular class-based Tripal Fields allowing for exchange of displays and widgets, as well as, data. Tripal continues to respond to the needs of biological web portals with a community-wide discussion ongoing for Tripal 4.

W1**105:** **Tripal Database Network and Initiatives**
Tripal Developer Toolkit: Facilitating Development in Tripal
Joseph Benjamin West, University of Tennessee, Knoxville, Knoxville, TN

W1**106:** **Tripal Database Network and Initiatives**
The South Green Rice Genome Hub
Manuel Ruiz, CIRAD, UMR AGAP, Montpellier Cedex 5, France

W1**107:** **Tripal Database Network and Initiatives**

BIMS (Breeding Information Management System) in Tripal for Efficient Management and Analysis of Breeding Data

Sook Jung[1], Taein Lee[2], Chun-Huai Cheng[2], Ksenija Gasic[3], B. Todd Campbell[4] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Clemson University, Clemson, SC, (4)USDA-ARS, Florence, SC

Breeding programs produce large amount of data that require efficient management systems to keep track of performance, pedigree, geographical and image-based data as well as genotyping data. The integration of breeding data with publicly available genomic and genetic data, as well as the integration of each breeder's own genotypic and phenotypic data in a database enhances genetic understanding of important traits and maximizes the marker-assisted breeding utility by breeders and allied scientists. We report the progress on BIMS in Tripal which we have implemented in in Genome Database for Rosaceae, CottonGEN, Cool Season Food Legume Database and Genome Database for Vaccinium. BIMS is a Drupal module and designed to work with Chado schema. BIMS uses Dojo Toolkit, a javascript library designed for rapidly creating JavaScript/Ajax-based websites and cross-platform applications. BIMS allows individual breeders to integrate their phenotypic and genotypic data with public genomic and genetic data and at the same time have complete control of their own breeding data and access to tools such as data import/export, data analysis and a data archive. BIMS incorporates the use of an Android App called Field Book, an open-source software for phones and tablets, which will allow breeders to replace hard-copy field books, thus alleviating the possibility of transcription errors while providing faster access to the collected data. The use of Field Book and BIMS promotes the use and development of standard trait descriptors and metadata as well. The current functionality includes manage breeding, data import, search and download and statistical analysis.

W1**108: Tripal Database Network and Initiatives**

Execution of Scientific Workflows for Tripal-Based Community Databases

Shawna Spoor[1], Connor Wytko[1], Ming Chen[2], Abdullah Almsaeed[2], Bradford Condon[3], Heidi Hough[1], Nic Herndon[4], Nick Mills[5], Margaret Staton[2], Jill L. Wegrzyn[4], Alex Feltus[6], Dorrie Main[1] and Stephen P. Ficklin[7], (1)Washington State University, Pullman, WA, (2)University of Tennessee, Knoxville, Knoxville, TN, (3)University of Tennessee, Knoxville, TN, (4)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (5)Electrical & Computer Engineering, Clemson University, Clemson, SC, (6)Clemson University, Clemson, SC, (7)Dept of Horticulture, Washington State University, Pullman, WA

Tripal is an open source toolkit for the construction of biological community databases that facilitate display and searching of genomic, genetic and breeding data. Advances in technology have seen the size of datasets and the need for more complicated analytical workflows grow. Therefore, it is an increasingly complex task the for the community databases to meet the computational demands of their users. The Tripal Galaxy extension module allows Tripal-based sites to use the Galaxy project workflow software to provide complex analytical tools for their end-users. The module provides an easy to use web interfaces for execution of complex analyses using the same look-and-feel as the site but powered by Galaxy behind-the-scenes. The module can be used in two ways. First, it can automatically generate a simple step-by-step web form for workflows on an existing Galaxy server, within an easy to understand graphical user interface. Second, site developers with programming experience can use the Application Programming Interface (API) to interact with Galaxy. Either approach gives end-users executable workflows within the Tripal site and site administrators additional administrative controls, and usage statistics. In addition, Aurora Galaxy Tools, is an R Markdown framework that can be included with Galaxy workflows to produce reports that summarize analysis results. It outputs reports as HTML documents with rich visualizations. A workflow developer can use Aurora Galaxy Tools to create workflows which when integrated with Tripal provides meaningful analysis reports.

W1**109: Tripal Database Network and Initiatives**

Beyond Search and Display - Analyze Tripal Data with CartograTree

Sean Buehler, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Studying the interaction between the phenotype, genotype, and environment currently follows a workflow similar to this: (1) identify the sources of public data, (2) download the data, (3) extract the environmental values at the geo-referenced locations where phenotypic and genotypic data was sampled, (4) upload the data to a high performance computer required for timely analysis, (5) analyze the data, and (6) generate charts to display the results. CartograTree proposes a more user-friendly workflow that takes advantage of the Tripal platform, and insulates the user from the details of extracting environmental variables and interacting with a Linux server. With CartograTree, a user (1) queries and selects the data to be analyzed using its web interface, (2) instructs CartograTree to submit these data for analysis to a Galaxy server, and then (3) evaluates the analysis results shown by CartograTree as an overlay on the map.

CartograTree is developed as a Drupal module to enable the seamless integration with Tripal, the toolkit for construction of biological community databases, and to take advantage of some of the features provided by the latter. It uses genotypic and phenotypic data provided by four clade organism Tripal databases – TreeGenes (23,860 trees from 36 species), Hardwood Genomics, Citrus Genome Database, and Genome Database for Rosaceae – complemented with environmental data from WorldClim, and other public repositories. Although designed for the analysis of forest trees, this Drupal module can be installed and used to analyze data housed in other Tripal databases.

W1**110: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**

A Candle in the Dark: A Reference Genome Assembly for Rye Highlights the Importance of Data Visualisation and Manual Editing

M. Timothy Rabanus-Wallace, Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Gatersleben, Germany

Rye is of great interest to the cereal crop community, owing to its remarkable tolerance of poor soils and harsh environmental conditions. A rye genome will help scientists understand the genetics underlying these desirable traits—knowledge that can be applied to related crops such as wheat and barley. However, assembling the rye genome has posed a major challenge owing to its large size, repetitive content, and inbreeding depression. A suite of modern technologies (such as molecule-linked reads, chromosome-conformation-capture sequencing, and optical mapping) help to overcome many of these challenges, but integrating the data from very diverse technologies is difficult to fully automate with optimal results. Several years of work by an international consortium of institutions has produced a new reference quality genome for rye, which was completed following the philosophy that the results of automated procedures are best taken as suggestions, to be carefully refined by a

human curator with access to an array of intuitive visualisations. Such close curation can markedly increase the quality of a genome assembly, and visually-intuitive representations of a genome assembly (and its relationship to the underlying data), are similarly valuable for those using the genome for downstream applications.

W1**111:** **Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**
Comparative Genomics of Wheat at an Unprecedented Scale
Curtis Pozniak, University of Saskatchewan, Saskatoon, SK, Canada

W1**112:** **Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**
An Improved Reference Sequence Assembly for Barley Cv. Morex
Cécile Monat, Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, Germany
The completion of a chromosome-scale reference sequence for barley cultivar Morex marks a milestone for barley genomics. A possible direction for future research in barley genomics is the assembly of genome sequence for representatives of global barley diversity. An important component in this endeavour will be a reproducible, open-source pipeline for sequence assembly, enabling fast and cost-effective assembly of a large number of genotypes. Here we present a new chromosome-scale assembly of Morex constructed with paired-end, mate-pair, Chromium 10X and Hi-C sequencing data. This updated reference genome has been obtained through a complete open-source pipeline that will also be presented here. The gene space completeness among the two versions of Morex assembly, validation by optical maps and Dovetail *In vitro* Hi-C data have been performed and will be shown in the presentation. In the context of pan-genomic studies in barley, we will also present first results obtained for four other genotypes among them one wild barley.

W1**113:** **Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**
Tilling Even More(x) Barley
Silvio Salvi, DipSA - University of Bologna, Bologna, Italy

W1**114:** **Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics**
Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference Genome Sequence to Wheat Genetic and Phenomic Data
Michael Alaux, URGI, INRA, Université Paris-Saclay, Versailles, France
URGI is a genomics and bioinformatics research unit at INRA (French National institute for Agricultural Research), dedicated to plants and crop parasites. We develop and maintain a genomic and genetic Information System called GnpIS that manages multiple types of wheat data.
The Wheat@URGI portal has been developed to provide the international community of researchers and breeders with access to the bread wheat reference genome sequence produced by the International Wheat Genome Sequencing Consortium.Genome browsers, BLAST, and InterMine tools have been established for in-depth exploration of the genome sequence together with additional linked datasets including physical maps, sequence variations, gene expression, and genetic and phenomic data from other international collaborative projects already stored in the GnpIS information system. The portal provides enhanced search and browser features that will facilitate the deployment of the latest genomics resources in wheat improvement.
Ref: Alaux *et al.*, *Genome Biology* 2018, https://doi.org/10.1186/s13059-018-1491-4

W1**115:** **Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**
A Tandem Kinase-Pseudokinase Protein Family Involved in Plant Immunity Revealed by Exploration of Wheat Reference Genomes
Valentina Klymiuk, Andrii Fatiukha and Tzion Fahima, Institute of Evolution and the Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel
The stripe rust resistance gene *Yr15,* derived from wild emmer wheat, encodes a putative kinase-pseudokinase protein, designated as Wheat Tandem Kinase 1 (WTK1) (Klymiuk et al., 2018; Nature Communications). The available reference wheat genomes, Chinese Spring, Svevo and Zavitan, have paved the way for the discovery of a unique protein family. *WTK1* orthologs and paralogs are found in all group 1 and 6 wheat chromosomes. The exon-intron structure of orthologues copies is similar to that of *Wtk1* from G25, but differ in numerous SNPs and indels that cause changes in reading frames. Although exon-intron structure of the paralogues copies (6A, 6B and 6D) is similar to *Wtk1,* the total number of exons was increased from six to seven due to the split of exon 4 into two.
The unique gene architecture of *WTK1* was found in 92 putative proteins across the plant kingdom, including the barley RPG1 and a candidate for *Un8*, suggesting that they are members of a distinct family of plant proteins, termed here tandem kinase-pseudokinases (TKPs). We found that 175 out of 184 kinase/pseudokinase domains of these TKPs were associated with receptor-like kinases (RLKs), suggesting that TKPs are involved in plant defense mechanisms. A further phylogenetic analysis indicated that TKP family members originated from either gene duplication or gene fusion events, suggesting a polyphyletic origin of the TKPs. The presence of kinase-pseudokinase structure in plant TKPs and animal Janus kinases (JAKs), is suggesting convergent molecular evolution of proteins involved in immunity in both of the kingdoms.

W1**116:** **Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**
Rapid Cloning of Disease Resistance Genes by Association Genetics on a Sequence-Configured Wild Wheat Diversity Panel
Kumar Gaurav, John Innes Centre, Norwich, United Kingdom

W1**117:** **Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**
Rapid Cloning of Rust Resistance Gene *Rph1* from Barley Cultivar Sudan
Peter Dracatos, University of Sydney, Sydney, Australia

Unravelling and exploiting mechanisms of disease resistance in cereal crops is currently limited by their large repeat-rich genomes and the lack of recombination or cultivar-specific sequence information. We cloned the first gene conferring resistance to leaf rust (Rph1) from barley using 'MutChromSeq', a recently developed molecular genomics tool for the rapid cloning of genes in plants. In this study we combined MutChromSeq with genetic mapping to rapidly clone Rph1 in barley cultivar (cv.) Sudan from a defined region on chromosome 2H. We also report on developing a cost-effective wild type sequence assembly with high contiguity (contig N50 >20.1 kb) and gene space representation (83% of genes with ≥90% query coverage).
Marker trait association in the CI 9214/Stirling doubled haploid population mapped Rph1 on the short arm of chromosome 2H to a physical region of 1.3 Mb relative to the barley cultivar Morex reference assembly. A sodium azide mutant population in cultivar Sudan was generated and 10 mutants were confirmed by progeny-testing. Flow-sorted chromosomes for 2H from Sudan (wild type) and six of the mutants were sequenced and compared to identify candidate genes for Rph1. MutChromSeq identified a single gene candidate encoding a coiled-coil NLR receptor protein that was altered in three different mutants. Further Sanger sequencing confirmed all three mutations and identified an additional two independent mutations within the same candidate gene. Phylogenetic analysis determined that Rph1 clustered separately from all previously cloned NLRs from Triticeae and was most similar (89%) to an RPM1-like protein from Triticum urartu.

W1**118:** **Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**
A Major QTL for Grain Weight in Wheat Is Associated with Increased Grain Length and Cell Size
Jemima Brinton, John Innes Centre, Norwich, United Kingdom
TBA

W1**119:** **Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**
Dwarfing Genes in Wheat and Barley: All about GA
Wolfgang Spielmeyer, CSIRO Agriculture & Food, Canberra, Australia

W1**120:** **Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification**
***LATE ELONGATED HYPOCOTYL1*, a Circadian Clock Gene Homolog, Coordinating with *TaPHS1* Regulates Pre-Harvest Sprouting Resistance in Wheat**
Hui Chen, Agronomy Department, Kansas State University, Manhattan, KS
Pre-harvest sprouting (PHS) causes serious reduction in grain yield and quality of wheat, and results in significant economic losses. Inadequate seed dormancy is the key factor for PHS when matured wheat spikes experience a long period of wet weather before harvest. Many quantitative trait loci (QTLs) for PHS resistance have been reported and two of them on the chromosome arms 3AS (*TaPHS1*) and 4AL (*TaMKK3-A*) have been cloned. However, the molecular mechanism of PHS resistance in wheat remains unknown. Previously, we cloned the *TaPHS1* gene as a key determinant of PHS resistance in white wheat. Here, we conducted RNA-seq and yeast-two hybrid (Y2H) analysis and found that a wheat circadian clock gene homolog, *LATE ELONGATED HYPOCOTYL*1 (*TaLHY1*), coupling with*TaPHS1* regulates PHS resistance in wheat. *TaLHY1* showed strong interaction with *TaPHS1* as demonstrated by planta bimolecular fluorescence complementation (BiFC) assays. Knocking down of *TaLHY1* using *Barley Stripe mosaic virus* (BSMV)-mediated gene silencing caused a significant change in seed dormancy, germination rate, and plant responses to exogenous hormones. Taken together, those data suggest that *TaLHY1* might play an important role in wheat seed dormancy and PHS regulation through direct interaction with *TaPHS1*. This work provides new insight into understanding molecular mechanisms of PHS resistance in wheat, in particular the role of a clock gene in regulation of PHS resistance.

W1**121:** **Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement**
Reconciling the Evolutionary Origin of Bread Wheats
Jérôme Salse, GDEC, INRA, Université Clermont Auvergne, Clermont-Ferrand, France
Jerome Salse[1], Caroline Pont[1], Thibault Leroy[2,3], Michael Seidel[4], Alessandro Tondelli[5], Wandrille Duchemin[1], David Armisen[1], Daniel Lang[4], Daniela Bustos-Korts[6], Nadia Goué[1], François Balfourier[1], Marta Molnar-Lang[7], Jacob Lage[8], Benjamin Kilian[9,10], Hakan Özkan[11], Darren Waite[12], Sarah Dyer[13], Thomas Letellier[14], Michael Alaux[14], WHEALBI consortium[15], IWGSC[16], Joanne Russell[17], Beat Keller[18], Fred van Eeuwijk[6], Manuel Spannagl[4], Klaus F.X. Mayer[4,19], Robbie Waugh[17,20], Nils Stein[20], Luigi Cattiveli[5], Georg Haberer[4], Gilles Charmet[1]
1. INRA-Université Clermont-Auvergne, UMR 1095 GDEC, 5 Chemin de Beaulieu, 63000 Clermont-Ferrand. France.
2. INRA-Université de Bordeaux, UMR 1202 BIOGECO, 69 Route d'Arcachon, 33612 Cestas. France
3. ISEM, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier, France
4. PGSB, Helmholtz Center Munich, Ingolstädter Landstraße 1 · D-85764 Neuherberg.
5. Council for Agricultural Research and Economics (CREA), Research Centre for Genomics and Bioinformatics, via S. Protaso, 302. I -29017 Fiorenzuola d'Arda PC. Italy.
6. Wageningen University, Droevendaalsesteeg 4, 6708 PB, Wageningen. Netherlands.
7. Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, Brunszvik u. 2, 2462. Hungary.
8. KWS UK Ltd, 56 Church St, Thriplow, Royston SG8 7RE. United Kingdom
9. Global Crop Diversity Trust, 53113 Bonn, Germany.
10. Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Gatersleben. Germany.
11. University of Çukurova, Faculty of Agriculture, Department of Field Crops, 01330 Adana.Turkey.
12. **Earlham Institute,** Norwich Research Park, Norwich NR4 7UZ, UK.
13. NIAB, Huntingdon Road, Cambridge CB3 0LE. United Kingdom.
14. URGI, INRA, Université Paris-Saclay, 78026, Versailles, France.
15. **WHEA**t and barley **L**egacy for **B**reeding **I**mprovement (WHEALBI) consortium, http://www.whealbi.eu/project/wheat-consortium/. *Coordinator*, Gilles Charmet (gilles.charmet@inra.fr), INRA UMR 1095 GDEC. France.

16. International Wheat Genome Sequencing Consortium (IWGSC), http://www.wheatgenome.org. *Executive Director*, Kellye Eversole (eversole@eversoleassociates.com), 2841 NE Marywood Ct Lee's Summit, MO 64086. United States.
17. The James Hutton Institute, Invergowrie Dundee DD2 5DA, Scotland.
18. Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zürich.
19. School of Life Sciences, Technical University Munich, Weihenstephan, Germany.
20. The University of Dundee, Division of Plant Sciences, School of Life Sciences, Dundee, DD1 4HN.
The origin of bread wheat (*Triticum aestivum*) has been a subject of intense investigation. Access to reference genome sequences for diploid (AA, DD), tetraploid (AABB) and hexaploid (AABBDD) wheats, together with exome sequencing data from world-wide panels from all across the geographical ranges of the wheat species complex, allowed us to investigate the ancestry of modern bread wheats. *We* explored the impact hybridization, domestication, selection and adaptation at the genic, chromosomal and subgenomic levels, underlying the genetic makeup of modern bread wheats and the structural asymmetry observed between the A, B and D subgenomes.

W1122: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Chromatin Accessibility and Balanced Gene Expression in Diploid and Hexaploid Triticeae
Michael Bevan, John Innes Centre, Norwich, United Kingdom
The wheat group comprises stably maintained tetraploid and hexaploid species that exhibit greater diversity, adaptability and potential for domestication than their diploid progenitors. Multiple types of genomic changes, including altered expression patterns of genes and transposable elements (TE), and epigenetic changes, have been proposed to contribute to this “genomic plasticity” of wheat. Nevertheless, little is known about mechanisms of genomic interactions that may contribute to improved traits in wheat or any other polyploid plant species. The bread wheat genome is a potentially informative experimental system for studying the interactions between closely related homoeologous loci as they form a relatively stable genomic framework. Recent progress in assembling and analysing Triticeae genomes provides new opportunities for detailed spatial analyses of gene expression of Triticeae chromosomes in diploid and polyploid contexts. Our analyses of gene expression in homoeologous chromosome arms of *Ae. tauschii* and bread wheat show a chromosome-arm wide pattern of 70% balancing gene expression and multiple instances of new patterns of tissue-specific expression, patterns consistent with whole genome analyses. Only a small proportion of these major differences was related to altered patterns of gene methylation. In contrast, we found major differences in patterns of chromatin accessibility in leaf tissues of *Ae tauschii* and hexaploid wheat. There was a chromosome arm- wide reduction in chromatin accessibility in hexaploid nuclei that was generally limited to promoter regions. In contrast, ATAC peaks were generally wider in diploid nuclei indicating a more accessible chromatin state. 83% of balanced genes showed reduced chromatin accessibility in hexaploid nuclei, consistent with reduced expression. Our analyses suggest new mechanisms that function rapidly to establish balanced gene expression in hexaploid wheat.

W1123: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Two New Sequencing-Based Genotyping Platforms that enable Cost Effect Variant Discovery, Saturation Mapping and Genomic Selection in Wheat
David Konkin, National Research Council Canada, Saskatoon, SK, Canada

W1124: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Unraveling Transcription Factor Networks that affect Yield in Barley
G.W. van Esse, Wageningen UR, WAGENINGEN, Netherlands

W1125: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
A Retrospective Look at the Impact of Genomics in the Triticeae for Crop Improvement
Catherine Feuillet, Inari Ag, Cambridge, MA

W1126: UCSC Genome Browser - a home for all organisms
UCSC Genome Browser - A Home for All Organisms
Robert Kuhn, University of California Santa Cruz, Santa Cruz, CA
The UCSC Genome Browser is a full-featured display engine for genomic annotations. Users may load their own data into the Browser and take advantage of the many visualization and analysis tools. The interface is consistent across all genome assemblies.
Users may upload their own data, such as RNA-seq, and view at all scales in a highly configurable representation. A new feature, Track Collections, allows multiple data sets to be configured together by setting parameters in one place. Track Collections also allow datasets to be shown together on a single axis with multiple colors in transparent overlay mode, so that at a glance, the dataset with the highest value will stand out in its color. Two tracks can be also be subtracted from each other, allowing display of the differences between treated and untreated, for example, on the same axis.
The assembly hub mechanism, described in the second half of the workshop, enables those with reference assemblies not hosted by UCSC, such as plants, to access these same features and all other features of the Genome Browser.

W1127: UCSC Genome Browser - a home for all organisms
UCSC Browser Assembly Hubs
Brian Lee, UC 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 Gene

Predictions (bigGenePred files), Multiple Alignments (bigMaf files), Pairwise Alignments (bigPsl files), Pairwise Interactions (bigInteract files) and a number of other data types, including, but not limited to bigWig, bigBed, bigBarChart, bigChain, VCF and BAM files.

W1**128:** **US National Animal Genome Research Program (NRSP8)**

The Enigmatic Role of Host Genetics in Disease Susceptibility: A Case Study of Porcine Circovirus 2

Lianna R Walker[1], Taylor Engle[1], Hiep Vu[1], Emily Tosky[1], Dan Nonneman[2], Timothy P.L. Smith[2], Tudor Borza[3], Thomas Burkey[1], Graham S. Plastow[4], Stephen D. Kachman[1] and **Daniel Ciobanu**[1], (1)University of Nebraska - Lincoln, Lincoln, NE, (2)USDA, ARS, USMARC, Clay Center, NE, (3)Dalhousie University, Truro, NS, Canada, (4)Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada

Porcine circovirus 2 (PCV2), the smallest known mammalian virus, is a circular single-stranded DNA virus responsible for a group of diseases collectively known as PCV2 Associated Diseases (PCVAD). Variation in the incidence and severity of PCVAD exists between pigs suggesting a host genetic component involved in pathogenesis. A genome-wide association study based on experimental infection (n=974), provided evidence of a host genetic role in PCV2 viremia, immune response and growth during challenge. Host genotype explained 64% of the phenotypic variation for overall viral load, with two major QTLs identified on chromosome 7 (SSC7) near the SLA class II locus and on the proximal end of chromosome 12 (SSC12). Dissection of the SSC12 QTL based on gene annotation, genomic and RNA-sequencing, suggested that a missense mutation in the *SYNGR2* (*SYNGR2 p.Arg63Cys*) gene is potentially responsible for the variation in viremia. PCV2 titer in PK15 cells decreased when the expression of *SYNGR2* was silenced by specific-siRNA or in a PK15 edited clone carrying a partial deletion of the second exon that harbors a key domain and the *SYNGR2 p.Arg63Cys*. Identification of a non-conservative substitution in this key domain of *SYNGR2* suggests that the *SYNGR2 p.Arg63Cys* substitution may underlie the observed genetic effect on viral load. The knowledge generated by this research will result in a reduction in susceptibility to PCV2 and improvement in the general health and welfare of pigs. USDA is an equal opportunity provider and employer.

W1**129:** **US National Animal Genome Research Program (NRSP8)**

Nrsp-8 Distinguished Lecturer

Ernest Bailey, University of Kentucky, Lexington, KY

W1**130:** **US National Animal Genome Research Program (NRSP8)**

Development of a Universal Sex Assay and Identification of y-Chromosome Haplotypes in Chinook Salmon

Garrett J McKinney, University of Washington, Seattle, WA

W1**131:** **US National Animal Genome Research Program (NRSP8)**

A Large Multi-Breed Reference Panel Maximizes Genotype Imputation Accuracy in Cattle

Troy N. Rowan, Division of Animal Science, University of Missouri, Columbia, MO, Robert D. Schnabel, Division of Animal Sciences, Informatics Institute, Columbia, MO and Jared E. Decker, Division of Animal Sciences, University of Missouri, Columbia, MO

W1**132:** **US National Animal Genome Research Program (NRSP8)**

Exploiting the Dimensionality of Genomic Information in Channel Catfish

Ivan Pocrnic[1], Daniela A.L. Lourenco[1], Andre Garcia[1], Geoff Waldbieser[2], Brian Bosworth[2] and Ignacy Misztal[1], (1)University of Georgia, Athens, GA, (2)USDA-ARS Warmwater Aquaculture Research Unit, Stoneville, MS

Genomic selection (GS) is successfully applied in animal breeding based on single-nucleotide polymorphism (SNP) markers. While the number of SNP can be very large, especially with sequence data, the dimensionality of gene content is limited by limited effective population size and subsequently by limited number of the independent chromosome segments. When number of SNP markers and/or number of genotyped animals is large enough, it is possible to estimate all chromosome segments precisely, leading to perfect accuracy of GS. In that case, the dimensionality can be obtained via the number of non-negligible singular values of gene content, or the number of non-negligible eigenvalues of the genomic relationship matrix (GRM) that explain 98% of the variation. When a small amount of data is available, the "98%" is depressed and 4 times the "90%" number is a better estimate. In channel catfish, the available genomic information consisted of 2911 fish genotyped for 57k SNP. Interpolated number of eigenvalues explaining 10, 50, 80, 90 and 98% of variation in GRM were 3.8, 71.9, 571.2, 1100.9 and 2140.9. Subsequently, the number of independent chromosome segments is around 4,400. Knowing the dimensionality of genomic information can be a useful tool to estimate the optimal number of SNP markers and/or genotyped animals needed for GS and variant discovery.

W1**133:** **US National Animal Genome Research Program (NRSP8)**

Functional Investigation of Putative Variant for Atypical Equine Thrombasthenia in Thoroughbreds

Anna R. Dahlgren, University of California - Davis, Davis, CA

Atypical Equine Thrombasthenia (AET) is a frequent cause of bleeding in Thoroughbreds, affecting one in every 150 horses. Aberrant cell signaling after thrombin stimulation prevents platelets from efficiently binding to fibrinogen, leading to abnormal bleeding. Affected Thoroughbreds commonly experience epistaxis during racing and prolonged bleeding after a vascular injury. Despite the negative effect on horse health and performance, the underlying etiology of AET is unknown, though pedigrees of affected horses indicate that AET is heritable. A whole genome association study using six affected and twelve control Thoroughbreds identified an associated 1.6 kilobase deletion within a long non-coding RNA (lncRNA) upstream of *suppressor/enhancer of lin-12-like (SEL1L)*. When *SEL1L* is knocked down in zebrafish, vascular leakage occurs leading to blood pooling. To investigate the putative role of this deletion in AET, the expression of the lncRNA and *SEL1L* was determined. The lncRNA expression was confirmed by isolating RNA from equine testes, the tissue with the highest expression in humans. After reverse transcription and PCR amplification, Sanger sequencing the product confirmed the lncRNA is expressed in horses. *SEL1L* expression

was also confirmed in the equine platelets. In a preliminary study, the protein expression of SEL1L in platelets was investigated using Western Blot in two affected horses and one unaffected horse. Both affected horses had lower levels of SEL1L protein than the control horse, suggesting this protein plays a role in the etiology of AET. Elucidation of the mechanism of AET in Thoroughbreds can lead to improved health and performance of the breed.

W1**134:** **US National Animal Genome Research Program (NRSP8)**
Functional Roles of Paternally-Linked mRNAs in Embryo Development in Bovine
Nicole Lee Gross, University of Wisconsin-Madison, Madison, WI

W1**135:** **US National Animal Genome Research Program (NRSP8)**
Genome-Wide Association Study to Identify Genetic Loci Associated with Resistance to Haemonchus Contortus in Katahdin Sheep
Gabrielle M. Becker, Department of Animal and Veterinary Science, University of Idaho, Moscow, ID
Haemonchus contortus is the most abundant gastrointestinal nematode (GIN) observed in small ruminants. Additionally, *Haemonchus contortus* has the highest prevalence of anthelmintic resistance among GIN, making it a crucial economic concern for sheep producers (Fleming et al., 2006). Previous research has shown that in at least some breeds of sheep, including Dorper, St. Croix, and Katahdin, GIN resistance is heritable (Burke and Miller, 2004; Kemper et al., 2011). A genome-wide association study was conducted to identify genetic loci associated with resistance to *Haemonchus contortus* in Katahdin sheep. Forty sheep were selected for high and low fecal eggs count estimated breeding values and genotyped using the Affymetrix Ovine 50K array. Following quality control, 46,268 single nucleotide polymorphisms (SNP) were included in subsequent analyses using a linear regression model in PLINK v1.90 and a Single-Locus Mixed Model in SNP and Variation Suite. Significance was determined using a Bonferroni correction for multiple testing ($p < 0.05$). A total of nine significant SNPs across chromosomes 2, 3, 16, 23, and 24 were identified, with one SNP on chromosome 2 reaching significance in both models. The linear regression model identified three nearby significant SNPs on chromosome 2, suggesting that this region is in linkage disequilibrium and may be related to GIN resistance in Katahdin sheep. This study identified genetic regions associated with GIN resistance that may eventually be used to help sheep producers select for GIN resistance in their flocks.

W1**136:** **US National Animal Genome Research Program (NRSP8)**
Host Genetics Influences Replication of Porcine Circovirus 2b
Lianna Walker[1], Taylor Engle[2], Hiep Vu[2], Emily Tosky[2], Dan Nonneman[3], Timothy P.L. Smith[3], Tudor Borza[4], Thomas Burkey[2], Graham S Plastow[5], Stephen D. Kachman[2] and Daniel C. Ciobanu[2], (1)UNL, Lincoln, NE, (2)University of Nebraska - Lincoln, Lincoln, NE, (3)USDA, ARS, USMARC, Clay Center, NE, (4)Dalhousie University, Truro, NS, Canada, (5)University of Alberta, Edmonton, AB, Canada
Porcine Circovirus 2 (PCV2) is the smallest known DNA virus capable of infecting mammals and the primary agent required for the development of a set of symptoms collectively known as Porcine Circovirus Associated Diseases. However, infection with PCV2 does not guarantee clinical disease with variation in severity observed between breeds and individuals. Experimental infection of ~1,000 pigs with PCV2b and genome-wide association analyses revealed two QTL for PCV2b viral load, located on SSC7 and SSC12. Dissection of the SSC12 QTL using *ab initio* gene prediction, RNAseq, and genomic sequencing identified 66 novel polymorphisms within and upstream of five positional candidate genes. Single marker association analysis of a subset of pigs with extreme high and low viral loads, identified two novel polymorphisms in high LD accounting for 21-23% of the phenotypic variation. One polymorphism is a missense mutation (*p.Arg63Cys*) within the second exon and critical domain of the *SYNGR2* gene. *In vitro* silencing of *SYNGR2* in PK15 cells via siRNA transfection resulted in a one-log reduction in PCV2b titer ($P<0.05$) compared to scramble siRNA and non-transfected control cells, indicating a direct role of this gene in PCV2b infection. Additionally, gene editing using CRISPR-Cas9 ribonucleoprotein complexes targeting *SYNGR2* generated a PK15 edited clone homozygous for a 106bp deletion within the second exon that exhibited a two-log reduction in PCV2b titer following infection compared to wild-type PK15 cells ($P<0.05$). Given these findings and that *SYNGR2 p.Arg63Cys* is the only missense mutation within this gene, *SYNGR2 p.Arg63Cys* is a plausible QTN for PCV2b susceptibility.

W1**137:** **US National Animal Genome Research Program (NRSP8)**
Predicting Chromatin States to Identify Distinct Active Enhancers within Bursa Tissue of Two Inbred Chicken Lines Under Ndv Infection and Heat Stress
Ganrea Chanthavixay, Animal Science, University of California, Davis, CA

W1**138:** **US National Animal Genome Research Program (NRSP8)**
Weaning Induces Rapid DNA Methylation and Transcriptional Changes in Piglet PBMCs Associated with Impaired Stress Response
Ryan J. Corbett, Genetics Graduate Program, Michigan State University, East Lansing, MI

W1**139:** **US National Animal Genome Research Program (NRSP8)**
Update on the Blueprint for Animal Genomics
Lakshmi Matukumalli, USDA-NIFA, Washington, DC

W1**140:** **US National Animal Genome Research Program (NRSP8)**
Plan for on-Going Support of NRSP-8

Molly E McCue, University of Minnesota, St. Paul, MN

W1**141:** **US National Animal Genome Research Program (NRSP8)**
NRSP-8 Business Meeting
Alison Van Eenennaam, University of California, Davis, Davis, CA

W1**142:** **Weedy and Invasive Plant Genomics**
Extrachromosomal DNA-Mediated Herbicide Resistance
Mithila Jugulam, Dal-Hoe Koo, Bernd Friebe and Bikram S. Gill, Kansas State University, Manhattan, KS
Evolution of herbicide resistance in weed species is a major constraint to crop production around the globe. Many agriculturally important weed species throughout the world have naturally evolved resistance to several major herbicides used in our agriculture. The investigation of physiological, genetic, and molecular mechanisms of weed resistance to herbicides have uncovered several novel, and exciting results related to fundamental, evolutionary mechanisms of herbicide resistance in weeds, specifically, regarding the evolution of resistance to glyphosate, one of the important herbicides used in crop production. With the introduction and wide acceptance of Roundup Ready crops in many countries, glyphosate has been used extensively for weed control, consequently, many weeds have developed resistance to glyphosate. The target site of glyphosate is 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), an important enzyme in shikimate pathway. Several types of mutations including amplification of *EPSPS* gene can bestow weed resistance to this herbicide. Recently, our molecular cytogenetic research indicated that the *EPSPS* gene amplification in glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*), one of the top problem weeds of the USA, was driven by extra-chromosomal, circular DNA (eccDNA) molecules. Each eccDNA carried one copy of the target gene *EPSPS*. However, freed from the rules of mitosis, EPSPS genes can multiply rapidly during the growth of the sporpohyte and produce copy number variation in somatic cells. The somatic cells with amplified *EPSPS* survive in the presence of the herbicide, and this acquired trait is transmitted to the germ cells and the progeny. Importantly, it appears that the eccDNA replicons are transmitted by an unknown mechanism of tethering to mitotic and meiotic chromosomes and modulate rapid glyphosate resistance response.

W1**143:** **Weedy and Invasive Plant Genomics**
Resistance Gene Directed Discovery of a Natural Product Herbicide
Qikun Liu, UCLA, Los Angeles, CA

W1**144:** **Weedy and Invasive Plant Genomics**
Distinguishing Characteristics of Herbicide Resistance in the *Amaranthus tuberculatus* (waterhemp) Transcriptome
Daniel Kohlhase[1], Jamie A. O'Rourke[2], Micheal D.K. Owen[1] and Michelle A. Graham[2], (1)Iowa State University, Ames, IA, (2)USDA-ARS-MWA-CICGRU, Ames, IA
Waterhemp (*Amaranthus tuberculatus* (Moq.) J.D. Sauer) is a problem weed commonly found in the Midwestern United States that causes crippling yield losses in major crop production systems. Since the discovery of *p*-hydroxyphenylpyruvate-dioxygenase (HPPD, EC 1.13.11.27) inhibitor herbicide resistance in waterhemp populations, studies have identified the likely mechanism of resistance and described its inheritance, however causal genes remain unknown. To date, no studies have examined genome-wide gene expression changes in response to HPPD herbicide treatment in herbicide-resistant and susceptible waterhemp. We developed a *de novo* waterhemp transcriptome from RNA-seq analyses of two waterhemp populations (HPPD-herbicide-resistant and susceptible), from herbicide-treated and mock-treated leaf samples collected three, six, twelve, and twenty-four hours after treatment (HAT). This allowed us to identify transcripts specific to a genotype, herbicide treatment, or time point. Our results indicate that both waterhemp genotypes responded rapidly to HPPD-inhibiting herbicide (within 3 HAT), long before a phenotype was observed. Little overlap in differentially expressed genes was observed between the resistant and susceptible genotypes, highlighting dynamic differences in response to herbicide treatment. Digital mapping of differentially expressed genes allowed identification of clusters of herbicide-responsive genes and suggests coordinated regulation. Stringent analytical methods identified candidate single nucleotide polymorphisms (SNPs) that distinguish the resistant and susceptible genotypes. The waterhemp transcriptome, herbicide-responsive genes and SNPs generated in this study provide valuable tools for future studies in waterhemp by numerous plant science communities. This collection of resources is essential to study and understand herbicide effects on gene expression in resistant and susceptible weeds.

W1**145:** **Weedy and Invasive Plant Genomics**
Introgression of Cultivar Genes into Wild Carrot Populations
Johanne Brunet, USDA-ARS, Madison, WI and Luciano Palmieri, ORISE-ARS, Madison, WI
Wild carrots are widespread in the USA, can be weedy and have been declared invasive in some states. Wild and cultivated carrots are commonly found in close proximity and can easily hybridize. Following their introduction into wild populations via hybridization, cultivar genes can spread both within and among populations, a process called introgression. The extent of cultivar gene introgression 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 increase the invasiveness of wild carrots.
We have previously compared the genetic diversity and genetic differentiation of four wild carrot populations located near and four located far away from cultivated carrots and found greater genetic differentiation and genetic diversity in the wild carrot populations in close proximity to cultivated carrots. This study also identified single nucleotide polymorphisms (SNPs) with great potential to detect introgression. In the current study, we further examine the pattern of introgression of cultivar genes into wild carrot populations. We sampled populations at incremental 300 meter distances along a line between one of the near and one of the far populations (near or far away from cultivars) to a distance up to 1800 meters. Leaf tissue was sampled from 20 individuals per population. This sampling process was repeated for three sets of near and far populations.

We extracted DNA from the leaf tissue and performed genotyping by sequencing (GBS). We identified single nucleotide polymorphisms (SNPs) from the combined set of wild populations. We used fastSTRUCTURE to examine the population genetic structure of these wild carrots. First, we only evaluated the four populations near and four populations far away from cultivars. Second, we analyzed the populations within each of the three 1800 meter lines between a near and a far populations; we analyzed each of the three lines separately. Lastly, we combined all populations. In addition, we calculated isolation by distance for populations at increasing geographic distances from cultivated carrots. We first used all SNPs identified and then only a subset of the SNPs predicted to be good detectors of introgression from a previous study. These results will increase our understanding of the pattern of spread of cultivar genes into wild carrot populations. This information could guide the design of methods to reduce the spread of cultivar genes into wild carrot populations and help prevent potential negative impacts such as an increased invasiveness of wild carrot populations.

W1**146: Weedy and Invasive Plant Genomics**

Distinguishing Among Native *vs.* Exotic Reed Canarygrass (*Phalaris arundinacea*) using GBS (DArTseqLD)

Andrzej Noyszewski[1], Neil O Anderson[2], Alan G. Smith[1], Andrzej Kilian[3], Diana Dalbotten[4], Emi Ito[5], Anne Timm[6] and Holly Pellerin[7], (1)University of Minnesota, Department of Horticultural Science, St. Paul, MN, (2)University of Minnesota, Department of Horticultural Science, Saint Paul, MN, (3)Diversity Arrays Technology Pty Ltd, Bruce, ACT, Australia, (4)University of Minnesota, St. Anthony Falls Laboratory,, Minneapolis, MN, (5)University of Minnesota, Department of Earth Sciences, Minneapolis, MN, (6)USDA Forest Service, Northern Research Station, Baltimore, MD, (7)University of Minnesota, St Anthony Falls Laboratory, Minneapolis, MN

Reed canarygrass (RCG, *Phalaris arundinacea*) is a wind-pollinated, wetland grass, cultivated in temperate regions around the globe as a forage and ornamental crop. It is also used for soil stabilization, bioremediation and bioenergy. RCG is considered invasive in N. America, particularly its exotic forms. Native vs. exotic status of RCG is not clear and past research suggests the existence of native RCG stands in N. America. Herbarium specimens (1889-1985), native in origin in N. America as well as comparative exotic samples from central Europe (Czech Republic), are used in the present study as a benchmark to determine current RCG populations' genetic background. The purpose of this study is to examine RCG wild MN and Czech populations along rivers, cultivars, N. American herbarium samples (n= 2,531) for their status to aid land managers in identifying native vs. exotic types. Genetic variation among and within RGC populations were assessed by DArTseqLD1.0 that produced 13,967 polymorphic SNPs. This is the first use of this technology in RGCs. Principal component and STRUCTURE analyses were used to assess SNP data. MN wild and Czech RCG wild collection are genetically distinct. MN wild population is panmictic, but the Czech populations have distinct grouping of genotypes for each major Czech river. Additionally, MN herbarium samples cluster with extant samples, indicating the persistence of potentially native RCG genotypes over time. It appears that MN cultivars ('Palaton' and 'Venture') are different than most MN wild samples. Based on our current data we conclude that RCG MN wild populations are genetically distinct from those of Czech origin and native N. American genotypes are persistent in MN wild population. It is critical to identify the extent of native vs. exotic reed canarygrass populations in MN using genetic testing for better management of exotic invasive populations and preservation of native populations in state and Tribal lands.

W1**147: Weedy and Invasive Plant Genomics**

Gene Drive for Agricultural Weed Control: Opportunities, Challenges and Constraints

Paul Neve, Rothamsted, Harpenden, United Kingdom