W001: Abiotic Stress

High Day and Night Temperature Response in Cereals

Krishna Jagadish, Kansas State University-Department of Agronomy, Manhattan, KS

Heat stress induces significant damage to key physiological processes during flowering and grain filling reducing seed numbers and seed weight, respectively. Grains crops have adapted to harsh environments by exercising different mechanisms to minimize these damages. For example, during flowering crops employ heat stress escape (early morning flowering), avoidance (transpiration cooling) or true tolerance (reproductive success) mechanisms. Flowering pattern in rice has been successfully altered to induce early morning flowering to reduce heat stress damage during flowering. However, upland crops such as sorghum and wheat seem to house this trait inherently. Time-of-day of flowering as an effective route to adapt to harsh environmental conditions will be discussed. An additional dimension to the heat stress research is the need to address high night temperature (HNT) that is increasing at a much rapid pace than the day-time maximum temperature. Conclusions from chamber based studies indicate both high day- and night-time temperatures affect same set of physiological processes such as pollen fertility, spikelet sterility leading to yield reduction. In contrast, a series of studies using unique field based tents provided novel insights i.e., reduction in carbohydrates, increased night respiration, reduced biomass, lower 1000 kernel weight and grain width, highlighting the chain of events that eventual lead to reduced yield and poor-quality grain under HNT. Breeding targets to induce greater HNT tolerance will be discussed taking rice as a case study. Similar efforts to minimize HNT damage on US wheat and rice, has been initiated in close collaboration between UNL, A-State and KSU.

W002: Abiotic Stress

High Night Temperature Impacts the Transcriptome and Metabolome of Rice in a Sensitivity-Dependent Manner

Ellen Zuther, Ulrike Glaubitz, Sandra Schaedel, Alexander Erban, Joachim Kopka and Dirk K. Hincha, Max-Planck-Institute of Molecular Plant Physiology, Potsdam, Germany

The faster increase of night compared to day temperatures resulting in asymmetric warming and a broad decline in the diurnal temperature range was shown to have detrimental effects on the yield of rice (*Oryza sativa L.*). However, little is known about physiological, metabolic and transcriptional changes occurring during the early response of rice to high night temperature (HNT) conditions. HNT tolerant and sensitive rice cultivars were identified based on leaf chlorosis estimates in the vegetative stage and physiological differences between tolerance-classes were described. Analysis of primary metabolite profiles revealed a highly activated TCA cycle and amino acid biosynthesis pathways branching off from the TCA cycle especially in sensitive cultivars. These findings are supported by enzyme activity measurements. Furthermore a metabolic pre-adaptation in tolerant cultivars was postulated. From transcript profiling data on leaves using microarrays six genes were shown to be central for HNT responses, encoding proteins involved in transcription regulation, signal transduction, protein–protein interactions, jasmonate response and the biosynthesis of secondary metabolites. Integrated data analysis using clustering based on one-dimensional self-organizing maps identified a sensitivity and a tolerance profile highly correlated with HNT sensitivity ranks. The occurrence of metabolites within these profiles such as GABA and *myo*-inositol provides links to the TCA cycle in sensitive and to jasmonate signaling in tolerant cultivars.

W003: Abiotic Stress

Changes in the Pigeonpea Genome under Abiotic Stress Responses

Rachit K Saxena, Lekha T. Pazhamala, Pallavi Sinha, CV SameerKumar, KB Saxena and Rajeev K Varshney, ICRISAT, Hyderabad, India

The ability of pigeonpea crop to cope with drought and temperature depends on flexible mechanisms in genome which remodels expression of specific genes. To understand molecular mechanisms involved in response to above mentioned abiotic stresses, we have analyzed transcriptional and genomic changes in the pigeonpea genome. In the case of drought 51 genes were selected for expression profiling on three pigeonpea lines having different levels of drought tolerance and haplotype analysis across reference set (292 lines). Whereas, role of temperature in fertility transition of pollens studied in a temperature sensitive male-sterile line following RNA-seq approach. Detailed analysis through qRT-PCR and re-sequencing data (for drought) and RNA-seq (for temperature) revealed promising candidate genes for conferring abiotic stress tolerance in pigeonpea.

W004: Abiotic Stress

Enhanced Plant Resilience using a Rare Silicon Transport Allele in Soybean

Rupesh Deshmukh, Humira Sonah, Caroline Labbe, Julien Vivancos, Julie-Anne Wilkinson, Genevieve Arsenault, Aliyeh Rasooli zadeh, Rachelle Frenette-Cotton, Paul Isenring, Francois Belzile and Richard R. Bélanger, University Laval, Quebec City, QC, Canada

Silicon (Si) is recognized as a beneficial element for plant growth, particularly under stress conditions. While plant species are known to have different abilities to absorb Si, no efforts have been made towards the genetic improvement of a species to enhance Si uptake and derived benefits. In this study, in search of intraspecies variation for Si uptake, we identified a soybean line having unique properties to accumulate Si. Interestingly, a screening of 140 soybean genotypes revealed almost no variation for Si accumulation except for Hikmok sorip, a Korean genotype showing Si levels twice as high (2% vs 1% d.w.). Subsequently, using a recombinant inbred line population derived from a cross between Majesta and Hikmok sorip, we identified a quantitative trait locus (QTL), Hisil, explaining over 66% of the phenotypic variation. The whole genome re-sequencing information of parental lines identified three candidate genes in the QTL. Haplotype analysis identified unique variants in one of the genes and heterologous expression studies conducted in Arabidopsis and *Xenopus oocyte* confirmed the superior role of this rare allele in Si transport. In phenotypic studies, RILs carrying the H1 haplotype showed enhanced resistance against water stress and *Phytophthora sojae*. Our result highlight the potential evolution of Si accumulation trait in plants and the mechanism regulating the natural variation of *Hisil* in soybean.

W005: Abiotic Stress

Network Rewiring in Transgressive Segregants for Stress Tolerance in Rice: Epistasis, Complementation, Regulatory RNA, and DNA Methylation

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

In rice, non-parental phenotypes arise in a minority of individuals in recombinant populations because of transgressive segregation. We hypothesized that transgressive stress tolerance is the outcome of reconfigured biochemical networks with both genetic and epigenetic components. In certain recombinant populations, we observed that cold tolerance can be modified relative to parental levels by optimally complementing and epistatic transcription factors and their novel epialleles that alter regulon composition. Our current data on another population that is segregating for salt and dehydration tolerance indicate that the epigenomic landscapes can be drastically altered in certain recombinants possibly due to the confrontation of the diverse parental genomes and epigenomes during recombination, leading to modified patterns of gene expression. Genomic hypomethylation in a transgressive salt-tolerant recombinant has drastic effects on Miniature Inverted-repeat Transposable Elements (MITEs), which are known to contribute to rapid sequence evolution within the upstream regions of genes. We have identified the critical hubs of regulatory networks that integrate hormonal, developmental, and stress-related responses that acquired novel super-upregulation signatures as a consequence of DNA hypomethylation. We have also identified a subset of transgressively expressed miRNAs that appear to contribute to the fine-tuning of critical network hubs that integrate hormonal, developmental, and stress-related responses. Based on the trends we have uncovered so far, it appears that changes in epigenomic landscapes have important contributions to regulatory network rewiring through direct and indirect effects on the expression of some of the most critical hubs for transcriptomic reconfiguration through transcriptional and post-transcriptional mechanisms.

W006: Abiotic Stress

Corn Drought Tolerance Improvement through Manipulation of Hormone Pathways

Norbert Brugiere, DuPont Pioneer, Johnston, IA

Drought stress is one of the main environmental problems growers face every year around the world. Reduction in arable land area and reduced water availability make it paramount to identify and develop strategies to allow crops to be more resilient in water limiting environments. Although plant breeders have been successful in improving crops' drought-tolerance, both conventional and biotechnological approaches will be required to maintain and surpass current genetic gains. We have previously shown that manipulating the plant hormone ethylene can positively affect yield performance in drought stress environments. Abscisic acid (ABA) is another plant hormone playing an important role in the response of plants to drought stress and modulation of the ABA pathway therefore represents an attractive avenue to improve the drought tolerance of crops. We identified two novel RING-H2 genes called ZmXerico1 and ZmXerico2 and showed that their overexpression in Arabidopsis and maize confers ABA hypersensitivity and improved water use efficiency which can lead to enhanced maize yield performance in a controlled drought stress environment. Overexpression of ZmXerico1 and ZmXerico2 in maize results in increased ABA levels and decreased levels of ABA degradation products. We show that ZmXerico1 is localized in the endoplasmic reticulum, where ABA 8'-hydroxylases have been shown to be localized, and that it functions as an E3 ubiquitin ligase. We demonstrate that ZmXerico1 plays a role in the control of ABA homeostasis through regulation of ABA 8'-hydroxylase protein stability, representing a novel control point in the regulation of the ABA pathway and an opportunity for abiotic stress tolerance improvement.

W007: Abiotic Stress

Summary and Conclusions

Henry T. Nguyen, University of Missouri, Columbia, MO

Highlights of the abiotic stress workshop presentations will be provided.

W008: African Orphan Crops

The African Orphan Crops Consortium, an Uncommon Collaboration

Howard-Yana Shapiro, University of California, Davis, CA

W009: African Orphan Crops

The Genomes of African Orphan Crops

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

W010: African Orphan Crops

Breeding *Gynandropsis gynandra*, a Model for Orphan Crops

Deedi Sogbohossou, Wageningen University & Research, Wageningen, Netherlands, Enoch G. Achigan-Dako, Laboratory of Genetics, Horticulture and Seed Sciences, Faculty of Agronomic Science, University of Abomey-Calavi, Abomey-Calavi, Benin, Svein Solberg, The World Vegetable Center, Tainan, Taiwan, Patrick M. Maundu, Kenya Resource Center for Indigenous Knowledge, National Museums of Kenya, Nairobi, Kenya, M. Eric Schranz, Biosystematics, Wageningen University & Research, Wageningen, Netherlands and Allen Van Deynze, Seed Biotechnology Center, University of California, Davis, Davis, CA; Seed Biotechnology Center, University of California, Davis, CA

Gynandropsis gynandra is a highly nutritious leafy vegetable species used in Africa and Asia with a great potential to contribute to food security and improved livelihoods. The species belongs to the Cleomaceae, the sister family of the Brassicaceae and is being developed as a model for omics-assisted breeding of orphan leafy vegetables. Our team recently collected a total of 59 from Kenya and the first 164 accessions from West Africa which were added to the collections at the World Vegetable Centre. Ethnobotanical surveys with farmers and consumers of the species in Benin and Kenya revealed late flowering, reduced bitterness, ability to regrow after cutting and nutritional value as the desired attributes for improved varieties. Towards breeding for those traits, a core set of 80 representative accessions was characterized based on phenotypic data and metabolomic profiling. The results revealed geographic patterns of variation: West African and Asian accessions formed

two separate clusters but were closer to each other than with East African accessions. High levels of carotenoids (α-carotene, β-carotene, lutein, violaxanthin) and tocopherols (α-tocopherol and δ-tocopherol) were detected in the leaves. All these compounds varied significantly across the genotypes. Whole-genome re-sequencing of the collection is underway by the African Orphan Crops Consortium (AOCC) to elucidate the genetic diversity in the species. Full-sib populations are being developed for the genetic mapping of QTLs affecting flowering time, carotenoids and tocopherols content in the species.

W011: African Orphan Crops

Breeding Bambara Groundnut for Abiotic Stress: Use of a Rapid Screening Technique for Drought Tolerance

Busiso O. Mavankeni, Department of Research & Specialist Services, Causeway Harare, Zimbabwe and Patrick Olusanmi Adebola, Agricultural Research Council-Vegetable and Ornamental Plants Institute, Pretoria, South Africa

Bambara groundnut is an important crop in the small holder sector in Southern Africa. It is a good source of carbohydrates, proteins and fats. The production potential of the crop is being hampered by severe droughts which ravage most parts of Africa. The main aim of this study therefore was to screen accessions of Bambara groundnut for drought tolerance using a quick screening method, with a view to identify accessions that can perform well under water stress conditions. Twenty Bambara groundnut accessions consisting of 2 released cultivars and 18 landraces from the ARC-VOPI gene bank were planted for drought screening in the glass house over a period of over 6 weeks during which water was withheld for two weeks to induce stress. Wilting was scored 7, 11, 15 and 17days after stressing the plants. Records were also taken on number of dead plants and number of days to wilting. The entry Ginane UOV was the most drought tolerant line surviving the stressing period without showing any signs of wilting. There was also between and within line variability in response to drought. Use of the box method was efficient as a rapid screening technique.

W012: African Orphan Crops

Developing Resources for Genomics-Assisted Breeding in Finger Millet

Damaris Achieng Odeny, ICRISAT, Nairobi, Kenya and Finger millet whole genome sequencing consortium

Finger millet (*Eleusine coracana* subsp. *coracana*; 2n=4x=36) is the third most important cereal crop in the semi-arid regions of the world. The grain has an impressive nutritional profile, excellent storage qualities, and can grow under diverse agro-ecologies, including degraded farmland. Despite its importance as a low input crop, finger millet lacks the most basic genomic resources. Genome analyses and resource development is complicated by the polyploid genetics and lack of conclusive information on the donor species of one of the two progenitor genomes. To enhance genetic and genomic analysis in finger millet and facilitate crop improvement efforts, we generated the draft genome sequence of an improved finger millet genotype, KNE796. We obtained a finger millet genome assembly of 1.26 Gb (84% of estimated genome size) with a scaffold N50 of 2.29 Mb. Repeat elements accounted for 56.4% of the genome. The completeness of the assembly was verified in three independent assessments using transcriptome mapping, genotyping-by-sequencing (GBS) data and BUSCO analysis. We predicted 52,574 protein-coding genes validated with transcript reads from nine tissues. We successfully assigned pathway information to 67.14% of the predicted genes. We report a draft genome assembly of finger millet with predicted repeat elements and gene models. For a food security crop that has previously received little research attention, the resources generated here will go a long way in facilitating the breeding process of this nutritious and climate resilient crop. The data will enhance population studies, genomics-assisted breeding and our understanding of the evolution of the genus *Eleusine*.

W013: African Orphan Crops

Research in Tree Crops

Ramni Jamnadass, World Agroforestry Centre, Nairobi, Kenya

Trees are key for the provision of important products such as foods, fodder, fuel, medicine, timber and services including soil health and fertility, and carbon sequestration. With respect to food and nutrition, sub-Saharan Africa contains some of the areas of highest 'hidden hunger' and nutritional insecurity in the world, with acute nutritional deficiencies often exacerbated by consumption of homogenized diets lacking essential micronutrients and vitamins. These deficiencies lead to serious disease and health related developmental underachievement. African countries' governments have highlighted diversification in the production of nutritious foods as part of their strategy to improve diets supporting alignment with the UN Sustainable Development Goal 2 (hunger). Food trees can interface between the economic, environmental and social dimensions of food production but are a seriously neglected resource. Bringing food trees into farms for production, with appropriate market development, has tremendous potential to increase not only food and nutrition but also improve the agroecology of farming systems, reduce pressure on wild resources and raise agricultural incomes. To have a desirable impact, research and knowledge is required for safeguarding tree genetic diversity, improving and domesticating high value tree species and deciphering delivery pipelines for quality germplasm. Cost and time-effective domestication approaches, developing nutritious cultivars of a wide range of tree species, decision support tools to allow growers to match production requirements (e.g. diversity to support diet diversity) with suitable planting material, etc. are some of the research approaches being explored to help mainstream tree foods into food systems.

W014: African Orphan Crops

Capacity Building for Breeding African Orphan Crops

Rita H. Mumm[1], J. Bruce Walsh[2], Iago Hale[3], Allen Van Deynze[4], Todd C. Wehner[5], Rale Gjuric[4], Kent J. Bradford[4] and Howard-Yana Shapiro[4,6], (1)University of Illinois, Urbana, IL, (2)University of Arizona, Tucson, AZ, (3)University of New Hampshire, Durham, NH, (4)University of California, Davis, CA, (5)North Carolina State University, Raleigh, NC, (6)Mars, Incorporated, McLean, VA

With the African Orphan Crop Consortium (AOCC) sequencing 101 African orphan crops, the UC Davis African Plant Breeding Academy (AfPBA) was launched to empower the use of the DNA sequence information to develop improved crop cultivars for Africa. This professional education program aims to upgrade the knowledge base and skill sets of 150 MS- and PhD-level active plant breeders throughout the African continent, providing the background and tools to enable informed choice of parents, precise performance evaluation, and genomics-assisted

selection. Approaches and methods proven to increase genetic gains and to speed the development of improved cultivars are highlighted. The intensive 6-week course is delivered in three 2-week sessions at the World Agroforestry Centre in Nairobi, Kenya, over a span of 13 months. A world-class team of experts in the areas of genetics, experimental design and data analysis, and plant breeding present the concepts and principles, which are demonstrated in action by accomplished scientists invited to share the details of their breeding programs focused on specific product targets. The course has become a forum for collaboration among its talented and highly motivated participants. Outcomes of such collaborations include: establishment of the African Association of Plant Breeders, substantial grant awards to support graduate training in Africa of the next generation of plant breeders, and other continuing education programs targeting training of assistant breeders and technical staff. Funded by Mars Incorporated and the Alliance for a Green Revolution in Africa, to date the program has trained over 80 scientists from 25 African countries.

W015: Allele Mining

Inferring Genotypes from Skim Sequence using a Graph-Based Approach: The Practical Haplotype Graph

Peter Bradbury[1], Daniel Ilut[2], Lynn C Johnson[3], Zachary Miller[3], Ramu Punna[3], M. Cinta Romay[3] and Edward Buckler[4], (1)USDA ARS-Cornell University, Ithaca, NY, (2)Cornell University, Ithaca, NY, (3)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (4)USDA-ARS, Cornell University, Ithaca, NY

The discovery and use of allelic variation for genetics and, in particular, for breeding applications depends on low-cost genotyping. The rapidly declining cost of sequencing makes sequencing based approaches to genotyping particularly attractive. We describe a general, graph-based, computational framework that can be used with a variety of sequencing methods to infer high-density genotypes directly from low-coverage sequence. The framework combines existing software with custom code implemented in a series of Docker modules that allows users to build custom analysis pipelines that can run on a variety of system architectures. The first step of the method loads haplotypes from a population to a relational database. To genotype an individual, a graph is constructed from the haplotypes stored in the database. Sequence from the individual is then used with an HMM (hidden Markov model) to identify the most likely path through the graph. The resulting path is then translated to variant calls and output in VCF format.

W016: Allele Mining

Exploring Allelic Diversity Underlying Breeding Progress in European Wheat

Kai P Voss-Fels, The University of Queensland, St Lucia, Australia

Despite the remarkable successes that were achieved in the history of wheat breeding, future wheat production remains challenging. Climatic changes that lead to unprecedented extreme weather scenarios are accompanied by a rising disease pressure and a declining fertiliser availability. While the dramatic global population growth necessitates a significant further improvement of wheat productivity in the upcoming decades, a stagnation of wheat yield increases has recently been reported in all major production areas worldwide. This has mainly been attributed to a drastic loss of genetic diversity in elite breeding pools due to strong selective breeding and intensive germplasm exchange. At the same time there are public concerns that modern agriculture can only sustain productivity under extremely high resource inputs involving chemical fertilisers and plant protection, while the actual impact of genetic improvements remains elusive.

Here, we present the first large-scale investigation of the impact of wheat breeding on all major trait complexes in a historic panel of almost 200 registered European winter wheat varieties, including important representatives of the last five decades of winter wheat production. Presenting phenotype data from multiple locations and three different cropping systems that range from fully extensive to fully intensive, we are able to demonstrate the great impact of genetic improvement on performance increase under any environmental scenario. Linking this to genome-wide marker information we are able to track the influence of artificial selection on genetic parameters throughout the history of wheat breeding and to define target regions with the highest impacts on agronomically important traits. Our study gives first insights into the genetic basis of the improvement of high-yielding winter wheat and assesses the potential for further genetic gain in the European elite germplasm pool in the short- and mid-term.

W017: Allele Mining

Identifying Useful Alleles for Crop Improvement: Applying State of the Art Genomic Tools, Methods and Approaches to Characterise the WHEALBI Barley Genetic Resource

Daniela Bustos-Korts[1], **Alessandro Tondelli**[2], Joanne Russell[3], Ian Dawson[3], Noemi Trabanco[4], Davide Guerra[2], Stefano Delbono[2], Stylianos Kyriakidis[3], Allan Booth[3], Chiara Ferrandi[4], Francesco Strozzi[4], Ezequiel L. Nicolazzi[4], Hakan Ozkan[5], Marta Molnar-Lang[6], Mária Megyeri[6], Mikó Péter[6], Benjamin Kilian[7], Nils Stein[8], Laura Rossini[4], Robbie Waugh[3], Luigi Cattivelli[2] and Fred A. van Eeuwijk[1], (1)Wageningen University & Research - Biometris, Wageningen, Netherlands, (2)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (3)The James Hutton Institute, Invergowrie, Dundee, United Kingdom, (4)Parco Tecnologico Padano, Lodi, Italy, (5)University of Cukurova, Adana, Turkey, (6)Agricultural Institute, MTA ATK, Martonvásár, Hungary, (7)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (8)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany

The EU-funded WHEALBI project (Wheat and barley Legacy for Breeding Improvement; http://www.whealbi.eu) is taking a multidisciplinary approach to identify, understand and utilise the genetic diversity available in wheat and barley cultivars, landraces and wild relatives. This is strategic to meet the challenge of increasing crop yields while reducing environmental impact. Here we focus on a carefully selected set of 400 barley accessions from extensive *ex situ* collections which cover the geographical and agro-ecological adaptive range of barley. Agronomic and life history traits, collected from multi-environment common gardens experiments across Europe, provide a unique dataset to decipher the genetic basis of adaptation to environmental conditions. A comprehensive molecular variant analysis by exome sequencing identified 1.75 million SNPs that have been used to investigate allelic variation at candidate genes driving phenotypic differences for heading date, plant height, grain weight, and awn length. We were able to relate geographical origins using site information to genotypic diversity, providing valuable information to identify novel ‘adapted’ alleles for future breeding under a changing climate.

W018: Allele Mining

GWA Mapping of *Oryza sativa* Median Lethal Low Temperature (LT50) QTL using the USDA Rice Diversity Panel 1 (RDP1)

Naoki S. Shimoyama, Marquette University, Milwaukee, WI

Improving crop productivity is vital to meet the needs of a growing world population. Rice is a staple crop feeding approximately half of the world population. Cold stress is a major factor limiting rice productivity due to its tropical origin. We measured seedling survivability at different temperatures for different rice cultivars with varying degrees of cold sensitivities to determine the Median Lethal Low Temperature (LT50) and used it for quantitative trait locus (QTL) mapping to identify potential mechanisms involved in cold stress tolerance. We calculated the LT50 of more than 300 cultivars from the USDA Rice Diversity Panel 1 (RDP1), because its cultivars were genotyped using the High Density Rice Array (HDRA), and 700,000 SNPs are currently available. The RDP1 and HDRA were used in an effort to produce data with great breadth in diversity and resolution, respectively. Using the open source genome-wide association (GWA) mapping pipeline based on the HDRA data, we were able to identify numerous QTL for the LT50 trait. The LT50 phenotypic data will also be used to bin the cultivars with similar median temperatures. By conducting GWA mapping and QTL analysis between these bins, we expect to identify mechanistically significant genes associated with certain LT50 bins. With this analysis, we expect to show that different LT50 bins may employ unique cold tolerance mechanisms and pathways. Genetic pathways linked to membrane composition, lipid synthesis, mechanosensing, chemosensing, and metabolism will be discussed.

W019: Allele Mining

Benchmarking Performance of GWAS Tools on the 3,000 Rice Genomes Dataset

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

The 3,000 rice genomes dataset (3kRG) has established itself as a resource for finding novel allelic variation in rice. Researchers who wish to use GWAS approach to search for causative variation in 3k lines face a problem of selecting best tool that would work well for the given phenotype and the volume of genotype data. Here, we benchmark accuracy and speed of several popular GWAS algorithms on the simulated quantitative phenotypes for the 3kRG data. The top performing tools will be made available for researchers working on 3k as part of a user-friendly 3k GWAS pipeline.

W020: Allele Mining

GWAS Reveals Genetic Diversity Associated with Variation in Plant Architecture and Leaf Photosynthesis

Sónia Negrão, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Increasing the yield potential of rice is a vital requirement to feed the world's growing population. One way this may be achieved is through improvements to photosynthetic efficiency. Very little is known about the factors underlying natural diversity in photosynthesis and how this is related to plant architecture.

Here we present a genome-wide association study (GWAS) with the aim to discover new genes associated with traits contributing to natural diversity in leaf photosynthesis under both controlled and field conditions. Our study includes high-throughput phenotyping and field trials. We used the *indica* diversity panel, which was established within the 'Phenomics of Rice Adaptation and Yield Potential' (PRAY) project. First, the *indica* panel was grown in controlled waterlogged conditions, and used to quantify several plant architectural traits such as plant compactness and volume using high-throughput phenotyping. Second, this panel was cultivated under irrigated field conditions in two different locations in the Philippines, to investigate leaf anatomy and several photosynthetic parameters. The use of the same *indica* panel in these two experiments may contribute to the improvement yield potential by discovering novel loci associated with plant architecture, photosynthesis, and leaf anatomy. We found an overlap of the best performing accessions between transpiration use efficiency under controlled conditions and stomatal conductance to water vapor along with intrinsic transpiration efficiency under field conditions. Moreover, we found that plant compactness and leafiness influences various photosynthetic traits.

Our latest data will be presented, and the linking results from GWAS under controlled and field conditions will be discussed. The final goal is to explore the genetic diversity present in the PRAY *indica* panel, and to develop higher yielding rice varieties.

W021: Analysis of Complex Genomes

Snapshots of Genome Evolution in Allopolyploid Grasses

Sean Gordon, DOE Joint Genome Institute, Walnut Creek, CA

Polyploid genomes are characteristic of grasses being developed as biomass crops and many grain crops. Therefore, a deeper understanding of gene regulation and genome evolution in polyploid genomes would be useful for developing improved crop varieties for both food and fuel. Despite their economic importance and being fascinating examples of evolutionary processes in motion, it is notoriously difficult to obtain high-quality whole genome assemblies for polyploids. We are developing both computational tools and experimental systems to study allopolyploids. I will first discuss our work developing *B. hybridum* and its extant progenitor-like species (*B. distachyon* and *B. stacei*) as a simple model system to study allotetraploid genome regulation and evolution. All three species have very compact genomes, small stature and are easily grown and manipulated in the laboratory. Chromosome-level assemblies have been developed for all three species, and we have performed matched RNA-Seq and methylation experiments to reveal sequence evolution and modifications in gene regulation within the allopolyploid relative to its progenitors. We are also exploring synthetic *B. hybridum* lines obtained from crosses between the progenitor species, *B. distachyon* and *B. stacei*. In the second part of the talk, I will discuss our progress in developing computational tools to dissect complex allopolyploids *in silico*, applied to a broad range of systems, and biological insights that this enables.

W022: Analysis of Complex Genomes

Accurate Prediction of Fiber Length using its Contributing Genes for Gene-Based Breeding in Cotton

Yun-Hua Liu, Department of Soil and Crop Sciences, Texas A&M University, College Station, TX

Phenotype prediction of quantitative traits is paramount to enhanced plant breeding. We previously cloned 474 *GFL* (*Gossypium* fiber length) genes significantly contributing to cotton fiber length using a novel and genome-wide high-throughput gene cloning technology (*gExpress*). Here we report the accurate prediction of fiber length using these *GFL* genes, especially using their number of favorable alleles, genotypes and expression profiles. We predicted the phenotype of fiber length directly using their number of favorable alleles with taking into account additive and dominant effects, or using their genotypes and expressions with nine prediction models widely used in genomic selection. When one of these three datasets was independently used for fiber length prediction, a prediction accuracy of $r = 0.78 – 0.85$ was obtained, approaching the maximal prediction accuracy. This has substantially improved the prediction accuracies of quantitative traits thus far achieved by genomic selection using genome-wide DNA markers, genome-wide transcriptomes or genome-wide metabolomes. When two or all of the three *GFL* datasets were jointly used, the prediction of fiber length was 100% ensured. Therefore, the *GFL* genes are capable of accurately predicting the phenotype of fiber length. Furthermore, we developed a gene-based breeding (GBB) system that selects parents, design crosses, performs progeny selection, and develop cultivars using 125 of the *GFL* genes critical to fiber length and conducted GBB for three years. The results have further confirmed the ability, utility and efficiency of the *GFL* genes for accurate prediction of fiber length, thus for enhanced and accelerated breeding in cotton.

W023: Analysis of Complex Genomes

Efficient Hybrid Breeding Based on the Modification of Meiosis using Virus-Induced-Gene-Silencing

Vanesa Calvo, WUR Wageningen University, Wageningen, Netherlands, Hans de Jong, Wageningen University & Research, Wageningen, Netherlands, C. Bastiaan de Snoo, Rijk Zwaan Breeding BV, Fijnaart, Netherlands, Arp Schnittger, University of Hamburg, Hamburg, Germany and Erik Wijnker, Wageningen University, Wageningen, Netherlands

Hybrid production in traditional breeding programmes is usually a lengthy process. Obtaining new inbred lines or finding favourable parental combinations requires a great investment of resources. To overcome these issues, we have developed a new breeding approach to produce hybrids more efficiently. First, it allows to select and fix any unknown hybrid genotype. Secondly, we can also obtain near-full hybrids that can perform either as the initial hybrid or potentially better. Our new breeding technique is based on the reduction of 80% of meiotic crossovers, by silencing MSH5 directly in the hybrid, using Virus-Induced-Gene-Silencing (VIGS). The major advantages of this new application are: i) absence of stable transgenes to modify gene expression ii) rapid generation of hybrids (only 3 generations from the initial hybrid) and iii) the production of new parental lines that are either chromosome substitution lines or low-recombinant lines (1 or 2 recombination events). This efficient hybrid breeding approach based on the modification of meiosis brings a new insight to hybrid production and evaluates hybrid performance by comparing full hybrids with rapidly obtained near-full hybrids.

W024: Analysis of Complex Genomes

Deciphering the Genome of the Insect Vector *Diaphorina citri* to Develop Solutions for the Citrus Greening Disease

Prashant S Hosmani[1], Mirella Flores-Gonzalez[1], Sherry Miller[2], Teresa Shippy[2], Wayne Hunter[3], Tom D'elia[4], Michelle Cilia[5], Susan J. Brown[6], Lukas Mueller[1] and **Surya Saha**[1], (1)Boyce Thompson Institute, Ithaca, NY, (2)Kansas State University, Manhattan, KS, (3)USDA Agricultural Research Service, Fort Pierce, FL, (4)Indian River State College, Ft. Pierce, FL, (5)Cornell University, Ithaca, NY, (6)Division of Biology - Kansas State University, Manhattan, KS

The Asian citrus psyllid (*Diaphorina citri* Kuwayama) is the insect vector of the bacterium *Candidatus* Liberibacter asiaticus (CLas), the causal agent for the citrus greening or Huanglongbing disease which threatens citrus industry worldwide. This vector is the primary target of approaches to stop the transmission of the pathogen. Accurate structural and functional annotation of the psyllid's gene models and understanding its interactions with the pathogenic bacterium, CLas, is required for precise targeting using molecular methods such as RNAi. We opted for manual curation of gene families in the draft genome of *D. citri* (Diaci v1.1, contig N50 34.4Kb) that have key functional roles in *D. citri* biology and pathology. The community effort resulted in Official Gene Set v1.0 with more than 500 manually curated gene models across developmental, RNAi regulatory, and immune-related pathways. More information about the annotation process is available here. Single copy marker analysis of the current genome shows a significant proportion of 3,350 markers conserved in Hemipterans to be missing (25%) with only 74% present in full-length copies. The manual genome annotation also identified a number of misassemblies and missing genes in the current genome. This is, in-part, due to the complexity introduced when assembling a heterogeneous sample containing DNA from multiple psyllids and is further exacerbated by the use of short reads. This challenge is common with insect genomes due to the size of individuals. To improve quality of genome assembly, we generated 36.2Gb of Pacbio long reads with a coverage of 80X for the 450Mb psyllid genome. The Canu assembler followed by Dovetail Chicago-based scaffolding was used to create an improved assembly (Diaci v2.0) with a contig N50 of 758.7kb and 1906 contigs. The assembly was polished with Pacbio and Illumina paired-end reads to remove indel and SNP errors followed by manual curation. We are employing Dovetail Chicago and 10X Illumina libraries generated from a single psyllid in conjunction with Bionano optical maps to achieve long-range scaffolding of the genome. We have also generated full-length cDNA transcripts from diseased and healthy tissue from multiple life stages with the Pacbio IsoSeq technology. This will be the first time all these methods have been applied to resolve a complex insect genome from a highly heterogeneous sample. The new assembly will be available on citrusgreening.org which is our portal for all omics resources for the citrusgreening disease. We are continuing with the manual curation effort using the improved genome. We will also present how the improved genome and annotation is contributing to the development of molecular interdiction methods to disrupt the transmission ability of *D. citri*.

W025: Analysis of Complex Genomes

Building Maize High Quality, Chromosome-Scale, *de novo* Genome Assemblies by Scaffolding Next-Generation Sequencing Assemblies with BioNano Maps Generated with the New Direct Labeling Enzyme

Yang Zhang, Bionano Genomics, San Diego, CA

W026: Analysis of Complex Genomes

Comparative Evaluation of Gene Annotation Methods in Conifer Megagenomes

Sumaira Zaman[1], Madison Caballero[2] and Jill L. Wegrzyn[1], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Connecticut, Storrs, CT
The emergence of improved sequencing technologies has brought us closer to understanding plant biodiversity. However, sequencing plants brings its own set of concerns, including: heterozygosity, polyploidy, and repetitive content. Specifically sequencing projects involving conifers are inherently challenging due to years of evolution which has introduced repetitive content and numerous pseudogenes. Hence, genome sizes in sequenced conifers can range from 16 to 32 Gbp. Genome annotation of conifers is equally challenging as they are non-model species with limited genetic resources. Furthermore, their incomplete gene space is convoluted with pseudogenes, complicating the identification of protein coding genes.
Genome annotation is further complicated by intermediary steps that can introduce bias and serve as sources of error downstream. Here, we examine the impact of using frame selected de novo assembled transcriptomes versus RNA reads used to train ab initio gene predictors. Both data types introduce error and we study the effects of these errors through specific criteria including gene length, intron length, protein annotation, and completion of orthogroups. Conifers have been known to have introns as long as 700 Kbp. Current annotation packages often cannot properly call these genes leading to fragmented or incomplete gene annotations. This study examines not only the impact of processing external evidence which is often sourced from a different genotype than used for the genome assembly, but also the primary applications used and their impact on the final gene annotation. These large and complex genomes serve as an ideal model to assess speed and accuracy of the available applications.

W027: Animal Epigenetics
Epigenetic Effects of Cowpea Polyphenols in Dairy Cattle
Mulumebet Worku and Sarah Adjei-Fremah, North Carolina Agricultural and Technical State University, Greensboro, NC
Polyphenols compounds including catechin, myricetin, quercetin, have antioxidant and anti-inflammatory properties, and have ability to prevent oxidative stress damage. Recent studies have indicated strong epigenetic effects of polyphenols especially in modulating epigenetic-related enzymes in the cell through either activation or inhibition. Dietary polyphenols exert their beneficial effects via chromatin remodeling through modulation of histone deacetylase (HDAC) and DNA methyl transferases (DNMT) activities. Flavonoids especially under in vitro conditions have been shown to be potent inhibitors of DNMTs. Furthermore, the anti-inflammatory properties of polyphenols have been associated with their ability to induce HDAC activity. Cowpea a highly nutritious leguminous plant is a rich source of polyphenols that is used as animal feed. Polyphenols within cowpea (CPE) have antioxidant, anti-inflammatory, properties, but the epigenetic effects of CPE in dairy cows are yet to be elucidated. The objective of this study was to evaluate the effects of cowpea polyphenols in modulating epigenetic-related enzymes in cow blood. Meta-analysis of previous microarray experimental data that compared cowpea polyphenol treatment and control in whole blood of dairy cows was conducted. Our results identified the expression and modulation of four epigenetic-related enzyme genes HDAC1, H3F3B, DNMT3A and methionine synthase reductase (MTRR) by CPE . There was an increase in mRNA expression of HDAC1 (Fold change, FC = 6.60), MTRR (FC=10.09) and H3F3B (FC=146.32), and downregulation of DNMT3A (FC= -5.86) after CPE exposure. Results from the current study suggest potential of cowpea polyphenols as a modulator of epigenetic targets HDAC1 and DNMT. Therefore, cowpea polyphenols can be further explored as a nutri-epigenetic intervention and a therapeutic strategy with impact on inflammation-related morbidities in dairy cows.

W028: Animal Epigenetics
Evaluating the Impact of Maternal Methionine Supplementation on Fetal Developmental Programming using Multi-Omics Data
Rocio Amorin, University of Florida, Gainesville, FL

W029: Animal Epigenetics
Non-Mendelian Genetic and Epigenetic Effects in Bovine Fetal Development
Ruijie Liu[1], Rick Tearle[1], John Williams[1], Cynthia Bottema[1], Dana Thomsen[1], Tong Chen[1], Timothy P.L. Smith[2] and Stefan Hiendleder[1,3], (1)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia, (2)USDA, ARS, USMARC, Clay Center, NE, (3)Robinson Research Institute, University of Adelaide, North Adelaide, Australia
Parent of origin-dependent genetic and epigenetic factors and their interactions are important determinants of prenatal development that remain largely unexplored. These are of considerable interest in animals and human as prenatal development impacts postnatal phenotype. We are using an intra-species large animal model based on purebred and reciprocal cross *Bos taurus taurus* (Bt) and *Bos taurus indicus* (Bi) that allows us to dissect X and Y chromosome and parent of origin effects, including allelic imbalances and imprinting. The phenotypic differences of purebred and hybrid genome combinations, including weights and growth rates, are significant with pronounced sex effects.
Transcriptome and miRNA data has been obtained from five fetal tissues: liver, brain, muscle, placenta and lung. For each tissue, three male and three female samples for each of four genetic groups, two pure bred and two reciprocal crosses, were examined. Transcriptome data from liver, muscle and lung showed clear separation of males and females and genetic groups. Separation according to genetics was less obvious in placenta and sex effects less obvious in brain samples. We identified nine differentially expressed (DE) genes across 5 tissues which are likely to be true sex-specific genes. Several of the 9 DE sex effect genes are located on both sex chromosomes and one on an autosome showed allele specific expression. We found 5 of 9 common DE genes are annotated as Y-linked genes on chromosome X in the current bovine reference genome (UMD3.1.1). Our analyses based on SNPs suggest Y-linked genes are more likely to be in the male-specific region. Our results suggest that the one terminal section of the X chromosome in UMD 3.1.1 is incorrect. Further studies will focus on assemblies of significantly improved X and Y chromosome sequences and integrate the transcription patterns of mRNA and small RNAs in tissues.

W030: Animal Epigenetics

Epigenomic Analysis of DNA Methylomes in Horse, Donkey and their Reciprocal Interspecific Hybrids Mule and Hinny from Whole-Genome Bisulfite Sequencing (WGBS-seq) Data
Xu Wang, Auburn University, Auburn, AL

W031: Animal Epigenetics
Evaluating the Role of Epigenomic Modifications in Host–Pathogen Interaction for Bovine Alveolar Macrophages Infected with *Mycobacterium bovis*
David E. MacHugh, University College Dublin, Dublin, Ireland

W032: Animal Epigenetics
Transcriptional and Genome-Wide Methylation Profiling in Fetal Pig Skeletal Muscle
Ryan J Corbett[1], Nancy E. Raney[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
Development, growth, and function of skeletal muscle are dynamic processes. Formation of skeletal muscle fibers occurs during pig fetal development in two waves: primary fibers form de novo (30-60dg) and secondary fibers form around primary fibers (54-90dg). Changing biological demands of tissues during development require coordinated transcriptional regulation; DNA methylation is one mechanism impacting this process. We used *Longissimus dorsi* muscle samples from pigs at 41dg and 70dg (n=3 per stage) to determine changes in CpG methylation status (whole-genome bisulfite sequencing, WGBS), and transcript abundance (RNA-seq and miRNA-seq) between stages. Sequencing was performed on the Illumina HiSeq 4000 platform. Bismark was used to align WGBS reads and obtain methylation rates at each CpG, and differential methylation (DM) analysis was performed using the methylKit R package. RNA-seq and miRNA-seq reads were mapped to the *S. scrofa* reference genome (v.11.1) using TopHat2 and miRDeep2, respectively; differential expression (DE) analysis was performed using DESeq2. Significant negative and positive correlations were observed between gene expression and methylation in promoters ($r = -0.25$, $p<0.001$) and gene bodies ($r = 0.28$, $p<0.001$), respectively. We identified 17,710 DM regions (3,518 hypermethylated and 14,192 hypomethylated at 70dg versus 41dg), which were enriched among gene promoters. DE genes (895 upregulated and 718 downregulated, FDR<0.05) were enriched among 3,911 DM genes. Our results agree with previously reported relationships between gene methylation and expression in other species, and reveal genes with synchronized changes in methylation and expression during pig skeletal muscle development that warrant further study.

W033: Animal Genomics and Adaptation to Climate Change
Investigating Host-Genetic Mechanisms Associated with Fescue- and Heat-Stress Tolerance in Cattle
Nick Serao, Department of Animal Science, Iowa State University, Ames, IA

W034: Animal Genomics and Adaptation to Climate Change
Genetics Response of Small Ruminants to Heat Stress
Joram M. Mwacharo, International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia
Predictions based on several scenarios suggest more frequent hot and fewer cold temperature extremes, on daily and seasonal timescales, as global mean temperatures rise. This will occasion more frequent droughts and increased water scarcity, which will substantially exacerbate food/feed insecurity and instability. The need for livestock that can support agricultural industries to meet projected increasing demands for animal source foods, and simultaneously cope with stresses arising from increasing global temperatures is thus a priority. Small ruminants (sheep and goats) are particularly important in this regard due to their resilience to adapt to a wide range of climes. Here, through the analysis of SNP genotype data, we present, at the genome-wide level, footprints of adaptation to biotic and abiotic stresses in desert-dwelling/adapted populations of indigenous sheep and goats. Our results suggest that the adaptation mechanisms involve a large network of interacting genes (found across several candidate genomic regions) and biological pathways. In particular, we reveal selection sweeps around candidate regions spanning genes associated with muscle function, energy metabolism, endocrine and nervous system function, thermo-tolerance and autoimmune and inflammatory response. The findings of the study offer a promising step towards mining the genetic potential of adaptable indigenous livestock as the foundation to breed appropriate small ruminants which can provide a viable option to mitigate against food insecurity and instability in increasingly volatile climatic events.

W035: Animal Genomics and Adaptation to Climate Change
Genomic Analysis of Responses to Heat Stress in Three Chicken Genetic Lines
Ying Wang[1], Perot Saelao[1], Colin Kern[1], Kelly Chanthavixay[1], Sihua Jin[1], Rodrigo A. Gallardo[2], David A. Bunn[1], Terra R. Kelly[2], Jack M. Dekkers[3], Susan J. Lamont[3] and Huaijun Zhou[1], (1)Animal Science, University of California, Davis, CA, (2)University of California, Davis, School of Veterinary Medicine, Population Health and Reproduction, Davis, CA, (3)Department of Animal Science, Iowa State University, Ames, IA
Heat stress results in significant economic losses to poultry industry. The objective of this study was to identify genes, signal pathways, and genomic regions associated with heat stress in liver, breast muscle, and hypothalamus through an integration of RNA-Seq analysis on two distinct, highly inbred chicken lines (Leghorn and Fayoumi) and a genome wide association study (GWAS) using commercial egg-layer Hy-Line Brown chicks. At 14 days(d) of age, birds were exposed to 38°C with 50% humidity for 4 hours, then 35°C until the conclusion of the study, while non-treated individuals were maintained at 29.4°C. The heat-treated birds were inoculated at 21d with Newcastle disease virus (NDV) La Sota strain to investigate the effects of heat stress and NDV infection. Blood components were measured using the iSTAT at 4 hours, 10 days and 14 days post treatment (hpt and dpt). Tissues were harvested at both 4hpt and 10dpt and used to identify differentially expressed genes (DEG) with an FDR < 0.05 and fold change > 1.5. Significant DEGs and pathways were identified, which are associated with heat stress response at 4hpt and with combined heat stress and NDV infection at 10dpt. Birds were genotyped using a 600K SNP chip and

GWAS was performed using the GenABEL R package on the phenotypes. GWAS revealed 11 unique SNPs associated with blood parameters at all three time points. Further analysis will explore the association of genetic variants with thermal tolerance, which can be used for potential genetic improvement of heat stress in poultry.

W036: Animal Genomics and Adaptation to Climate Change
Genomic Analysis of African Local Chickens to Elucidate Tropical Adaptation
Almas Gheyas[1], Adriana V. Trujillo[2], Adebabay Kebede[3], Ahmed Al-Jumaili[2], Dessie Tadelle[4], Nick Sparks[1], Jacqueline Smith[1] and Olivier Hanotte[2], (1)The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom, (2)The University of Nottingham, Nottingham, United Kingdom, (3)Addis Ababa University, Addis Ababa, Ethiopia, (4)International Livestock Research Institute, Addis Ababa, Ethiopia
Local breeds of chicken used in backyard farming in Africa are adapted to their tropical environmental contexts and have high tolerance to both biotic and abiotic stressors like incidence of infectious diseases, exposure to high tropical heat, and access to marginal feed. However, these local chickens generally have much lower productivity compared to elite breeds used for commercial farming. Genetic characterisation of local breeds of African poultry can provide insights into their adaptations to tropical environment and opportunities for developing improved chicken lines by incorporating desirable genetic factors by either marker-assisted selection, introgression or genome editing.
In a large collaborative project between the Roslin Institute and ILRI under the Centre for Tropical Livestock Genetics and Health (CTLGH), we are undertaking genome characterisation of many local chicken populations from Ethiopia, Nigeria and Tanzania (total 385 birds from 39 ecotypes). In the first phase, sequence data from 12 populations of Ethiopian chickens from different geographical regions is being analysed. We are studying genetic relationship among these populations, their adaptive diversity, and genomic imprints of positive selection in response to local environmental contexts. Preliminary results indicate presence of many candidate regions of selection signatures across the genome showing reduced heterozygosity. The project is expected to generate knowledge on genetic regulation of different adaptive and phenotypic traits. The candidate genes and variants identified would be used for breed improvement operations under African Chicken Genetic Gains (ACGG) program through genome editing technology.

W037: Animal Genomics and Adaptation to Climate Change
Detection of Selection Signatures Among Brazilian, Sri Lankan, and Egyptian Chicken Populations under different Environmental Conditions
Muhammed Walugembe, Department of Animal Science, Iowa State University, Ames, IA

W038: Animal Genomics and Adaptation to Climate Change
Studies in Genetics of Heat Stress in Dairy, Beef and Pigs
Ignacy Misztal, University of Georgia, Athens, GA
Production environments are expected to change, mostly to hotter climates but also possibly more extreme and drier. This raises a question whether farm animals should be specifically selected for the changing conditions. Specific studies in various species show complexities of defining and selecting for heat tolerance. In dairy, the genetic component of heat stress on production is relatively small in the first parity but increases strongly to second and again to third parity. In hot but less intensive environments the effects of heat stress on production is minimal although the effect on fertility remains. Mortality shows a peak under heat stress, again increasing in parity, but extensive data editing is required for any research in mortality, and the definition of mortality changes due to new regulations. In Angus, the effect of heat stress can be strong only in a few regions, partly due to adaptation of calving seasons to local conditions. The maternal effect seem to have a very low heat stress component perhaps due to dams shielding calves from environmental challenges. In pigs, the effect of heat stress is strong in commercial but almost none in nucleus operations. This is due to lower pig density and better heat abatement in nucleus farms. Genomic evaluation for heat stress is currently done for dairy in Australia and for pigs in the U.S. Few regions were associated with heat tolerance, e.g., “slick hair” gene in dairy. The need for genetic improvement in heat stress can be partially offset by improved management.

W039: Aquaculture
Assembly and Computational Use of Aquatic Genome Models
Wesley Warren, McDonnell Genome Institute at Washington University, St. Louis, MO

W040: Aquaculture
The National Center for Biotechnology (NCBI) Genome Annotation Resources for Aquaculture Species
Nuala A. O'Leary, National Center for Biotechnology Information (NCBI), Bethesda, MD
NCBI’s Eukaryotic genome annotation pipeline (ncbi.nlm.nih.gov/genome/annotation_euk/) incorporates genomic, transcript, and protein sequence records, including RNA-seq data available in SRA, to provide comprehensive annotations of public genome assemblies submitted to NCBI's Assembly resource (ncbi.nlm.nih.gov/assembly). To date, this pipeline has been used to annotate more than 420 eukaryotic genomes across diverse array of taxa. Among these annotated genomes are several economically important aquaculture species including catfish, salmon, tilapia, scallop and oyster. The annotations provided by this pipeline are available in various NCBI resources, including Reference Sequence (RefSeq) sequence databases, Gene, BLAST databases, FTP and in NCBI’s Genome Data Viewer. All genome annotations produced by this pipeline are in scope for manual curation by the RefSeq curation group. Curators correct sequence or feature annotation errors that are identified by quality assurance tests, generate additional splice variants, and add feature annotation and data attributes. These curated RefSeq transcript, proteins and genomic regions, designated by NM_NP_, NR_, or NG_ accession prefixes, serve as reagents to NCBI’s Eukaryotic genome annotation pipeline, thereby contributing iteratively to improvements in our genome annotation products. This presentation will describe some of the computational and manual curation procedures used in NCBI's genome annotation process and provide guidance on how the resources can be accessed and utilized by the aquaculture research community.

W041: Aquaculture
The Eastern Oyster Genome: A Resource for Comparative Genomics in Shellfish Aquaculture Species
Marta Gomez-Chiarri, University of Rhode Island, Kingston, RI

W042: Aquaculture
The Sea Cucumber Genome Provides Insights into Morphological Evolution and Visceral Regeneration
Lina Sun, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

W043: Aquaculture
The Genome of the Coho Salmon
Eric B. Rondeau[1], Jong S. Leong[1], David R. Minkley[1], Kris A. Christensen[2], Cody A. Despins[1], Anita Mueller[1], Robert H. Devlin[2], Ruth E. Withler[3], Terry Beacham[4], Kerry A. Naish[5], Jose Manuel Yanez[6], Roberto Neira[7], Louis Bernatchez[8], William S. Davidson[9] and Ben F. Koop[1], (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, Pacific Biological Station, Nanaimo, BC, Canada, (4)Molecular Genetics Lab, Pacific Biological Station, Nanaimo, BC, Canada, (5)University of Washington, Seattle, WA, (6)Aquainnovo, Puerto Montt, Chile, (7)University of Chile, Santiago, Chile, (8)Université Laval, Québec, QC, Canada, (9)Simon Fraser University, Molecular Biology and Biochemistry Department, Burnaby, BC, Canada
Coho (*Oncorhynchus kisutch*) is a culturally and economically important salmon species that spawn in rivers that flow to the Northern Pacific Ocean. Wild coho salmon stocks have seen significant declines over the past quarter century, with few signs of recovery. An emerging species in aquaculture in its native range, the species is an established component of the aquaculture industry in Chile. To assist in wild stock management and to further develop hatchery and aquaculture stocks, genomic tools for coho salmon are being developed. Such efforts are complicated by the salmonid whole-genome duplication event approximately 80-100 MYA which resulted in a large, highly similar, pseudo-tetraploid genome. In this work, we present a chromosome-level assembly of the Coho salmon genome. This first release of the assembly has a contig N50 of 58 kbp, a scaffold N50 of 1,266 kbp and a longest scaffold of 15,030,138 bp. The assembly was anchored to a previously published linkage map, with 71.1% of assembled bases assigned to a chromosome. A transcriptome was generated from fifteen tissue-specific libraries for genome annotation and to generate a gene expression atlas. Sixty coho sourced from twelve locations throughout the North American range have been re-sequenced to evaluate variation and to develop a high-density SNP array for consistent genotyping in further populations. The Coho salmon genome and the development of genomic resources will aid in management and conservation of wild populations, the analysis of hatchery efficacy and improvements in aquaculture production.

W044: Aquaculture
Functional Genomic Basis of Host Resistance to Sea Lice in Atlantic Salmon
Ross D. Houston, The Roslin Institute and R(D)SVS, Edinburgh, United Kingdom
Sea lice are parasitic copepods that cause large economic losses to salmon aquaculture worldwide. Alternative methods to control this parasite, such as selection for host resistance, are increasingly important. Insight into the host-parasite interaction and mechanisms of host resistance can lead to improvements in selective breeding for resistance, and potentially novel treatment targets. The aims of this study were to characterise the functional genomic basis of host resistance to lice, and to identify potential functional polymorphisms underlying resistance. To achieve this, challenge experiments were performed on a population of salmon from a Chilean breeding program, from which a large genome-wide association study was performed using a SNP array. In addition, salmon from resistant and susceptible families were compared using RNASeq of attachment sites and healthy skin, and using whole genome sequencing. Analyses of the gene expression signature of host resistance revealed genes and pathways associated with resistance, and these results were cross referenced with the GWAS and WGS data to identify candidate functional resistance genes and polymorphisms. These results improve our understanding of host response to lice in salmon, and highlight potential functional genomic variants that could be used to enhance genomic selection for host resistance to lice in salmon breeding programs to help tackle this major disease problem.

W045: Aquaculture
GWAS Analysis of Disease Resistance Against Enteric Septicemia of Catfish (ESC)
Huitong Shi[1], Tao Zhou[1], Suxu Tan[1], Xiaozhu Wang[1], Yujia Yang[1], Shikai Liu[2], Lisui Bao[1], Ning Li[1], Zihao Yuan[1], Yulin Jin[1], Dongya Gao[1], Rex Dunham[1] and Zhanjiang Liu[3], (1)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL, (2)Auburn University, Auburn, AL, (3)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY
Disease resistance is one of the most important traits for aquaculture industry. Enteric septicemia of catfish (ESC), caused by the bacterial pathogen *Edwardsiella ictaluri*, causes enormous economic losses for the domestic catfish industry every year. However, molecular mechanisms of disease resistance to ESC are still not clear. In this study, three significant quantitative trait loci (QTL), with two of them located on LG1 and one on LG26, and three suggestive QTL located on LG1, LG3, and LG21, respectively, were identified to be associated with ESC resistance by genome-wide association study (GWAS) using the 690K catfish SNP arrays. With a well assembled reference genome sequence, genes around the involved QTL regions can be readily identified. On these genes, 37 genes had known functions in immunity, making them potential candidate genes for ESC resistance in channel catfish. Notably, *nlrc3* and *nlrp12* were also reported for ESC resistance in hybrid catfish, suggesting this QTL was operating within channel catfish populations and within interspecific hybrid populations. Many of the genes with functions in immunity were involved in Class I MHC pathway of the mediated antigen processing and presentation, indicating that this pathway was significantly associated with ESC resistance in channel catfish. This study validated one QTL previously identified using the fourth generation of the backcross progeny populations and F2 generation of backcross progenies, and identified five additional QTL

among channel catfish families. Taken together, it appears that there are only a few major QTL for ESC disease resistance, making marker-assisted selection an effective approach for genetic improvements of ESC resistance.

W046: Aquaculture

Differential Expression of Apoptosis Pathway Gene Families in Response to Immune Challenge in *Crassostrea gigas* and *Crassostrea virginica*

Erin M. Roberts and Marta Gomez-Chiarri, University of Rhode Island, Kingston, RI

The eastern oyster, *Crassostrea virginica,* and Pacific oyster, *C. gigas*, are affected by disease outbreaks which threaten industry sustainability and ecosystem function. Oysters rely on a complex innate immune system characterized by several significantly expanded innate immune gene families. The genetically controlled pathway of programmed cell death, apoptosis, plays significant roles in immunity. Expansion of gene families involved in apoptosis is confirmed in *C. gigas* and supported by *C. virginica de novo* transcriptomic studies. The role of apoptosis in disease resistance in these species is unknown. This study utilized the recently available *C. virginica* genome to perform differential gene expression analysis on publically available transcriptomes of oysters challenged with a variety of stimuli, including Oyster Herpesvirus OsHV-1, several *Vibrio* spp., *Micrococcus luteus*, *Alliroseovarius crassostreae* CV919-312 (cause of Roseovarius Oyster Disease), and the probiotic bacterium *Bacillus pumilus* RI-695. Intra-species analysis reveals unique apoptosis pathway responses between challenges, and changes in functionally enriched pathways identified by Gene Set Enrichment Analysis. Bacterial challenge significantly enriched metabolic processes and chromosomal maintenance genes. Viral challenge significantly enriched signal transduction and apoptosis inhibitors, with Inhibitor of Apoptosis 2 (IAP2) and suppressor of cytokine signaling 2 most significantly differentially expressed. Transcripts from the expanded families IAP and GTPase of the Immune Associated Proteins (GIMAP) show distinct patterns of expression depending on the nature of the immune challenge. Cross-species comparison of apoptosis pathway expression indicates promising genetic targets for pathway manipulation under disease challenge and potential candidates for disease resistance markers.

W047: Aquaculture

GWAS for Detecting QTL Associated with Columnaris Disease in Two Rainbow Trout Breeding Populations

Rafael M.O. Silva[1,2], Jason P. Evenhuis[2], Roger L. Vallejo[2], Guangtu Gao[2], Kyle E. Martin[3], Ignacy Misztal[1], Timothy D. Leeds[2], Daniela A.L. Lourenco[1] and Yniv Palti[2], (1)University of Georgia, Athens, GA, (2)USDA-ARS-NCCCWA, Kearneysville, WV, (3)Troutlodge, Inc., Sumner, WA

The purpose of this study was to prospect genomic regions that explain large portion of the additive genetic variance for resistance against Columnaris disease (CD) in rainbow trout. Two important aquaculture populations were investigated. The National Center for Cool and Cold Water Aquaculture (NCCCWA) odd-year line, which was previously selected for bacterial cold water disease resistance; and the Troutlodge, Inc., May odd-year (TLUM) nucleus breeding population. The number of fish in the pedigree was 54,350 and 36,265, respectively; in which 8,453 and 3,986 fish had phenotypes recorded for CD resistance, respectively. Fish that survived to 21 days post immersion challenge were recorded as resistant. Genotypes for 57k SNPs (Affymetrix Axiom®) were available for 1,185 and 1,137 fish from NCCCWA and TLUM, respectively. The SNP effects and variances were estimated using the weighted single-step genomic BLUP approach for genome-wide association (WssGBLUP), which uses pedigree, genotypes, and phenotypes from genotyped and ungenotyped animals. The weighting strategy accounted for 1Mb moving SNP-windows along each of the 29 chromosomes in the reference genome. Genomic regions that explained more than 1% of the additive genetic variance were considered associated with CD resistance. A total of 13 windows located on six chromosomes were found to be associated with CD resistance in the NCCCWA population. Two windows, located at 59-60 Mb and 61-62 Mb on chromosome Omy17, explained 12% and 11.33% of the genetic variance for CD resistance, respectively. In the TLUM population, a total of 16 windows located on nine chromosomes were detected. Only three similar windows (located on two chromosomes) were detected in both populations. The results suggest that CD resistance has an oligogenic architecture, and the SNP windows found to be associated with CD are not informative enough for selection decisions across populations. In the next steps, we will assess strategies for genomic selection by predicting and comparing the accuracy of genomic evaluations generated using lower-density SNP panels and a panel composed solely from QTL-associated SNPs.

W048: Aquaculture

Genome Wide Association Analysis for Resistance to the Causal Agent of Bacterial Kidney Disease in a North American Commercial Atlantic Salmon

Melissa K. Holborn[1], Keng P. Ang[2], J.a.K. Elliott[3], Frank Powell[3] and Elizabeth G. Boulding[1], (1)University of Guelph, Guelph, ON, Canada, (2)Cooke Aquaculture, Saint John, NB, Canada, (3)Cooke Aquaculture, Oak Haven, NB, Canada

Bacterial kidney disease (BKD), caused by the pathogen Renibacterium salmoninarum, is economically costly to the Atlantic salmon farming industry. BKD mortality of infected fish may not occur for one to two years after infection, after a substantial monetary investment. Our goal was to identify associations between single nucleotide polymorphisms (SNPs) and resistance to BKD in the Saint John River strain of Atlantic salmon that could be used to select for resistant fish. 652 fish from 63 families were experimentally infected with R. salmoninarum by intraperitoneal injection. Fish were held in controlled tanks and daily mortalities were recorded. After 102 days, all surviving fish were euthanized and sampled. 576 fish were selectively genotyped on a custom 50K SNP chip designed for North American Atlantic salmon. 508 fish and 44,346 SNPs passed quality control. Genome wide association analysis was performed using the GenABEL package. A mixed model approach using a polygenic model was used to account for familial relatedness and population structure. A single SNP was found to have chromosome wide significance after Bonferroni correction when analyzing survival as the BKD resistance trait. Two SNPs were found to have chromosome wide significance after correction when analyzing time to death and the BKD resistance trait. These 3 SNPs explained a small percentage of the phenotypic variance and were located on different chromosomes, 4.0%, 3.8%, and 3.6% respectively. The results of these two association analyses indicating that BKD resistance, both as overall survival and time to death, has a polygenic trait architecture.

W049: Aquaculture

Applied Genomics for Conservation of Distinct Stocks and Phenotypic Diversity in Chinook Salmon

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

Chinook salmon (*Oncorhynchus tshawytscha*) is an anadromous fish species with considerable ecological, economic and social value, and has been a cultural icon that has sustained native people of western North America for millennia. This species has experienced dramatic long-term declines in abundance due to anthropogenic impacts, yet a broad portfolio of phenotypic diversity in natural organisms can buffer against exploitation and increase species persistence in disturbed ecosystems. In the Columbia River, genetic monitoring enables detection of distinct stock abundance and return timing to assist with management that is based on intensive genotyping with low cost, high-throughput methods (GT-seq). Genomic tools that allow for conservation of diverse phenotypes are still early in development but a novel reference genome was recently assembled for a diploid male Chinook salmon (2.36 GB) from the interior Columbia River to enable association mapping of life history variation and phenotypic traits. Whole genome resequencing of populations with distinct life history traits provided evidence that divergent selection was extensive throughout the genome within and among phylogenetic lineages, suggesting a broad portfolio of phenotypic diversity exists in this species that is related to local adaptation and life history variation. Incorporating genetic markers associated with specific traits such as thermal tolerance, premature migration, adult run-timing, and age-at-maturity into high-throughput genotyping panels is expected to allow for expanded monitoring to ensure phenotypic diversity is maintained in natural populations and unintentional artificial selection is avoided in conservation aquaculture programs.

W050: Aquaculture

Investigating the Effects of Early-Rearing Environment on Sperm DNA Methylation Programming in Hatchery Reared Steelhead (*Oncorhynchus mykiss*)

Mackenzie Gavery[1], Krista M. Nichols[2], Giles Goetz[1], Mollie Middleton[1] and Penny Swanson[2], (1)University of Washington, Seattle, WA, (2)NOAA Fisheries, Seattle, WA

Relative reproductive success studies have documented substantial fitness loss for wild steelhead after a single generation of rearing in the hatchery, but the relative contribution of genetic selection and/or environmentally-induced heritable epigenetic changes passed through the germline are relatively unknown. The aim of this work is to examine the effects of early-rearing environment on epigenetic programming in steelhead. In an initial study, we described epigenetic variation in hatchery and natural-origin (wild) steelhead from the Methow River. We identified significant differences in DNA methylation between hatchery and natural-origin fish, but also observed a high degree of epigenetic variation among individuals necessitating studies on how epigenetic and genetic variation interplay to promote such differences, and how much epigenetic variation is inherited. To limit the potential confounding effects of genetic variation, a second study using controlled genetic backgrounds and simulated ‘hatchery’ and ‘natural’ environments was performed. Steelhead embryos from 20 families were split across hatchery and natural treatments. After 8 months in the treatment environments fish were tagged and raised to maturity in a common environment. Sperm samples collected from 60 fish were analyzed using RRBS. Hierarchical clustering of genome-wide methylation patterns shows strong clustering within family regardless of rearing environment. These results highlight a major challenge in DNA methylation studies in natural populations, where population structure and kinship among samples is typically not known, let alone controlled for. Our findings emphasize the importance of understanding the effects of kinship among studied individuals in order to properly analyze and interpret DNA methylation data.

W051: Aquaculture

Egg Transcripts Associated with Family Fertility in Rainbow Trout (*Oncorhynchus mykiss*)

Hao Ma, USDA-ARS-NCCCWA, Kearneysville, WV

W052: Aquaculture

Small RNAs Involvement in *Flavobacterium psychrophilum*-Rainbow Trout Host Pathogen Interactions

Pratima Chapagain, Middle Tennessee State University, Murfreesboro, TN

W053: Aquaculture

Identification of Sexually Differentially Methylated Regions in Channel Catfish Provides Evidence of Epigenetic Control of Its Sex Determination

Yujia Yang[1], Yang Liu[1], Lisui Bao[1], Changxu Tian[1], Zihao Yuan[1], Dongya Gao[1], Rex Dunham[1] and Zhanjiang Liu[2], (1)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL, (2)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY

Channel catfish is a dominant aquaculture species in the United States. It is known to have a XY sex determination system, but environmental factors such as temperature can affect sex phenotypes. Although genotypes are determined at the time of fertilization, its sex phenotypes are uncommitted before 19 days post fertilization (dpf). In this study, we conducted methylation mapping analysis in genotypic female and male channel catfish before the commitment of sex phenotypes at 9, 12, and 16 dpf. A total of 2,683 sexually differentially methylated CpG sites were identified. Interestingly, 51.28%, 34.01%, and 42.95% of differentially methylated CpG sites were located on sex chromosome (chromosome 4) at 9, 12, and 16 dpf, respectively. The sex control region had the highest density of methylated sites, spanning a physical distance of approximately 5 Mb (Chr4:15Mb-20Mb). This region was significantly more hyper-methylated in females than in males, suggesting epigenetic regulation of sex control. A total of 1,271 genes were annotated nearby differentially methylated CpG sites, many of which had functions associated with sex determination, sex chromosome evolution, gonadogenesis, and gonad differentiation. Detailed analysis of methylation patterns within a set of genes related to sex determination such as idh2, sema4b, chd2, rasgrf1, was conducted. Preliminary results suggested that intragenic regions and promoters were drastically differentially methylated between females and males. Along with analysis of sex determination genes, this work provides insights into the mechanisms of sex determination, and provides evidence that epigenetic control is involved in the sex determination in channel catfish.

W054: Aquaculture

Development of a 50K Transcribed Gene SNP-Chip Identifies Major QTL for Growth in Rainbow Trout

Rafet Al-Tobasei[1], Ali R. Ali[2], Yniv Palti[3], Timothy D. Leeds[3] and Mohamed (Moh) Salem[2], (1)University of Alabama at Birmingham, Birmingham, AL, (2)Middle Tennessee State University, Murfreesboro, TN, (3)USDA-ARS-NCCCWA, Kearneysville, WV

Coding/functional SNPs change the biological function of a gene, therefor, may lead to identification of causative alleles within QTLs and development of genetic markers with large-effects on phenotypes. Two bioinformatics pipelines, GATK and SAMtools, were used to identify ~21K coding/functional SNPs with allelic-imbalances associated with important aquaculture production traits including WBW, muscle yield, muscle fat content, shear force, and whiteness in addition to resistance/susceptibility to bacterial cold-water disease (BCWD). SNPs were identified from pooled RNA-Seq data collected from ~620 fish, representing 98 families from a growth- and 54 families from a BCWD-selected lines with divergence phenotypes. In addition, ~29K SNPs without allelic-imbalances were strategically added to build a 50K Affymetrix SNP-chip. SNPs selected included 2 SNPs per gene from 14K genes and ~5K non-synonymous SNPs. The SNP-chip was used to genotype 1728 fish. The average SNP calling rate for samples passing QC (1641 fish) was ≥ 98.5%. GWAS analysis on 783 fish (representing 200 families from 2 generations) X 40K polymorphic markers (passing QC) identified a 25-SNP window on chromosome 13 explaining ~25% of the genetic variance of WBW. PLINK analysis on the same set of data identified 1783 SNPs significantly associated with WBW (P-value <4.99E-08). Majority of the SNPs identified by GWAS had allelic imbalances with WBW in the original SNP data set used to build the SNP-chip indicating high success rate of the bioinformatics pipelines in calling informative SNPs with allelic-imbalances from pooled samples and utility of the SNP-chip in GWAS studies in rainbow trout.

W055: Aquaculture

Advanced Black Tiger Prawn Breeding Using DNA Markers and High-Throughput Phenomics from Commercial Ponds

Mehar S Khatkar, Sydney School of Veterinary Science, The University of Sydney, and ARC Research Hub for Advanced Prawn Breeding, Camden, Australia

W056: Aquaculture

Genomic Evaluation for Harvest Weight and Residual Carcass Weight in Channel Catfish using Single-Step Genomic BLUP

André L. S. Garcia[1], Brian Bosworth[2], Geoff Waldbieser[3], Shogo Tsuruta[1], Ignacy Misztal[1] and Daniela A.L. Lourenco[1], (1)University of Georgia, Athens, GA, (2)USDA-ARS Warmwater Aquaculture Research Unit, Stoneville, MS, (3)USDA - Agricultural Research Service, Stoneville, MS

Catfish production is the largest aquaculture segment in the US. Since 2006 selection has been based on traditional BLUP evaluations and with the recent availability of genomic information the objectives of this study were: to investigate the feasibility of using genomic selection in US catfish and to identify major SNP associated with harvest weight and residual carcass weight. Phenotypes were available for harvest weight (n=27,160) and residual carcass weight (n=6020), and the number of fish in the pedigree was 36,365. After quality control, genotypes on 54,837 SNPs were available for 2911 fish. Genomic and pedigree predictions were calculated in a 5-fold cross validation approach, using single-trait models. Single-step genomic BLUP (ssGBLUP) was the method of choice for genomic predictions. Ability to predict breeding values was calculated as the correlation between adjusted phenotypes based on complete data and EBV or genomic EBV (GEBV). Inflation was assessed as the regression coefficient (b1) of adjusted phenotype on (G)EBV. The GEBV were backsolved to SNP effects and the percentage of variance explained by each SNP was calculated as SNP effect squared. Predictive ability for both traits increased 8 points and bias was reduced when genomic information was used. The proportion of variance explained by windows of 20 SNP was at maximum 2.2% for harvest weight and 3.3% for residual carcass weight. Both traits appear to be polygenic with no major SNP. Using genomic information is beneficial in catfish selection because of higher predictive abilities and it also allows to identify superior individuals within families.

W057: Aquaculture

A Strategy to Assemble High-Quality Reference Genomes for All Vertebrate Orders

Adam M. Phillippy, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

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. Long-read sequencing technologies from Pacific Biosciences and Oxford Nanopore have begun to correct this deficiency and enabled the automated reconstruction of reference-quality genomes at relatively low cost. Further combination of these technologies with complementary scaffolding and phasing approaches such as chromatin conformation capture (Hi-C) will soon enable the complete reconstruction vertebrate haplotypes. I will review our recent application of these approaches, and present a strategy for the automated assembly of high-quality vertebrate reference genomes. This strategy is under development for the Genome10K Vertebrate Genomes Project, which aims to generate a finished reference genome for all vertebrate orders within the next few years.

W058: Arabidopsis Informatics

Arabidopsis Informatics: An Overview

Blake Meyers, Donald Danforth Plant Science Center, St. Louis, MO

W059: Arabidopsis Informatics

Araport: Current Content and Future Directions

Christopher D. Town, J. Craig Venter Institute, Rockville, MD

W060: Arabidopsis Informatics
Interpreting MapMan4 Annotations *via* Araport
Marie E Bolger, IBG-2, Forschungszentrum Jülich, Jülich, Germany

W061: Arabidopsis Informatics
Metomgraph
Eve Syrkin Wurtele, Iowa State University, Ames, IA

W062: Arabidopsis Informatics
Defining the Translational Landscape using Ribosome Profiling
Polly Yingshan Hsu, Michigan State University, East Lansing, MI
The translational landscape reveals the proteins being synthesized in real time and reflects ongoing activities in the cell under specific conditions. Ribosome profiling, i.e., the deep sequencing of ribosome footprints, has emerged as a powerful method for identifying and quantifying translation events in diverse organisms. Recently, several experimental approaches and analytical tools have been developed to advance our understanding of the translational landscape and of translational regulation in gene expression. I will highlight our work on combining ribosome profiling and transcriptome assembly to uncover novel translated open reading frames (ORFs), including small ORFs in presumed non-coding RNAs (ncRNAs) and upstream ORFs in the 5' UTRs of protein-coding genes in Arabidopsis. With the improvement of both experimental procedures and computational tools, ribosome profiling could offer unparalleled sensitivity and yield detailed information, uncovering novel molecular mechanisms and improving genome annotation.

W063: Arabidopsis Informatics
Biocuration Efforts with Gene Trees
Monica C. Munoz-Torres, Phoenix Bioinformatics, Fremont, CA
Biocuration at The Arabidopsis Information Resource facilitate discovery for thousands of plant scientists around the globe. We here offer an update on our latest efforts to improve the quality and depth of the tools for our community: phylogenetics based functional assignments and phenotype integration.

W064: Arabidopsis Informatics
Helping the Community Transform High-Throughput Imaging Data into Phenotype Information
Edgar P. Spalding, University of Wisconsin-Madison, Madison, WI
Many researchers have discovered that digital imaging can capture plant phenotypes with high detail and throughput. Many researchers have also discovered that converting those digital images into useful data can be the hardest part of a phenotyping project. Arabidopsis researchers find collaborators or teach themselves to create suitable computer programs for quantifying the phenotypes present in an image. In some cases, a standard desktop computer can execute the program to produce data. In many cases, the project requires processing more images than is feasible on a desktop. Our Phytomorph project has created several image analysis tools for measuring phenotypes present in images including time series of images. The tools are designed to execute on high-throughput computing infrastructure. The community can access the tools through CyVerse's Discovery Environment. A user uploads images to the data store folder associated with their user account, then selects the appropriate analysis tool from a menu. The tool in docker form executes automatically on an HTCondor pool, returning results in CSV files to the user's data folder. Currently available tools measure seed size and shape, root growth rate, root tip angle, and calcium reporter dynamics. A tool for measuring root growth kinematics (axial strain rate profile) is near completion. Our project has also deployed tools for maize ear, cob, kernel, and tassel traits. Tools for measuring carrot shoot and root architectures are near completion.

W065: Arthropod Genomics and Genome Engineering
Targeted "*in utero*" Delivery of Cas9 for Heritable Germline Gene Editing
Jason Rasgon, Penn State, University Park, PA

W066: Arthropod Genomics and Genome Engineering
Genome Editing in the Whitefly, *Bemisia tabaci*
Peter W Atkinson[1], Simran K. Sandhu[2], Anna-Louise A. Doss[2], Michael E. Han[2] and Linda L. Walling[2], (1)University of California, Riverside, Riverside, CA, (2)University of California, Riverside, CA
The whitefly, *Bemisia tabaci*, is a significant pest of global agriculture, ranking in the top 100 of international insect pests. It is extremely polyphagous and invasive and is present in tropical, subtropical and temperate regions throughout the world. It damages plants through feeding on phloem and by transmitting pathogenic viruses to them. Whiteflies have developed resistance to chemical insecticides and new control strategies are needed to combat their ongoing spread. Whiteflies have remained refractory to genetic analysis, frustrating attempts to develop genetic-based technologies in them. We have developed rearing and microinjection procedures that enable the efficient delivery of nucleic acids and proteins to developing embryos of *B. tabaci*. We show that genome editing using the CRISPR/Cas9 system can be achieved in *B. tabaci* thereby bringing the power of contemporary genetics to this important pest of global agriculture.

W067: Arthropod Genomics and Genome Engineering
Genome Editing in Ants
Hua Yan, NYU School of Medicine, New York, NY
The chemosensory system is key to establishing and maintaining social structure in eusocial insects. Ants exhibit cooperative behaviors with an extensive dependency on communication. The perception of cuticular hydrocarbons (CHCs) as pheromones is mediated by odorant receptor

neurons (ORNs). ORNs express specific odorant receptors (ORs) encoded by a dramatically expanded *Or* gene family in ants. The biological features in a few ant species, such as *Harpegnathos saltator*, allow CRISPR-Cas9 gene targeting to generate a germline mutation. This facilitates the genetic analysis of the *orco* gene that encodes the obligate co-receptor whose mutation should significantly impact ant olfaction. Our results show that Orco exhibits a conserved role in the perception of general odorants but also a role in reproductive physiology and social behavior plasticity in ants. Surprisingly, and in contrast to other insect systems, the loss of OR functionality also dramatically reduces the development of ant ORNs and antennal lobe glomeruli. Taken together, the ant genetics will provide inroads towards understanding the function of genes in regulating complex social behavior.

W068: Arthropod Genomics and Genome Engineering
A Genomic Approach to Developing Genetic Sexing Strains *de novo*
Sheina Sim, USDA-ARS, Hilo, HI

W069: Arthropod Genomics and Genome Engineering
Antisense Oligonucleotides, FANA, for Treatment of Microbes and Arthropod Pests of Agricultural and Medical Importance

Wayne Hunter[1], Jackie Lin Metz[2,3], Kevin B. Temeyer[4], Beto Perez de Leon[4], Kristen Pelz-Stelinski[5], Andres F. Sandoval Mojica[5], Greg McCollum[6] and Veenu Aishwarya[7], (1)USDA Agricultural Research Service, Fort Pierce, FL, (2)University of Florida, Ft. Pierce, FL, (3)AUM LifeTech, Inc., Philadelphia, PA, (4)USDA, Kerrville, TX, (5)Univeristy of Florida, Citrus Research and Education Center, Lake Alfred, FL, (6)USDA, Fort Pierce,, FL, (7)AUM LifeTech, Inc, Philadelphia, PA
New synthetic biotechnologies provide potent, sequence specific methods for RNA-specific silencing making them tremendously attractive applications for human, animal, and plant therapeutics. Treatments with sequence specific, synthetic antisense oligonucleotides, FANA_ASO (2'-deoxy-2'-fluoro-D- arabinonucleic acid)_(antisense oligonucleotides) were shown to reduce bacteria pathogens in woody fruit crops, insect endosymbionts, as well as being able to suppress mRNA in arthropod vectors: the Asian citrus psyllid, *Diaphorina citri*; the Glassy-winged sharpshooter leafhopper, *Homalodisca vitripennis,* and whitefly, *Bemisia tabaci* (Hemiptera); Coleopterans (Curculionidae) and Ticks (Ixodidae). The results demonstrate that the improved stability of FANA_ASO enables broader use and applications when wanting to suppress RNA targets. Thus FANA_ASO can be used as a plant-delivered treatment to reduce plant pathogens and insect pests of fruit trees and other crops, or for the protection of domesticated animals and livestock.

W070: Arthropod Genomics and Genome Engineering
A Population-Genomic Approach to Delineate Host Plant use by an Insect Vector
Daisy (Zhen) Fu[1], Brendan Epstein[2] and William E. Snyder[1], (1)Washington State University, Pullman, WA, (2)College of Biological Sciences, Saint Paul, MN

Winged herbivorous insects can readily move between habitats and/or host plants, track high quality plant resource through space and time. Analysis of genomic variation has been one tool to understand insect dispersal patterns. The potato psyllid, *Bactericera cockerelli*, is of applied interest as a minute herbivore vectors the devastating pathogen of zebra chip disease on cultivated potato (*Solanum tuberosum*). Our previous population genomic studies (2012-13 sampling) suggested the interbreeding of psyllids hosted on potato crops and a perennial solanaceous plant, bittersweet nightshade (*Solanum dulcamara*). Meanwhile, a psyllid population from potatoes that was genetically distinct from others suggested additional potential non-crop host plants to be identified. Here we used Nextera-tagmented reductively-amplified DNA ("NextRAD") to investigate the fine-scale movement patterns of the potato psyllid from an extended sampling network in the Pacific Northwest of the USA. We identified 6,529 polymorphic loci among psyllids that were sampled between 2012 and 2016. Multiple population genetic analyses suggested that *Lycium* sp, a non-native solanaceous perennial which withstands arid environment, was more likely the source of the potato psyllids invading potato fields. This was further supported by the results that late in the season psyllids from *Lycium* sp. were more genetically similar to potato-collected psyllids than the early season comparison. Altogether, our results provide evidence that a population genomic approach form an effective means to delineate complex patterns of insect movement across landscapes.

W071: Avian Genomics - Going Wild!
Avian Influenza: The Winners and the Losers
Jacqueline Smith[1], Katrina Morris[1], Anamika Raut[2], Nikki Smith[1], Ashwin Raut[2], Lonneke Vervelde[1], Paul Digard[1] and David Burt[3], (1)The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom, (2)National Institute of High Security Animal Diseases, Bhopal, India, (3)University of Queensland, St. Lucia, Australia
Highly pathogenic avian influenza (HPAI) H5N1 affects several avian species, including domesticated chickens, turkeys, quails and guinea fowl as well as wild birds. However the response to infection varies widely. Ducks and waterfowl are often resistant i.e. they become infected but are capable of clearing the virus or carry the virus without symptoms, and act as reservoirs. In contrast, poultry are highly susceptible i.e. they become infected and are not able to clear the virus, which results in high mortality. Historically, Avian Influenza outbreaks have mainly affected domestic chickens and turkeys. Strikingly, in recent years, high mortality has also been observed in crows and ducks, previously assumed to be resistant to H5N1 infection. The mechanisms that promote high pathogenicity of these recent clades of AIV in normally resistant species are not completely understood, but may be caused by viral mutations that facilitate virus replication and dysregulation of the host immune response. Six avian species, chickens/turkeys (highly susceptible with heavy mortality), geese/pigeons (carriers with sporadic mortality) and ducks/crows (resistant to most AIV infections but having differential response to virus of different clades), were infected with H5N1 from different clades, as well as LPAIV H9N2. Global responses to infection have been studied by transcriptomic analysis of lung, ileum and brain tissues. Comparison of pathways involved in host responses in susceptible/resistant species are being used to identify putative

host resistant genes. The knowledge gained from these comparisons can be used to develop sustainable strategies to control Avian Influenza infections in domestic poultry.

W072: Avian Genomics - Going Wild!
Two Tickets to Paradise: Genome Variation in the Feral Fowl (*Gallus gallus*) of Bermuda and Kauai
Eben Gering, Michigan State University, East Lansing, MI

W073: Avian Genomics - Going Wild!
The Genomic Complexity and Diversity of a Chromosome-Wide Inversion in a Songbird
Vinicius H. da Silva[1,2], Veronika N. Laine[2], Mirte Bosse[1], Kees van Oers[2], Martijn F.L. Derks[1], Marcel E. Visser[2], Richard P.M.A Crooijmans[1] and Martien A.M. Groenen[1], (1)Wageningen University & Research Center, Wageningen, Netherlands, (2)The Netherlands Institute of Ecology (NIOO), Wageningen, Netherlands
Chromosomal inversions represent an important class of polymorphisms that are of particular interest in evolutionary studies. Inversions are found in various bird species and are associated with traits related to sexual behavior and sperm motility. Great tit (*Parus major*) is a songbird that is extensively used as a model in ecology and genetics. Using PCA clustering followed by FST on ≈500k SNPs from 2,296 birds, we identified a pericentric inversion overlapping ≈90 % of chromosome 1A (64.2 Mb). This inversion is present in a heterozygous state in 5 % of the population (117 birds). We identified a 60 kb copy number gain that is associated with the inversion and located close to the tentative downstream breakpoint of the inversion. This copy number gain (≈10 copies), validated using whole genome sequencing data, results in an increase of the size of chromosome 1A of ≈600 kb. The allele distribution in the inverted phase differs throughout the chromosome, suggesting some degree of recombination/mutation in a block with approximately 30 Mb in the middle of the inversion. As recombination is unexpected between inverted and normal phases, more recent haplotypes of the inversion may have a genomic interval in the middle which is in the normal orientation. However, further analyses are required to clarify existing haplotypes, structural complexity and the biological effect of the inversion.

W074: Avian Genomics - Going Wild!
Comparative Genomics and Genome Evolution in Birds of Paradise
Stefan Prost, Program for Conservation Genomics, Stanford University, Stanford, CA

W075: Avian Genomics - Going Wild!
Genomics of Hybridization and the Maintenance of Islands of Divergence: Mallards & American Black Ducks of Eastern North America
Philip Lavretsky, University of Texas at El Paso, El Paso, TX
Recent radiations are often complicated by extensive sharing of genomic polymorphisms due to ancestry and/or gene flow that stifle genetic diagnosability. Without species specific markers, statistical diagnosability requires increased sampling efforts of individuals and the genome. Here, I present the most comprehensive molecular study of mallards (*Anas platyrhynchos*) and American black ducks (*A. rubripes*; "black duck") to date. Using double-digest restriction site associated DNA sequencing (ddRAD-seq) across 290 samples, I recovered 3,200 ddRAD-seq loci, covering the Z-sex chromosome and 27 autosomal chromosomes, as well as the mitochondrial DNA (mtDNA) control region. Our data set provided sufficient statistical support to successfully differentiate between these two closely related and previously genetically undiagnosable species. Next, a more accurate assessment of hybridization and gene flow in our dataset was performed by first simulating and establishing expected assignment probabilities for F1 hybrids and nine generations of backcrosses. Using simulated indices, I re-categorized samples by assignment probabilities and found that ~80% of all phenotypically identified mallards and black ducks, as well as only ~60% of phenotypically identified hybrids were correct. Next, using our genetically vetted dataset, I report no identifiable genetic structure within black ducks, but find genetic structure among mallards that I characterize as western and non-western. Given that the genetics of eastern North American mallards has been potentially complicated by the well documented release of game-farm mallards on the eastern seaboard since the early part of the 20th century, I provide several lines of evidence that are consistent with demarcated non-western mallards originating from game-farm stock. Given that all collected samples were from birds harvested in the wild, and not from shooting preserves, our data strongly suggests that released game-farm mallards have not only significantly contributed to the genetics of their wild counterparts, but perhaps have established a viable wild [feral] population. Finally, using genetically vetted black ducks and western mallards, I demarcated several putative speciation regions on the Z sex chromosome, as well as autosomal chromosomes 1-5, 12, and 21. I conclude that despite high rates of observed interspecific hybridization between these species in the mid part of the 20th century, our results do not support the predicted hybrid swarm and subsequent genetic extinction of the black duck. Instead, I conclude that secondary contact between these two closely related species may have reinforced or pushed them closer to the completion of speciation.

W076: Avian Genomics - Going Wild!
Comparative Analysis Examining Patterns of Genomic Differentiation across Multiple Episodes of Population Divergence in Birds
Kira E. Delmore[1], Juan Lugo[2], Benjamin Van Doren[3], Max Lundberg[4], Staffan Bensch[4], Darren Irwin[5] and Miriam Liedvogel[2], (1)Max Planck Institute for Evolutionary Biology, Plön, Germany, (2)Max Planck Institute for Evolutionary Biology, Ploen, Germany, (3)Edward Grey Institute, Oxford, United Kingdom, (4)Lund University, Lund, Sweden, (5)University of British Columbia, Vancouver, ON, Canada
Heterogeneous patterns of genomic differentiation are commonly documented between closely related populations and there is considerable interest in identifying factors that contribute to their formation. These factors could include genomic features (e.g., areas of low recombination) that promote processes like linked selection (positive or purifying selection that affects linked neutral sites) at specific genomic regions.

Examinations of repeatable patterns of differentiation across population pairs can provide insight into the role of these factors. Birds are well suited for this work, as genome structure is conserved across this group. Accordingly, we re-estimated relative (F_{ST}) and absolute (d_{XY}) differentiation between eight sister pairs of birds that span a broad taxonomic range using a common pipeline. Across pairs, there were modest but significant correlations in window-based estimates of differentiation (up to 3% of variation explained for F_{ST} and 26% for d_{XY}), supporting a role for processes at conserved genomic features in generating heterogeneous patterns of differentiation. This suggestion was reinforced by linear models identifying several genomic features (e.g., gene densities) as significant predictors of F_{ST} and d_{XY} repeatability. F_{ST} repeatability was higher among pairs that were further along the speciation continuum (i.e., more reproductively isolated), suggesting that early stages of speciation may be dominated by positive selection that is different between pairs and replaced by processes acting according to shared genomic features as speciation proceeds.

W077: Banana Genomics

Mosaic Genome Structure and Chromosome Segregation in Polyploid Interspecific Plantain Bananas and Derived Breeding Accessions

Franc-Christophe Baurens[1], Guillaume Martin[1], Catherine Hervouet[1], Célestin Njembele[2], Remy Habas[1], Frederic Salmon[3], Sébastien Ricci[4], Mathieu Rouard[5], Nabila Yahiaoui[1] and **Angelique D'Hont**[6], (1)CIRAD, Montpellier, France, (2)CARBAP, Douala, Cameroon, (3)Cirad, Capesterre, France, (4)CIRAD, Capesterre, France, (5)Bioversity International, Montpellier, France, (6)CIRAD, UMR AGAP, Montpellier, France

Almost 40% of the world banana production relies on triploid interspecific cultivars deriving from hybridization between *Musa. acuminata* (Genome A, 2n=22) and *M. balbisiana* (Genome B, 2n=22). These cultivars were classified based on morphological characteristics in genomic groups (AAB and ABB) subdivided into various subgroups. They include the cooking banana Plantain group (classified as AAB) that represents 18% of the banana production worldwide. The origin of these cultivars, their chromosome structure as well as its impact on chromosome recombination and segregation are still poorly known.

We analyzed using Genotyping By Sequencing (GBS), the A/B chromosomes composition of a few cultivars and showed that it deviates in several regions from the conventional genomic classification. For example Plantain, classified as 'AAB,' has six genomic regions with an AAA chromosome composition and one entire chromosome set with an ABB composition. We compared the global chromosome structure of A and B genomes through the construction a high density SNP genetic map of *M. balbisiana* and its comparison with the *M. acuminata* reference sequence assembly. We identified a large reciprocal translocation between chromosome 1 and chromosome 3 and a large inversion on chromosome 5. We analyzed the A/B chromosomes recombination and segregation in a progeny from an 'AAAB' Plantain-derived tetraploid breeding accession. This revealed frequent recombination between A and B chromosomes all along the genome to the main exception of the inverted segment on chromosome 5. We observed 62% of aneuploids in the progeny that mainly involved the three chromosomes displaying large structural variations between A and B genomes. Implication of these results will be discussed.

W078: Banana Genomics

Detection of Variety Specific Alleles and Correlation to their Phenotype: A Proof of Principle in the Bioversity International Collection

Sebastien Carpentier, KULeuven, Leuven, Belgium

W079: Banana Genomics

Resistance Gene Candidates to *Fusarium oxysporum* f.sp. *cubense* Race 1 Resistance in Indonesian *Musa acuminata* var. *malaccensis*

Fajarudin Ahmad, Research Center for Biology, Indonesian Institute of Sciences, Bogor, Indonesia

Resistance gene candidates to *Fusarium oxysporum* f.sp. *cubense* Race 1 resistance in Indonesian *Musa acuminata var. malaccensis*

Fajarudin Ahmad[1,2], Yuyu S Poerba[1], Hugo A Volkaert[3], Harold JG Meijer[4,] Hans de Jong[1], Henk J Schouten[5], Gert HJ Kema[4,6]
[1]Wageningen University & Research, Laboratory of Genetics, [2]Research Center for Biology, Indonesian Institute of Sciences, [3]Kasetsart University, Biotechnology for Biodiversity Research and Conservation Laboratory, [4]Wageningen University & Research, Wageningen Plant Research, [5]Wageningen University & Research, Department of Plant Breeding, [6]Wageningen University & Research, Laboratory for Phytopathology

Banana is one of the most important fruit crops in the world and a staple food in many tropical countries. The current annual production is over 100 million tons, but this is threatened by one of the most devastating fungal diseases: Fusarium wilt. The causal agent is *Fusarium oxysporum* f.sp. *cubense* (*Foc*). During the 1950s the Gros Michel banana based industry was devastated by *Foc* Race 1. However, the epidemic was quenched and the banana industry was saved by replacing Gros Michel with the Race 1 resistant Cavendish bananas. Surprisingly, the responsible gene(s) for resistance are still unknown. We used s self-compatible wild banana accession *Musa acuminata* var. *malaccensis* (*Mam*, AA, 2n=22) from the Sumatra *Musa* population to generate a mapping population and to investigate the inheritance of resistance to *Foc* Race 1. Initial greenhouse bio-assays confirmed that *Mam* is resistant to Race 1. The F1 population was generated from 272 pollinated flowers to produce 3,458 seeds and 718 of them were embryo rescued, in the end 255 genotypes survived by tissue culture. The population was genotyped (N=244) using DArTseq markers and subsequent phenotyping (N=225) revealed segregation for resistance. After strict filtering, 4,171 SNP markers were used for genetic mapping. Analyses of the genotyping and phenotyping data showed the inheritance of a single dominant resistance gene that mapped near the top of chromosome 10, based on the reference genome of doubled haploid 'Pahang', which is also a *Mam* accession. The recombination between the markers among the selected recombinants, together with the position of the putative resistance gene, were further analyzed using graphical genotyping and resulted in markers that flank a 360 kb genetic region containing at least 14 NBS-LRR like resistance gene candidates, including the identified resistance gene resistance.

Keywords: *Musa acuminata* var. *malaccensis*, *Fusarium oxysporum* f. sp. *cubense*, Race 1, banana, DArTseq, resistance gene.

W080: Banana Genomics

RNAi-Based Management of Fusarium Wilt of Banana

Shulang Fei[1], Jonathan Peters[2], Elizabeth Aitken[1] and Neena Mitter[2], (1)The University of Queensland, Brisbane, Australia, (2)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia

Fusarium oxysporum (Fo), a soil-borne fungal pathogen, is pathogenic to many crop species, causing root rot or wilt symptoms. Particularly banana, susceptible to *Fusarium oxysporum* f.sp. *cubense* (Foc), is facing huge challenge to survive in the overwhelming spread of the most aggressive Foc strain, tropical race 4 (TR4). Unfortunately, effective control methods to combat Fo remain limited. This project uses an RNA interference (RNAi) strategy, Host-Induced Gene Silencing (HIGS), to confer durable resistance against Fo. The rationale of HIGS is to introduce dsRNAs of particular Fo gene sequences into the plant so that the plant RNAi machinery is activated upon infection to inhibit the expression of targeted essential Fo genes, resulting in reduced virulence of the pathogen. To date four conserved regions from Fo gene coding sequences have been chosen as HIGS targets. Fo *in vitro* culture treated with dsRNAs corresponding to these regions showed significantly inhibited growth. Fo inoculation assay on transgenic Arabidopsis lines expressing these dsRNAs presented significant improvement with regards to survival rate and reduced wilting symptoms. The generation of the expected functional siRNAs was detected by small RNA northern blot. The *in planta* expression level of the targeted genes will be tested to confirm the resistance is conferred by the RNAi mechanism as hypothesized. A Red Fluorescence Protein (RFP)-expressing Fo strain will also be employed to investigate the Fo infection patterns in the transgenic lines. As the chosen target sequences are highly conserved among a series of Fo formae speciales which are the cause of a range of host crops, the HIGS system established in this study is expected to confer resistance in all of these hosts.

W081: Banana Genomics

The Benefits, Challenges and Prospects of Genomic Prediction in Polyploid Banana

Moses Nyine[1,2], Brigitte Uwimana[1], Nicolas Blavet[3], Eva Hřibová[3], Helena Vanrespaille[4], Michael Batte[1], Violet Akech[1], Allan Brown[5], Jim Lorenzen[6], Rony Swennen[4,5] and Jaroslav Dolezel[3], (1)International Institute of Tropical Agriculture, Kampala, Uganda, (2)Faculty of Science, Palacký University, Olomouc, Czech Republic, (3)Institute of Experimental Botany, Olomouc, Czech Republic, (4)Laboratory of Tropical Crop Improvement, Division of Crop Biotechnics, Katholieke Universiteit,, Leuven, Belgium, (5)International Institute of Tropical Agriculture, Arusha, Tanzania, United Republic of, (6)The Bill & Melinda Gates Foundation, Seattle, WA

The interploidy breeding approaches practiced in banana limit the application of classical marker assisted selection strategies. Yet, there is an ultimate need to improve the efficiency of conventional crossbreeding and reduce the selection cycle to respond more rapidly to abiotic and biotic stresses. The development of sequencing and genotyping technologies such as genotyping by sequencing (GBS) are leveraging the breeders to explore genomic prediction-based approaches. In this work, the performance of six genomic prediction models was evaluated in banana under different cross validation strategies using data from a genomic selection training population comprising 307 genotypes. The population consisting of diploid, triploid and tetraploid genotypes was phenotyped under two different field management conditions and genotyped using GBS. Sequence data were processed through a bioinformatics workflow and single nucleotide polymorphisms (SNPs) were called using the genomic analysis tool kit (GATK). A custom R script was developed to process the SNP data prior to input into the models. The genotypic data were both bi-allelic SNP and allele dosage SNP markers. The total number of SNP markers varied from 5574 to 10807 depending on cross-validation strategy. Phenotypic data collected for four years on 15 traits under plant stature, suckering behavior, black leaf streak resistance, fruit bunch and fruit filling were used in cross validation. We compared the effect of accounting for allele dosage in SNP markers on the predictive ability of genomic prediction models. The results permit the evaluation of benefits, challenges and prospects of applying genomic prediction in banana, an important polyploid clonally propagated crop.

W082: Banana Genomics

Musabase : A Phenotyping and Breeding Database for Banana

Guillaume J. Bauchet[1], Nicolas Morales[1], Margaret Karanja[2], Alex C. Ogbonna[1], Bryan Ellerbrock[1], Rhiannon Crichton[3], Inge Van den bergh[3], Robooni Tumuhimbise[4], Brigitte Uwimana[5], Allan Brown[6], Trushar Shah[2], Rony Swennen[7] and Lukas A. Mueller[1], (1)Boyce Thompson Institute, Ithaca, NY, (2)IITA, Nairobi, Kenya, (3)Bioversity International, Montpellier, France, (4)National Agriculture Research Organization, Kawanda, Uganda, (5)International Institute of Tropical Agriculture, Kampala, Uganda, (6)International Institute of Tropical Agriculture, Arusha, Tanzania, United Republic of, (7)KU Leuven, Leuven, Belgium

Banana is a key staple crop for million people around the world, notably in Africa. Despite its important production, banana has not reach its full breeding potential. Several characteristics including various ploidy level, genetic groups, multiple and long developmental cycles, plant size, biotic and abiotic stresses, make banana breeding among the most complex and challenging ones across the plant kingdom.
Taking an integrative approach, musabase (https://musabase.org/) is a central repository for banana breeding phenotypic data collection, genotypic data. It ambitions to bridge gaps between different breeding programs and existing databases, notably the Musa Germplasm Information system (MGIS), the database for global ex situ –held banana genetic resources, Field Book App and ODK collection tools. Through its web interface, Musabase is a globally accessible resource, enhancing data sharing and communication within banana research community in Africa and overseas. User feedback drives new developments and improvements through an active communication system. Current tools cover phenotyping data and trait ontology dictionary, breeding management (trial, trait, pedigree search, barcode, crosses). All current developments and data are open source and available to the community (https://github.com/solgenomics/). In this workshop, we will provide an overview of the tools offered by Musabase. We will highlight challenges encountered and future developments.

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], John Vogel[1] and Daniel S. Rokhsar[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 phytozome.jgi.doe.gov. 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
New Capabilities and Technologies in Plant Functional Genomics at the Joint Genome Institute
Juna Lee, Joint Genome Institute, Walnut Creek, CA

W085: BER Plant Genomic Science
Brachypodium Encode – Deciphering the Regulation of Drought Control
Sarit Weissmann[1], Madeline A. Wiechert[2], John Gierer[2], Philip J. Ozersky[2], Michael J. Mohan[2], Kerrie W. Barry[3], Jeremy Schmutz[4] and Todd Mockler[1], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)Donald Danforth Plant Science Center, Saint Louis, MO, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)HudsonAlpha Institute for Biotechnology, Huntsville, AL

Brachypodium distachyon is a model plant closely related to less genetically tractable feedstock grasses such as switchgrass and Miscanthus and important crops like wheat, rye, and barley. We initiated the *B. distachyon* ENCODE (Encyclopedia of DNA Elements) to build a comprehensive map of functional elements involved in drought stress in the Brachypodium genome. We chose a susceptible accession of *B. distachyon*, Bd3-1, that showed an intense and immediate phenotypic and transcriptional response to drought. We detected distinct differences in regions of open chromatin between the treatments that combined with gene expression patterns, suggests the activation/inactivation of specific genes in response to drought. Combined with a detailed array of Histone modifications, we will create a public database that will portray the regulatory landscape of this important model crop.

W086: BER Plant Genomic Science
Genome-Wide Association Mapping using High-Density Drone-Based Phenotyping for *Sorghum bicolor* Biomass Traits under Drought Conditions
Jennifer Spindel, DOE Joint Genome Institute, Walnut Creek, CA

W087: BER Plant Genomic Science
Open Green Genomes: Expanding Plant References across the Kingdom
Jim Leebens-Mack, University of Georgia, Athens, GA

W088: BER Plant Genomic Science
The DOE Systems Biology Knowledgebase: KBase for Plant Research
Robert W. Cottingham, Oak Ridge National Laboratory, Oak Ridge, TN

The U.S. Department of Energy Systems Biology Knowledgebase (KBase, http://kbase.us) provides a computational environment to tackle the grand challenge of systems biology: predicting and designing biological function at scales ranging from the biomolecular to the ecological. KBase enables researchers to collaboratively generate, test, compare, and share hypotheses about biological function. Researchers can upload their own data or access public data to execute and share customized, ordered analyses that target their specific systems biology hypotheses. In KBase, computational experiments are captured in dynamic, interactive documents called Narratives that promote collaboration and reproducibility of scientific results.

This overview will briefly cover plant resources available in KBase including commonly used data types and reference data integrated with associated analysis tools and capabilities for visualization, exploration, and predictive analysis designed to accelerate understanding of microbes, plants, and their communities.

The KBase platform has extensible analytical capabilities that currently include (meta)genome assembly, annotation, comparative genomics, transcriptomics, and metabolic modeling. KBase supports the sharing and integration of reference and experimental data with analysis tools that enable researchers to build new knowledge, interpret missing information necessary for predictive modeling, test hypotheses, design experiments, and share findings that can be reproduced and extended by other researchers.

KBase tutorials (http://kbase.us/tutorials) such as *Build Plant Metabolic Model* provide a good starting place to understand how KBase might be useful in your research and how to get started.

W089: Big Data: Manage your data before your data kills you

Big Data, Small Graduate Student: Developing Data Management Skills for Graduate Students

Caryn Johansen, University of California Davis, Davis, CA

Big data and multi-lab collaborations are exciting opportunities for advancing science in new directions, but also can be rife with pitfalls and distractions. Sharing big data both between groups and within a lab presents unique problems, and these problems can be particularly challenging for a graduate student entering an established collaborative effort. Organization, reproducibility and user-friendliness should be made priorities in order for collaborative efforts to be effective, and for ease of on-boarding new collaborators. As a graduate student in two data-heavy labs and with experience building in-house databases, I will discuss my own lessons learned on personal project organization and reproducibility, and on working with collaborators. These lessons are from more than five years of personal experiences learning to work with large data sets, and the from the collective experiences of the Ross-Ibarra lab at University of California at Davis. I will outline simple, immediately applicable methods that principal investigators or graduate students can begin using to avoid mistakes concerning personal data and project organization, and for working with and sharing data with collaborators. The ability to generate data quickly, easily and cheaply is exciting and can propel research to new areas and answer so-far unasked questions. However, without deliberate practice of data management, labs can easily be bogged down with wasted effort, lost data, and dramatic miscommunication. Here, with cautionary tales of cringe-worthy mistakes, I will present some thoughts on simple steps both labs and graduate students can take to become big data masters.

W090: Big Data: Manage your data before your data kills you

Obstacles Faced when Merging and Integrating SNP Data

Anne Brown, USDA-ARS, Ames, IA

In both plant and animal research the number of dense genotyping studies and the associated data has significantly increased. At the same time the call to share SNP data has increased because integrating genotypic data across experiments can lead to improved GWAS studies. Comparisons between these datasets are complicated, as challenges remain due to the lack of consistency in how the data is represented and/or organized. The complexity of these challenges ranges from simple variation in chromosome naming (i.e. Chr vs Gm) to more complicated challenges such as the reference allele not matching the reference genome used. These challenges can be tackled with proper knowledge and training in how to handle large SNP datasets. In this talk I will cover the obstacles I faced when trying to combine and compare soybean genotyping datasets from multiple sources and how these obstacles can be avoided.

W091: Big Data: Manage your data before your data kills you

The Importance of Getting Genome Assemblies into Genbank, and How to Do It

Margaret Woodhouse, ISU, Ames, IA, Ethalinda Cannon, Iowa State University, Ames, IA and Carson M Andorf, USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA

Given the large volume of genomes now being sequenced and assembled, it is more important than ever to standardize genome nomenclature, quality, and metadata. Essential to this process is the submission of genomes to GenBank, which has quality-control procedures and metadata requirements to ensure each submitted genome meets a minimum standard. Having genomes available at a central location in a common format, with consistent metadata, and correct provenance, enables researchers to easily find and use genome assemblies in their research. This talk discusses the important aspects of GenBank genome data submission, and outlines protocols and procedures to do so efficiently.

W092: Big Data: Manage your data before your data kills you

Increasing the Long-Term Value of your Research Data: Examples from HymenopteraMine and BovineMine

Christine G. Elsik, Division of Animal Sciences, University of Missouri, Columbia, MO; Division of Plant Sciences, University of Missouri, Columbia, MO

The availability of high-throughput genomic technologies has accelerated the publication of genomic datasets and fostered the growth of research communities. Often multiple research groups work toward similar goals, so it makes sense that researchers frequently wish to compare their data with published results. Repositories at NCBI and EBI, as well as the availability of supplemental data sets on journal websites, provide opportunities for meta-analysis. Many organism-focused data resources curate large data sets to facilitate their re-use. I will demonstrate how meta-analysis is performed using HymenopteraMine (http://hymenopteragenome.org/hymenopteramine) and BovineMine (http://bovinegenome.org/bovinemine). I will also describe the challenges we have faced both in curating data and in locating sufficiently described and suitably formatted data sets to perform meta-analysis. I will show you how to increase the long-term value of your research data by ensuring that it is optimally formatted and described.

W093: Big Data: Manage your data before your data kills you

Managing Data for Non-Model Plant Systems: Integrating Resources for Genomic, Phenotypic, and Environmental Interactions

Jill L. Wegrzyn, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

The challenges of big data include storage, access, and integration. These hurdles are not domain specific and genomics is not immune to their impact. As the research community generates increasing volumes of sequence and phenotypic data through more efficient and cost effective means, the challenge of integration and access becomes more daunting. The TreeGenes database provides a web-based genomics and phenomics resource for over 1700 forest tree species. The majority of these species do not have a well assembled or annotated reference genome. Studies of relevance to the forest tree research community include transcriptomics as well as reduced representation genomics, with experimental designs based on: genotyping by sequencing, genotyping assays, or exome capture. The majority of these studies are published without any reference to the source data aside from the original sequence reads in NCBI's Sequence Read Archive (SRA). This results in a substantial body of work that is not reproducible or extensible as the intermediate assemblies, SNP calls, and/or phenotypic values/measures are not captured. To contend with these challenges, TreeGenes has developed a workflow in Tripal that is dedicated to association genetics and landscape genomics studies. The Tripal Plant PopGen Submission module (TPPS) captures metadata, georeferenced tree locations, intermediate assemblies, variant calls, phenotypic definitions/values, and associated environmental data. The submission eases the burden on

the researcher by providing transparent integration of community derived metadata standards. Here, we will demonstrate the power of integrated metadata for non-model plant species, as implemented for forest trees in TreeGenes.

W094: Big Data: Manage your data before your data kills you
Towards Making the "Big Data" Useful - Developing Databases in Terms of Data Re-Synthesis for the Ease of Use of the Big Data in the Future
James Reecy and Zhiliang Hu, Iowa State University, Ames, IA
The accelerated growth of livestock QTL/association/correlation data in the past 20+ years has positively impacted animal production industries. This phenomenal data increase has mainly been due to advances in molecular biology and the efforts to elucidate the genes underlying economically important traits in livestock species. The new data from animal genomic studies includes phenotypes, genotypes, and their associations validated by various studies. Given the increasing number of animal traits to examine (hundreds), the expanding number of genomic structural variations being accumulated from active genome studies (hundreds of thousands of SNPs per genome), and the number of additional factors involved in analysis, such as breeds, environments, locations, times, etc., the growth in data is exponential. However, the usefulness of the resulting “big data” varies greatly depending on how the data are harvested, saved, and archived. For example, there is a huge difference between organized versus unorganized, curated versus uncurated, and annotated versus unannotated data, in terms of data reusability. Our experiences developing the Animal QTLdb and CorrDB have emphasized that a properly designed and continually improved database, and diligent data curation efforts, are essential to making big data useful. This is because only when the data attributes and data links are well annotated and data integrity preserved, can the historical big data from genetic studies be useful for data comparisons and combined analyses, from which new knowledge may be synthesized. We consider this database development and data curation to be data resynthesis processes because properly identified data from different sources and experiments can be viewed differently and used in meta-analysis to draw new conclusions. Therefore, the importance of laying a strong foundation for the aggregation of big data in a scientific manner, such as in database development practices, can never be over-emphasized.

W095: Bioenergy Grass Genomics
A Chromosome-Scale *Miscanthus sinensis* Genome
Therese Mitros, UC Berkeley, Berkeley, CA

W096: Bioenergy Grass Genomics
Genomics of *Miscanthus sinensis*.
Tim Weijde, Andres Torres, Claire L Alvim Kamei, Oene Dolstra and **Luisa M Trindade**, Wageningen UR Plant Breeding, Wageningen, Netherlands
Miscanthus is a perennial energy grass characterized by a high productivity and efficiency in the use of natural resources, such as water and nutrients. It is therefore an interesting feedstock for the production of cellulosic biofuels and a wide range of other biobased (high-value) products. However, the large-scale commercialization of converting biomass into cellulosic biofuel is hindered by our inability to efficiently deconstruct the plant cell wall. The plant cell wall is a complex and dynamic structure and its components are extensively cross-linked into a rigid matrix. During my talk I will provide insights into the variation in biomass quality properties in miscanthus. I will also discuss our understanding of the molecular, genetic and environmental factors influencing its conversion efficiency into biofuel and provided tools to exploit these factors to expand the use of miscanthus as a lignocellulose feedstock

W097: Bioenergy Grass Genomics
Domestication of *Miscanthus* through Genomic Index Selection
Gancho T. Slavov, Rothamsted Research, Harpenden, United Kingdom, Chris Davey, Institute of Environmental Biological and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, United Kingdom, Paul R. H. Robson, Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, United Kingdom, Iain Donnison, Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth University, United Kingdom and Ian J Mackay, NIAB, Cambridge, United Kingdom
Miscanthus has potential as a biomass crop but the development of robustly high-yielding varieties that are superior to the natural hybrid *M. x giganteus* has been challenging, largely because of strong G x E interactions with respect to climatic conditions. We demonstrate that a combination of index selection and genome-wide prediction can overcome the limitations imposed by the inherent complexity of biomass yield. Furthermore, we explore the potential of this approach to simultaneously achieve multiple breeding targets, while also monitoring correlated responses for non-target traits. Finally, we review progress on genome-wide association studies and prediction in an expanded experimental population, including 484 *M. sinensis*, 326 *M. sacchariflorus* and 100 *M. sinensis* x *M. sacchariflorus* hybrid genotypes.

W098: Bioenergy Grass Genomics
Utilizing the Sorghum Pan-Genome to Accelerate Candidate Gene Discovery and Breeding Approaches
Scott Lee, Danforth Center for Plant Science, St. Louis, MO

W100: Bioinformatics
Towards Haplotype-Resolved Genomes with Canu
Sergey Koren[1], Arang Rhie[1], Sarah B Kingan[2], Timothy P.L. Smith[3], John Williams[4] and Adam M. Phillippy[1], (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)Pacific Biosciences, Menlo Park, CA, (3)USDA, ARS, USMARC, Clay Center, NE, (4)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia

A complete and accurate genome sequence forms the basis of all downstream genomic analyses. While most sequenced animal or plant genomes are diploid or polyploid, reference assemblies collapse them to a single sequence that is not likely to accurately represent any existing haplotype. This artificial merging masks variation and confounds downstream analysis. Long-read and long-range sequencing technologies have begun to correct this deficiency by providing de novo phased assemblies. However, the phasing blocks are often short and only represent a fraction of the genome. We developed a novel method for resolving haplotypes during the assembly process, which is capable of creating independent, completely phased haplotype assemblies at genome-scale. We demonstrate our method on both a plant F1 hybrid and the human genome. Each haplotype is recovered accurately in its entirety. Combining this method with complementary scaffolding and phasing approaches, such as chromatin conformation capture (Hi-C), may soon enable the complete reconstruction of vertebrate haplotypes.

W101: Bioinformatics
Proximity Ligation Data for Genome Scaffolding and Haplotype Resolution.
Richard E. Green, Dovetail Genomics, Santa Cruz, CA

W102: Bioinformatics
In Pursuit of Perfect Genome Sequencing: Accurate Identification and Phasing of Structural Variations using Short, Long, and Linked Reads
Michael C. Schatz, Johns Hopkins University, Baltimore, MD

W103: Bioinformatics
Deep Learning Approach to Eukaryotic Promoter Prediction
Tatiana Tatarinova, University of La Verne, La Verne, CA
Computational analysis of promoters is hindered by the complexity of their architecture. In less studied genomes with complex organization, false positive promoter predictions are common. Accurate identification of transcription start sites and core promoter regions remains an unsolved problem. We present a comprehensive analysis of genomic features associated with promoters and show that probabilistic integrative algorithms-driven models allow accurate classification of DNA sequence into “promoters” and “non-promoters” even in absence of the full-length cDNA sequences. These models may be built upon the maps of the distributions of sequence polymorphisms, RNA sequencing reads on genomic DNA, methylated nucleotides, transcription factor binding sites, as well as relative frequencies of nucleotides and their combinations. Positional clustering of binding sites shows that the cells of *Oryza sativa* utilize three distinct classes of transcription factors: those that bind preferentially to the [-500,0] region (188 “promoter-specific” transcription factors), those that bind preferentially to the [0,500] region (282 “5′ UTR-specific” TFs), and 207 of the “promiscuous” transcription factors with little or no location preference with respect to TSS. For the most informative motifs, their positional preferences are conserved between dicots and monocots. We propose a machine learning approach to predict position of transcription start site discuss accuracy and applications.

W104: Bioinformatics
Improved Automatic Gene Prediction
Mark Borodovsky, Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA
The text will be submitted before the deadline on December 31

W105: Bioinformatics
The Critical Assessment of Function Annotation: A Large Scale Collaboration to Improve Computational Annotation Methods
Iddo Friedberg, Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA

W106: Brachypodium Genomics
BrachyNet: mRNA Expression Atlas and Comparative Co-Expression Analyses for Brachypodium
Marek Mutwil, Nanyang Technological University, Singapore, Singapore and Richard Sibout, Sebastian Proost, Bjoern Oest Hansen, Neha Vaid, Federico M. Giorgi, Severine Ho-Yue-Kuang, Frédéric Legée, Laurent Cézart, Oumaya Bouchabké-Coussa, Camille Soulhat, Nicholas Provart, Asher Pasha, Philippe Lebris, David Roujol, Herman Hoft
While *Brachypodium distachyon* is an emerging model for grasses, no expression atlas and gene co-expression network is available. Such tools are of high importance to provide insights into the function of Brachypodium genes. We present a detailed Brachypodium expression atlas, capturing gene expression in its major organs at different developmental stages. The data was integrated into large-scale co-expression database (gene2function.de), enabling identification of duplicated pathways and conserved processes across ten plant species, thus allowing genome-wide inference of gene function. We highlight the importance of the atlas and the platform through the identification of duplicated cell wall modules, and show that a lignin biosynthesis module is conserved across angiosperms. We identified and functionally characterized a putative ferulate 5-hydroxylase gene through overexpression of it in Brachypodium, which resulted in an increase in lignin syringyl units and reduced lignin content of mature stems, and led to improved saccharification of the stem biomass. Our Brachypodium expression atlas thus provides a powerful resource to reveal functionally related genes, which may advance our understanding of important biological processes in grasses.

W107: Brachypodium Genomics
Genomic Dissection of Nonhost Resistance to Wheat Stem Rust in *Brachypodium distachyon*

Rafael Della Coletta[1], Candice N. Hirsch[2], Aaron Lorenz[2], Matthew N. Rouse[3] and David F. Garvin[4], (1)University of Campinas, Campinas, Brazil, (2)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (3)USDA-ARS Cereal disease laboratory, St. Paul, MN, (4)USDA-ARS, Plant Science Research Unit, St. Paul, MN

Wheat stem rust caused by the fungus *Puccinia graminis* f.sp. *tritici* (*Pgt*) is a devastating disease that has largely been controlled for decades by the deployment of resistance genes. However, new races of this pathogen have emerged that overcome many important wheat stem rust resistance genes used by breeding programs, and their spread toward major wheat production areas poses a threat to global wheat production. Nonhost resistance in plants, which provides durable and broad-spectrum resistance to non-adapted pathogens, holds great promise for helping to control wheat stem rust, but the genetic and molecular basis of nonhost resistance is poorly understood. This study employed the model plant *Brachypodium distachyon* (Brachypodium), a nonhost of *Pgt*, to genetically dissect nonhost resistance to wheat stem rust. Using bulked segregant analysis, next-generation sequencing, cumulative allele frequency differences and statistical analysis, seven quantitative trait loci (QTL) that contribute to stem rust resistance were identified in a recombinant inbred population derived from a cross between two Brachypodium genotypes with differing levels of resistance. The QTL effects vary in their magnitude, and act both additively and in some cases interact, indicating that the resistance is genetically complex. The delineation of regions of the Brachypodium genome that harbor these QTLs will guide future research aiming to identify genes essential to the nonhost resistance response and their mechanisms of action.

W108: Brachypodium Genomics

Detoxification of Deoxynivalenol and Resistance to FHB : From the Model Cereal Species *Brachypodium distachyon* to Bread Wheat

Miriam Gatti[1], Jean-Claude Pasquet[1], Frédéric Choulet[2], Catherine Macadre[1], Florence Cambon[3], Florence Guerard[1], Thierry Langin[3] and Marie Dufresne[1], (1)Institute of Plant Sciences Paris-Saclay, Orsay, France, (2)GDEC, INRA, UCA, Clermont-Ferrand, France, (3)GDEC, INRA, Clermont-Ferrand, France

Fusarium head blight (FHB) caused by fungi of the *Fusarium* genus is a widespread disease of wheat (*Triticum aestivum*) and other small-grain cereal crops. The main causal agent of FHB, *Fusarium graminearum*, can produce mycotoxins belonging to type B trichothecenes, such as deoxynivalenol (DON) that can negatively affect humans, animals and plants. Several quantitative trait loci (QTLs) for resistance to FHB have been identified some of which have been correlated with efficient DON detoxification, through the conjugation of DON into DON-3-*O*-glucose (D3G), a reaction catalyzed by UDP-glucosyltransferases (UGTs). Nevertheless, only few studies have conducted functional analyses to directly correlate DON glucosylation and resistance *in planta* and none were performed on wheat UGT gene(s). The search for UGT candidates able to conjugate DON into DON 3-*0*-glucoside (D3G) in the cereal model species *Brachypodium distachyon* resulted in the identification of the *Bradi5g03300* gene. Functional analyses of this gene showed increased sensitivity of the mutant lines to the toxin and to *F. graminearum*. Furthermore, lines overexpressing this gene showed a tolerance to the toxin and quantitative resistance to the fungal pathogen. These results were positively correlated with the detection of increased amounts of D3G, further reinforcing the ability of the *B. distachyon* to conjugate DON *in planta*. Using a synteny approach between *B. distachyon* and bread wheat genomes we identified a wheat locus carrying wheat genes orthologous to the *B. distachyon Bradi5g03300* gene. One homeolog was selected as the best ortholog by examining the gene expression pattern during wheat infection. It was therefore introduced by transformation into *B. distachyon* to rapidly determine its ability to conjugate DON into D3G *in planta* and its involvement in FHB resistance. These results contribute to increase the knowledge concerning the functional relationship between DON glucosylation and FHB resistance in different cereal species and provide candidate genes to include in selection processes in wheat.

W109: Brachypodium Genomics

Molecular Response to Varying Drought Conditions in *Brachypodium distachyon*

Sarit Weissmann[1], Skyler J Mitchell[1], Cesar Lizarraga[1], Madeline A. Wiechert[1], Erica Agnew[1], John Gierer[1], Stuart Marshall[1], Kerrie W. Barry[2], Jeremy Schmutz[3] and Todd Mockler[4], (1)Donald Danforth Plant Science Center, Saint Louis, MO, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)Donald Danforth Plant Science Center, St. Louis, MO

Drought is a major environmental stress limiting the productivity of crop plants in most parts of the world. Plants have evolved several mechanisms to survive water deficit or drought stress conditions. However, most of these complex mechanisms are poorly understood, and prevent the development of new varieties with enhanced drought resistance. Brachypodium distachyon is a model plant closely related to important crops like wheat, rye, and barley and also to less genetically tractable feedstock grasses such as switchgrass and Miscanthus. We chose three representative B. distachyon accessions: Bd3-1 (highly susceptible), Bd21 (susceptible), and Bd1-1 (tolerant), for analysis of phenotypic and molecular response to drought treatments. Combining gene expression analysis and network analysis, we point the involvement of specific genes and pathways in drought response. We show that susceptible accessions of B. distachyon intense and immediate phenotypic and transcriptional response to drought, while tolerant accessions are less responsive to changes in water level. Our multidimensional phenotypic and molecular approach allows us to identify drought related genes and their regulatory elements at a high resolution.

W110: Brachypodium Genomics

New *Brachypodium* Resources to Study Polyploidy, Perenniality and Gene Function: Four New Reference Genomes and Nearly One Million Mutations.

John P. Vogel[1,2], Sean Gordon[1], Pilar Catalan[3], Bruno Contreras-Moreira[4], David M. Goodstein[1], Shengqiang Shu[1], Patrick Davidson[1], Richard D Hayes[1], Amy Cartwright[1], Richard Sibout[5], Debbie Laudencia-Chingcuanco[6], Jeremy Schmutz[7], Jerry Jenkins[7], Virginia Markham[2], Lifeng Liu[1], Joel Martin[1], Daniel Woods[8], Richard Amasino[8], John Doonan[9], Luis Mur[9] and Ana Caicedo[10], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)University of California Berkeley, Walnut Creek, CA, (3)Department of Agriculture (Botany), High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain, (4)Fundación ARAID, Zaragoza, Spain, (5)INRA-IJPB, Versailles, France, (6)USDA-ARS, Western Regional Research Center, Albany, CA,

(7)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (8)University of Wisconsin-Madison, Madison, WI, (9)Aberystwyth University, Aberystwyth, United Kingdom, (10)Biology Department, University of Massachusetts, Amherst, MA
The DOE Joint Genome Institute (JGI) plays a leading role in developing resources for the *Brachypodium* research community. This talk will present an overview of recent developments and insights learned from several sequencing projects at the JGI including: Comparisons of genomes of the allopolyploid *B. hybridum* and its two diploid progenitors, *B. stacei* and *B. distachyon*, that have revealed multiple origins for *B. hybridum* spread over time creating a natural time course to study polyploid genome evolution. The genome of the perennial species *B. sylvaticum*. The identification of over 800,000 SNP mutations (available on Phytozome) by a project that aims to sequence 2,000 sodium azide, EMS and fast neutron mutant lines. Finally, the creation of a database (BrachyPan, https://brachypan.jgi.doe.gov) that allows users to query and download a *B. distachyon* pan-genome based on *de novo* assemblies from 54 lines as well as the genomes of individual lines.

W111: Brassicas

Genome Structural Variation in *Brassica napus* Germplasm

Lenka Havlickova, Zhesi He and Ian Bancroft, Centre for Novel Agricultural Products (CNAP), York, United Kingdom
Polyploidy plays an important role in the evolution of angiosperms and has conferred genetic preconditions for successful domestication of many major plant crops, including rapeseed (*Brassica napus* L.). The difficulties in genome assembly of polyploid crops arises from the genomic complexities induced by combining two or more evolutionarily diverged genomes into a single nucleus, frequent sequence exchanges between constituent genomes and by the significant size of polyploid genomes. With an aim to capture the entire gene set in the *Brassica napus* genome, we developed a new pan-transcriptome reference sequence. This was derived from the *Brassica* A and C coding DNA sequence (CDS) gene models of the progenitor species; *B. rapa* and *B. oleracea* respectively, complemented by the interpolation of *B. napus*-specific CDS models. By having a genomics platform based on a set of genes assigned with high confidence to their genome, due to the very high density *B. napus* genetic linkage map, we have embarked upon a series of analyses aimed at understanding genome structural variation in *B. napus* resulting from traditional breeding approaches. Our newly developed genome visualization approach by Transcriptome Display Tile Plots (TDTPs) had enabled us to uncover and visualize very frequent homoeologous exchanges (HEs), which underlies trait variation in rapeseed.

W112: Brassicas

Demographic History of Morphotype Diversification in *Brassica oleracea*

Sarah D. Turner[1], Makenzie E. Mabry[1], J. Chris Pires[1] and Timothy M. Beissinger[2], (1)Division of Biological Sciences, University of Missouri, Columbia, MO, (2)USDA-ARS, University of Missouri, Columbia, MO
The vegetables encompassed by *Brassica oleracea*, commonly known as cole crops, can be categorized into six distinct and diverse morphotypes: kale, cabbage, Brussels sprouts, kohlrabi, broccoli, and cauliflower. These crops are an especially interesting model for domestication; wild mustard, a small, unpalatable plant, was selected for a wide range of appearances, flavors, and nutritional attributes. Despite widespread use of these crops to demonstrate the power of human selection on crop domestication, the demographic history of *B. oleracea* remains poorly understood. Using resequenced genomes from the major morphotypes in *B. oleracea*, we present a demographic analysis to estimate ancestral population sizes and population splits. Results from this work will expand our understanding of how humans have impacted the evolution of cole crops, providing insight into Brassica evolution and facilitating future crop improvement efforts.

W113: Brassicas

Structural Genome Variation Associates with Quantitative Disease Resistance in *Brassica napus*

Iulian Gabur, Department of Plant Breeding, Justus Liebig University Giessen, Germany, Giessen, Germany, Regine Delourme, UMR APBV, Le Rheu, France, Sebastien Faure, BIOGEMMA, Mondonville, France, Christophe Jestin, Terres Inovia, Thiverval-Grignon, France, Emmanuelle Dyrszka, Syngenta, Saint-Sauveur, France, Andreas von Tiedemann, Department of Crop Sciences, Plant Pathology and Crop Protection Division, Georg August University, Goettingen, Germany, Frank Breuer, KWS Saat AG, Einbeck, Germany, Rod Snowdon, Justus Liebig University, Giessen, Germany and Christian Obermeier, Plant Breeding, Giessen, Germany
Within the French-German GeWiDis consortium ("Exploiting genome wide diversity for disease resistance improvement in oilseed rape"), analyses of structural organization and allelic diversity associated with resistance to three important fungal oilseed rape diseases were performed. Disease resistance screening of Blackleg (*Leptosphaeria maculans*), Sclerotinia stem rot (*Sclerotinia sclerotiorum*) and Verticillium stem striping (*Verticillium longisporum*) was done using a *B. napus* Nested Association Mapping (NAM) panel in greenhouse and field experiments at different locations across Germany and France. Genome-wide association analysis (GWAS) revealed significant marker trait associations including a number of new regions for resistance to all three diseases. GWAS identified overlapping quantitative trait loci (QTL) for multiple disease resistance. To optimize the strategy of breeding multi-resistant varieties, we investigated the localization of resistance factors and their relationships with regard to polyploidy, duplications, homeologous translocations, or other structural rearrangements. Analyses of resistance factors involved in this three diseases and their positive/negative correlations in regions that are subjected to structural variation is a promising tool that allows a better understanding of the genetic basis of quantitative resistance in *B. napus*. New valuable alleles were identified from diverse Brassica genetic resources which might be useful for introgression of new resistance variability into oilseed rape.

W114: Brassicas

Tissue-Specific Transcriptome Analysis Reveals Genes Involved in the Root Biomass Increase of *Raphanus sativus*, a Root Crop in Brassicaceae

Ji-Young Lee, School of Biological Sciences, Seoul National University, Seoul, South Korea
A root serves as an essential organ in plant growth by up-taking nutrients and water from soil and supporting the rest of a plant body. Some plant species also utilize roots for storage. Many of these, including sweet potatoes (*Ipomoea batatas*), cassava (*Manihot esculenta*), and radish (*Raphanus sativus*) are important crops, however how they increase root biomass has remained elusive. Radish is a close relative of *Brassica*,

and extensively cultivated in eastern Asia. Radish cultivars selected in these regions show a remarkable increase of root biomass in short growth periods. We thus characterized radish root growth in details and found that the cambium zone established before radish roots initiate active radial growth is the key place that affects root growth and yields. To discover genes important for the radish root growth, we selected two radish inbred lines that show contrasting root radial growth and yields. We then established laser capture microdissection techniques and performed Illumina sequencing for RNA libraries prepared from cambium and neighboring tissues in the radish roots, which were harvested after 5, 7, and 9 weeks post seed planting. From these data, additional 3,515 protein coding transcripts were found, which have not been annotated in the radish genome database (http://radish-genome.org/). From a total of 50,029 transcripts, 4,602 transcripts were selected as the ones differentially expressed in the cambium and further classified based on their expression dynamics. Gene groups showing high expression in the cambium at the stage of rapid root biomass increase were enriched with genes involved in growth, development, and cell cycle regulation. Interestingly, our analysis indicated that a contrasting difference in root biomass of two selected inbred lines might be related to the differential expression of genes involved in stress responses. Collectively, our tissue-specific transcriptome profiling data provide a novel insight into the yield-related genes in a representative root crop.

W115: Brassicas

Genetic Diversity in Oilseed Rape Root Morphology

Christian Hermans[1], Julien Louvieaux[2], Laszlo Kupcsik[1], Jiajia Xu[1], Rod Snowdon[3] and Nathalie Nesi[4], (1)Université Libre de Bruxelles, Brussels, Belgium, (2)CARAH/HEPH-Condorcet and Université libre de Bruxelles, Ath, Belgium, (3)Justus Liebig University, Giessen, Germany, (4)UMR 1349 IGEPP, INRA, le Rheu, France

Nitrogen (N) is the quantitatively most important nutrient in cropping systems. However, a considerable N fraction is lost through runoffs with detrimental consequences for the environment and human health. One way to reduce N fertilizer input is to breed for crops with better Nitrogen Use Efficiency (NUE). Increasing the plant N uptake by optimizing the degree of root branching for exploring a larger soil volume in search of the mobile nitrate resource may contribute to that purpose. Rapeseed (*Brassica napus* L.) is a major oil crop showing a poor NUE, which makes its production highly dependent on mineral N fertilization. Our aim is to understand the genetic control of root system architecture and how it is impacted by nitrate supply. We are currently exploiting the genetic diversity of root morphology in a panel of >400 double haploid accessions. Seedlings grew for seven days on vertical germination paper imbibed with a nutrient solution containing low or moderate nitrate concentrations. Seedlings cultivated with low nitrate supply developed more numerous and longer lateral roots than those grown with high supply. There was a wide diversity within the accessions for the root morphological traits. We are identifying genomic regions associated those traits by performing Genome Wide Association Studies. That information is complemented with root transcriptome sequencing data which will allow detecting gene copies differentially expressed between accessions and between N nutrition conditions. A measurable outcome in the mid-term will be to provide genetic markers for selecting new crop genotypes for smart farming.

W116: Brassicas

Analysis of Transcriptional and Epigenetic Regulation in Hybrid Vigor of *Brassica napus* uncovers Key Roles for Small RNAs

Longjiang Fan, Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, Hangzhou, China

W117: Buffalo genomics

Introduction

John Williams, Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia

W118: Buffalo genomics

Chromosome Level Assembly of the Water Buffalo Genome

Lloyd Low, Davies Research Centre, Roseworthy, Australia, Rick Tearle, University of Adelaide, School of Animal and Veterinary Science, Roseworthy, Australia, Derek M. Bickhart, Dairy Forage Research Center, USDA-ARS, Madison, WI, Benjamin D. Rosen, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, Sarah B Kingan, Pacific Biosciences, Menlo Park, CA, Timothy P.L. Smith, USDA, ARS, USMARC, Clay Center, NE, Paolo Ajmone Marsan, Inst. of Zootechnics, Università Cattolica del S. Cuore, Piacenza, Italy and John Williams, Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia

Water buffalo (*Bubalus bubalis*), also known as Asian buffalo, is a globally important species for agriculture. A good reference genome of this species will facilitate understanding its biology, managing genetic diversity, and applying new genome-based selection methods for genetic improvement. We previously reported a *de novo* assembled genome based on 454 and Illumina short reads. Due to the limitations of using only short reads in the assembly, the publicly available genome consists of 366,983 scaffolds, with a scaffold N50 of 1,412,388 bp and an L50 of 581 scaffolds. Here we report an improved diploid assembly of the same individual (Olimpia) using PacBio Sequel and RSII reads (> 69-fold coverage), assembled with the FALCON-Unzip assembler. Despite using a highly inbred animal with a high homozygosity level, 58% of its genome could be phased with a haplotig N50 of 0.394 Mbp and the longest haplotig is 2.77 Mbp. Scaffolding using Chicago and HiC reads was done using the Dovetail HiRise software along with custom scripts to remove false contig breaks that were erroneously introduced. Improvement in the latest assembly scaffold N50 is 117,187,264 bp which is 83-fold longer than the previous short read based assembly, and has an L50 that is made up of only 9 scaffolds. The current draft of the assembly is still evolving as we continue to correct potential mis-orientation within scaffolds and examine conservation of synteny with the cattle and goat genomes, before final gap filling, assembly polishing and annotation.

W119: Buffalo genomics

Comparative Sequence Alignment with Cattle Reveals Water Buffalo Structural Genomics Differences

Wenli Li, U.S. Dairy Forage Research Center, Agricultural Research Service, USDA, Madison, WI, Derek M. Bickhart, Dairy Forage Research Center, USDA-ARS, Madison, WI, Luigi Ramunno, Department of Agriculture, University of Naples "Federico II", Portici (NA), Italy, Daniela Lamartino, Parco Tecnologico Padano, Via Einstein, 26500, Lodi, Italy, Italy, John Williams, Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia and George E. Liu, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

Water buffalo (*Bubalus bubalis* L.) is a significant livestock species worldwide with high economic importance. Like many other livestock species, buffalo lacks a high quality and contiguous reference genome assembly which is necessary for in high resolution comparative genomics studies. There are two subspecies of domestic water buffalo: river buffalo and swamp buffalo. This study sought to characterize differences in gene content, regulation and structure between taurine cattle and river buffalo using the extensively annotated UMD3.1 cattle reference genome as a basis for sequence comparisons. After alignment of 14 river buffalo WGS datasets to the cattle reference, we identified 127 deletion CNV regions representing 5 annotated cattle genes, which are present in at least 13 of the investigated buffalo individuals. We also characterized 583 merged mobile element insertion (MEI) events within the upstream regions of annotated cattle genes. To assess the functional impact of the structural differences identified, we performed comparative gene expression analysis between cattle and buffalo using publicly available RNA-sequencing data. Transcriptome analysis in various tissue types in river buffalo confirmed the absence of four cattle genes that are predicted to be completely deleted in buffalo. Four genes which may be related to phenotypic differences in meat quality and color, had upstream MEI predictions and were found to have significant elevated expression in river buffalo compared with cattle. Our comparative alignment approach and gene expression analysis suggested a functional role for many genomic structural variations, which may contribute to the unique phenotypes of river buffalo.

W120: Buffalo genomics

Allele-Specific Expression of Macrophage Expressed Genes in Water Buffalo

Prasun Dutta[1], Rachel Young[1], Lucas Lefèvre[1], Stephen J. Bush[1], Eileen Wall[2] and David A. Hume[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)Scotland's Rural College (SRUC), Edinburgh, United Kingdom

Tuberculosis or TB (caused by *Mycobacterium bovis*) and brucellosis (caused by *Brucella abortus*) are zoonotic diseases affecting livestock species, including water buffalo (*Bubalus bubalis*). This project aims to identify and validate polymorphisms in genes associated with susceptibility and resistance to TB and brucellosis in Indian domestic water buffalo. As both pathogens replicate inside macrophages, genes associated with differences in susceptibility are likely to be expressed specifically in these cells. Genes with enriched expression in macrophages were identified using an atlas of gene expression, created using RNA-sequencing data and spanning multiple tissues and cell types. DNA and RNA sequence data was used to identify protein-coding variation in macrophage-expressed genes, including potential null mutations. Variants that affect the expression level of macrophage-enriched genes were detected by their allelic imbalance; that is, unequal levels of transcription from two alleles. The extreme of allelic imbalance, monoallelic expression, is when one allele is not expressed at all. This could arise by imprinting or a loss of function. In these cases, we examine the corresponding genomic sequence to explore the likely mechanism. Overall, this work will explore the prevalence of allele-specific expression in buffalo macrophages, and show how allelic imbalance can help identify *cis*-acting regulatory variations in candidate macrophage-expressed genes.

W121: Buffalo genomics

Identification of Genetic Markers Associated with Beta Carotene Levels in Buffalo and Dairy Cattle Milk: An Opportunity to Improve Milk Quality in India

Francesca Bertolini[1], Josue Chinchilla Vargas[1], J.R. Khadse[2], Alok Juneja[2], Prasad Deshpande[2], Vinod Potdar[2], Kaustub Bhave[2], A.B. Pande[2] and Max F. Rothschild[1], (1)Department of Animal Science, Iowa State University, Ames, IA, (2)BAIF Development Research Foundation, Pune, India

Beta-carotene, a precursor of vitamin A, can be consumed in buffalo and cattle milk but intake is too low in developing countries. The main genes related to beta-carotene/vitamin A production are *BCMO1*, *BC02* and *SCARB1*. In this work, a Sequenom panel was built using Single Nucleotide Polymorphisms (SNPs) derived by next generation sequencing data from the coding sequence of the three genes. The panel included 44 SNPs for cattle and 23 SNPs for buffalo. A total of 1,421 buffalo (Jaffarbadi, Murrah, Pandharpuri, Mehsana, Surti) and 2,312 cattle (Sahiwal, Tharparkar, Gir, Jersey-cross and Holstein-cross) from India were genotyped with this panel and milk beta-carotene content was measured. The beta-carotene analyses demonstrated a significant difference among cattle and buffalo with buffalo showing a lower beta-carotene content. Among buffalo breeds, the average of beta-carotene levels range from 1.19 mcg/100ml in Mehsana to 4.25 mcg/100ml in Jaffarbadi. For Mehsana and Surti, no suggestive or significant association has been detected. For Pandharpuri, SNPs in both the *SCARB1* and *BCMO1* genes were associated with beta-carotene levels. For Murrah, SNPs in *BCMO1* and *BC02* and for Jaffarbadi, at least one SNP for each gene were associated with beta carotene levels. Several SNPs of the three genes in cattle breeds were associated as well. These markers may be useful to develop genetic selection strategies that can increase beta-carotene content in milk in Indian animals and the panel may be useful also in other developing countries. Funding for this project was kindly provided by Bill & Melinda Gates Foundation.

W122: Buffalo genomics

Genome-Wide Association Study on Milk Production Traits in Philippine Dairy Buffaloes

Ester Flores, Philippine Carabao Center, Philippines

A dense SNP panel for buffalo was tested on four different swamp and riverine buffalo breeds that are available in the Philippines. The number of polymorphic SNPs in the swamp buffaloes were considerably lower at <20,000 compared to >50,000 for the riverine buffalo breeds. Moderate linkage disequilibrium (LD) at 60-70 kb pairwise distance was estimated for the riverine buffaloes. This suggests that the SNP panel was dense enough for genome-wide association studies (GWAS) and genomic selection. Previous work on GWAS on buffaloes involved the Italian and Brazilian buffalo breeds. For the Philippine dairy buffaloes, GWAS was done using the genotypes from the Axiom 90k Buffalo

Genotyping Array and the first parity milk yield (MY), fat yield (FY) and protein yield (PY) record of 503 buffalo cows from institutional herds managed by the Philippine Carabao Center (PCC). GWAS was done using the lm function of R to associate MY, FY and PY with the genotypes. Bonferroni correction of 5%was applied resulting in 3 significant SNP markers. Two SNPs were associated with MY (AX-85102445 and AX-85152230) and one was associated with PY (AX-85118021). Heritabililty estimates were 0.23, 0.13 and 0.27 for MY, FY and PY, respectively. The correlation of genomic breeding values with that of BLUP breeding values derived from random regression model was 0.51, 0.28 and 0.46 for MY, FY and PY, respectively. The identification of significant SNPs associated with MY and PY can potentially be used in marker assisted selection. This preliminary result should be validated further using more data.

W123: Cacao Genomics Workshop

Transcriptome Sequencing of Two *Theobroma cacao* Varieties Reveals Candidate Resistance Genes to *Phytophthora megakarya*

Noah P. Winters[1], Désiré Pokou N'da[2], Andrew S. Fister[1], Tahi G. Mathias[2], Klotiolonan Coulibaly[2], Aswathy Sebastian[1], Eric K. Wafula[1], Claude W. dePamphilis[1], James H. Marden[1], Siela N. Maximova[1] and Mark J. Guiltinan[1], (1)The Pennsylvania State University, University Park, PA, (2)Centre National de Recherche Agronomique, Abidjan, Cote d'Ivoire

Black pod (PB) disease of *Theobroma* cacao is caused by the oomycete pathogen *Phytophthora megakarya* and is one of the most serious threats to cacao cultivation in Africa. Genetic resistance to PB has been documented, but identification of causal loci has been limited. Here, resistance to PB was investigated by performing a time course RNA-sequencing experiment that examined model resistant (SCA6) and model susceptible (NA32) cacao cultivars at 0, 6, 24, 48 and 72 hours post *P. megakarya* inoculation. We identified 5213 genes as differentially regulated in response to *P. megakarya* infection in the resistant and susceptible cultivars. Among those differentially regulated genes, several were implicated in pathogen detection and response, including nucleotide binding leucine rich repeat receptors (NB-LRRs), receptor-like kinases (RLKs) and mitogen activated protein kinases (MAPKs). This transcriptome-wide analysis gives insight into the molecular and genetic basis of host defense against *P. megakarya,* and will contribute to the identification of loci governing resistance in this important agricultural crop.

W124: Cacao Genomics Workshop

Validation of Low Cost, Long Read Sequencing for Crop Pangenomics

Joe Morrissey[1], Joseph C. Stack[2], Rebecca Valls[1] and Juan Carlos Motamayor Arias[3], (1)Mars Inc, Miami, FL, (2)Mars, Inc., Miami, FL, (3)MARS Inc., Miami, FL

The Minion is a portable, long read sequencer that can run off of a standard laptop computer. This versatility opens the possibility of genomics-guided breeding in the remote and under-resourced locations where cacao is grown. Here we validate the Minion's application to crop genomics by creating a de novo genome assembly of the heterozygous 'Pound 7' clone of cacao, and compare it to the previous two cacao reference genomes. We found that generating a high quality assembly of a heterozygous plant genome was possible for less than $5000, in contrast to previous cacao reference genomes, which required many years and millions of dollars to produce.

W125: Cacao Genomics Workshop

Application of CRISPR/Cas9 Mediated Genome Editing to Enhancement of Disease Resistance in Cacao

Andrew S. Fister, Lena Landherr, Siela Maximova and **Mark Guiltinan**, The Pennsylvania State University, University Park, PA

Theobroma cacao, the source of cocoa, suffers significant losses to a variety of pathogens resulting in reduced incomes for millions of farmers in developing countries. Development of disease resistant cacao varieties is an essential strategy to combat this threat, but is limited by sources of genetic resistance and the slow generation time of this tropical tree crop. In this study, we present the first application of genome editing technology in cacao, using Agrobacterium-mediated transient transformation to introduce CRISPR/Cas9 components into cacao leaves and cotyledon cells. As a first proof of concept, we targeted the cacao *Non-Expressor of Pathogenesis-Related 3 (TcNPR3)* gene, a suppressor of the defense response. After demonstrating activity of designed single-guide RNAs (sgRNA) *in vitro*, we used *Agrobacterium* to introduce a CRISPR/Cas9 system into leaf tissue, and identified the presence of deletions in ~30% of *TcNPR3* copies in the treated tissues. The edited tissue exhibited an increased resistance to infection with the cacao pathogen *Phytophthora tropicalis* and elevated expression of downstream defense genes. Analysis of off-target mutagenesis in sequences similar to sgRNA target sites using high-throughput sequencing did not reveal mutations above background sequencing error rates. These results confirm the function of NPR3 as a repressor of the cacao immune system and demonstrate the application of CRISPR/Cas9 as a powerful functional genomics tool for cacao. Several stably transformed and genome edited somatic embryos were obtained via *Agrobacterium*-mediated transformation, and ongoing work will test the effectiveness of this approach at a whole plant level.

W126: Cacao Genomics Workshop

Identification of QTLs Associated with Frosty Pod Rot and Black Pod Rot Resistance in an F_1 Population of *Theobroma cacao*

Osman A. Gutierrez[1], Alina S. Puig[2], Wilbert Phillips[3], Donald Livingstone[4], Guiliana Mustiga[5], Stefan E. Royaert[6], Raymond J. Schnell[7] and Juan Carlos Motamayor Arias[7,8], (1)USDA ARS SHRS, Miami, FL, (2)USDA-ARS Subtropical Horticulture Research Station, Miami, FL, (3)Centro Agronomico Tropical de Investigacion, Cartago, Costa Rica, (4)Mars Inc, Miami, FL, (5)MARS, MIAMI, FL, (6)Mars Center for Cocoa Science - MARS Inc., Itajuipe, Brazil, (7)MARS Inc., Miami, FL, (8)MARS Inc, Miami, FL

Cacao (*Theobroma cacao* L.) is an understory tree cultivated worldwide by small farmers in tropical regions whose center of origin is the Amazon Basin. Yearly this crop suffers severe losses due to pests. In the Americas, frosty pod (FP) [*Moniliophthora roreri* (Cif. and Par.)], and black pod (BP) [*Phytophthora palmivora* (Butl.) Butl.] caused significant yield reductions. The use of traditional measures such as fungicide applications and pod removal to control the spread and damage has not been very successful due to the increase in production costs as well as environmental concerns. The utilization of genomics-assisted breeding methodology is the most effective way to develop cacao clones with

resistance to both diseases. The goals of this study were to: (i) find single nucleotide polymorphism markers linked with quantitative trait loci for frosty pod and black pod resistance, and (ii) place associated single nucleotide polymorphism markers to specific chromosomes. A cross between 'Pound 7', a clone susceptible to frosty pod and resistant to black pod, with 'UF 273', which is resistant to frosty pod and moderately susceptible to black pod, was made at CATIE, Turrialba, Costa Rica and produced a heterozygous F_1 mapping population of 179 individuals. F_1 progenies were genotyped using 5,149 single nucleotide polymorphism markers in a 5K Illumina chip and screened for resistance to these two diseases for5 years. A dense linkage map composed of 10 linkage groups was constructed using 2,944 single nucleotide polymorphic markers. Single and composite interval mapping was used to use to map quantitative trait loci for resistance to frosty pod and black pod. Results show that two major and seven minor QTLs associated with FP and BP resistance were identified. These loci associated with disease resistance were located in chromosomes 2, 4, 7, 9 and 10. Further fine mapping is being conducted to identify candidate resistance genes within mapped quantitative trait loci.

W127: Cacao Genomics Workshop

The Genomes and Transcriptomes of the Major Cacao Pathogens: Our Current State of Knowledge and Prospects for their Future Use.

Bryan Bailey[1], Shahin Ali[2] and Lyndel Meinhardt[1], (1)Sustainable Perennial Crops Lab, USDA-ARS, Beltsville, MD, (2)Sustainable Perennial Crops Lab, USDA-ARS

Progress is being made in determining the genome and transcriptome sequences of the major pathogens of cacao. The *Moniliophthora perniciosa* genome was published in 2008, and other than the sequence of the cacao swollen shoot virus, represented the first major cacao pathogen genome released. In 2014, the *M. roreri* genome and transcriptome sequence was published, verifying the close genetic relationship between the two *Monilophthora* species causing disease on cacao. Early in 2016, the *Phytophthora palmivora* and *Phytophthora megakarya* genomes and transcriptomes were published together, here also showing the close genetic relationship between these two *Phytophthora* species. Two unique findings from this study include the observation that *P. palmivora* is a tetraploid having twice as many functional genes as is typical for the genus and the fact that *P. megakarya* has specifically amplified gene families that target/suppress plant defense (effectors). The knowledge concerning effectors in *Phytophthora* species is in sharp contrast to that of the *Moniliophthora* species. Although candidate effectors have been identified in the *Moniliophthora* species, we have little current knowledge of their putative functions. Several additional cacao pathogen genomes are in the process of being assembled and should soon be available. All these sequencing efforts have taken advantage of collaborations. When combined with the increasing knowledge of the *Theobroma cacao* genome, a unique opportunity exists to dissect the interactions between multiple plant pathogens on a single host and identify their species specific and shared aspects. By understanding these interactions genetic resources for managing diseases can be optimally exploited.

W128: Cacao Genomics Workshop

Differential Response of *Phytophthora palmivora* and *Phytophthora megakarya* to Temperature Stress.

Alina S. Puig, USDA-ARS Subtropical Horticulture Research Station, Miami, FL

Phytophthora megakarya and *Phytophthora palmivora* cause black pod rot of cacao, one of the most economically significant diseases affecting this crop. They are closely related but differ in terms of aggressiveness and response to temperature. *P. palmivora* has a higher temperature maximum (34°C) than *P. megakarya* (30°C). In culture, *P. megakarya* lost viability after 3 days at 32°C while *P. palmivora* viability began decreasing after 3 days at 36°C. This study provides insight into the effects of temperature stress on these pathogens and the underlying resistance mechanisms by monitoring expression changes in genes involved in abiotic stress responses in other species. When exposed to heat, genes encoding heat shock proteins and chaperones were most commonly induced in both species. When inducible in *P. megakarya*, these genes were almost exclusively heat responsive, tending toward lower expression at cold temperatures. In *P. palmivora* these heat responsive genes also tended to be induced at low temperatures (9°C and 11°C). Although there were some temperature induced differences in metabolite profiles between the species, *P. palmivora* maintained higher metabolite pools overall than *P. megakarya*. Due to these differences between *P. megakarya* and *P. palmivora*, temperature is likely to influence epidemiology and the potential of these pathogens to spread to new areas.

W129: Cacao Genomics Workshop

Evidence for Extreme Ongoing Diversification within the Cacao Swollen Shoot Badnaviral Species Complex

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

Cacao swollen shoot virus (CSSV) *Cacao swollen shoot CD virus* (CSSCDV) and *Cacao swollen shoot Togo A virus* (CSSTAV) are recognized causal agents of cacao swollen shoot disease (CSSD) in West Africa. Diverse leaf and shoot-swelling symptoms have been associated with CSSD since its discovery nearly a century ago, however, failure of various diagnostic tests to detect badnavirus in as many as 50-70% symptomatic trees, has led to the recent hypothesis that additional, undiscovered badnaviral pathogens may be involved. To address this possibility, total DNA was purified from symptomatic cacao samples from Cote d'Ivoire, Ghana, and Nigeria using the CTAB method, and subjected to Illumina sequencing, which requires no *a priori* sequence knowledge. All apparently full-length genome sequences were assembled, and ranged in size from 6.8-7.2 kilobase pairs. The sequence accuracy was confirmed using sequence-specific abutting primers for PCR amplification, and Sanger DNA sequencing with primer walking. Pairwise nucleotide (nt) identity analysis of the Illumina sequences with 7 GenBank reference CSSD-associated badnavirus genomes revealed 70-100% shared identity, indicating substantial divergence. Pairwise distance analysis of the respective RT-RNase H region sequence from each genome indicated 72-100% shared identity, indicative of five distinct species. The genome arrangement comprises either four, five, or six predicted open reading frames with similar conserved protein domain (CPD) architectures. Among the 63 available CSSD-badnaviral full-length genome sequences, the following five species are represented: Cacao red vein virus, Cacao red vein-banding virus, and the previously recognized CSSCDV, CSSTAV, and CSSV. Additional species are expected to be discovered genomic pathology tools.

W130: Cacao Genomics Workshop
Cacao Breeding in Colombia, a Trip in History
Caren Dayana Rodriguez Medina, Corpoica, Palmira, Valle del Cauca, Colombia

W131: Camelids
Variation in Expression of κ-Casein Gene (CSN3) and its Genetic Polymorphism in *Camelus dromedaries* Population of United Arab Emirates
Abdullah Al-Mutery, University of Sharjah, Sharjah, United Arab Emirates

W132: Camelids
A *de novo* Hybrid Assembly of a Dromedary Camel
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], Adel Khalili[6], Douglas Antczak[2] and Samantha A. Brooks[1], (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
The dromedary camel is of great economic and cultural importance in many countries, often selected for meat and milk production, draught, riding, and racing traits. The global population is estimated to be over fifteen million, many of which are in developing nations. Camels possess an array of unique physiological adaptations that have enabled their survival in an arid desert environment. The two currently available dromedary reference genomes were assembled using only Illumina short reads, containing roughly thirty thousand scaffolds with a N50 of single digit megabases. Our goal was to produce an assembly incorporating long read technologies to improve contiguity, which will facilitate genomic selection and a comparative genomic study of desert adaptation in mammals. We selected a male dromedary camel from the US to serve as our genome reference, establishing a fibroblast cell line as a continuous source of high molecular weight DNA. Hybrid assembly incorporated 74x of paired end Illumina reads and 15x PacBio long reads. Scaffolding was accomplished by incorporating 30x coverage of 10X Genomics Chromium sequencing. Scaffolds were organized into putative chromosomes with comparative genomics using the RACA method. Resulting chromosomes were evaluated and assigned using alpaca genetic maps and dromedary radiation hybrid maps. A *de novo* transcriptome assembly was available to assist with gene annotation. The resulting assembly is a valuable resource for future genomic studies in the dromedary camel.
This study was made possible in part by NPRP Grant 6-1303-4-023 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

W133: Camelids
Identification of Genetic Variants Affecting Color and Phenotypic Variation between Dromedary Populations Using Whole Genome Sequencing (*Camelus dromedarius*)
Fahad Alshanbari, Texas A&M University, College Station, TX
Coat color serves many purposes in the wild, but is also a target of selection by humans. The genetic mechanisms underlying coat color vary by species, but are typically regulated by few major genes and further modified in pattern and intensity by other genes, which ultimately regulate levels and distribution of eumelanin and pheomelanin. The melanocortin 1 receptor (MC1R) and agouti signaling protein (ASIP) are associated with coat color variation, but nothing has yet been documented in the dromedary. Recently, we showed that a nonsense mutation in *ASIP* is associated with black, and a missense mutation in *MC1R* with the white color in the dromedary. TaqMan assay analysis in large cohorts suggests additional modifier genes to produce a range of different shades of black/brown/red/beige/white. We pooled DNA by color phenotype from 17 black, 27 white and 31 red Saudi Arabia dromedaries and sequenced by Illumina NextSeq500. Using a sliding window zF_{ST} outlier approach combined with intra-population nucleotide diversity, we observed high F_{ST} (Z>8) and low nucleotide diversity surrounding *ASIP* in black compared to white and red dromedaries, confirming our prior findings. However, we did not observe this pattern surrounding *MC1R*, perhaps due to lack of strong selection or sequence complexity. Additionally, 150 putative selective sweep regions were identified, indicating that these moderately isolated populations retain genomic differentiation despite some gene flow. The artificially imposed demography leads to differences related to production traits such as milk and meat production, representing genetic resources useful for producers to assist in breeding programs.

W134: Camelids
Towards a Gene Catalogue of the Alpaca Y Chromosome
Matthew Jevit[1], Andrew Hillhouse[1], Mark F. Richardson[2], Brian W. Davis[3], Rytis Juras[1], Malcom Ferguson-Smith[4], Ahmed Tibary[5], Vladimir Trifonov[6] and Terje Raudsepp[1], (1)Texas A&M University, College Station, TX, (2)Deakin University, Geelong, Australia, (3)Texas A&M, College Station, TX, (4)University of Cambridge, Cambridge, United Kingdom, (5)Washington State University, Pullman, WA, (6)Institute of Molecular and Cellular Biology, Novosibirsk, Russian Federation
In order to improve the understanding on the genetic factors affecting male sexual development and fertility in the alpaca (*Vicuna pacos*) we initiated systematic gene discovery in the alpaca Y chromosome by cDNA selection procedure. We extracted total RNA from alpaca testis, selected mRNA and converted it into an Illumina-compatible cDNA library. Next, we flow sorted 20,000 copies of alpaca Y chromosome, amplified Y DNA and labeled it with biotin. After testing the Y DNA for specificity by FISH on alpaca chromosomes, the biotin-labeled Y DNA was hybridized with the testis cDNA library. The Y-specific testis transcripts were captured with streptavidin-coated paramagnetic beads and sequenced as 2 x 300 bp reads on the Illumina MiSeq platform. The Y-testis transcriptome was denovo assembled with Trinity software. Transcript lengths ranged from 297-11,000 bp. Analysis of the Y transcriptome by BLASTX and BLASP is in progress. We have identified several known mammalian Y genes, like *SRY*, *ZFY*, *DDX3Y*, and *TSPY*, but also those with no known mammalian homology. The latter are of particular interest because they may represent novel alpaca Y-specific sequences. Additionally, potential ampliconic and multi-copy genes are

being discovered by quantifying transcripts with the RSEM package in R. We predict that significant outliers may be indicative of the ampliconic or multicopy genes. These putative ampliconic genes will be verified using ddPCR. The Y-transcriptome study is a part of a larger project with an ultimate goal to produce an annotated sequence assembly of the alpaca Y chromosome.

W135: Camelids

Generating Y-Chromosomal Shotgun Assemblies for Old World Camelids to Study their Male Genealogies

Sabine Felkel[1], Barbara Wallner[1], Hussain Bahbahani[2], Faisal Almathen[3] and **Pamela Burger**[4], (1)Institute of Animal Breeding and Genetics, Vetmeduni Vienna, Vienna, Austria, (2)National Unit for Environmental Research and Services (NUERS), Faculty of Biological Sciences, Kuwait University, Kuwait, Kuwait, (3)King Faisal University, Al-Hasa, Saudi Arabia, (4)Vetmeduni Vienna, Research Institute of Wildlife Ecology, Vienna, Austria

Polymorphic on the markers on the male specific part of the Y-chromosome (MSY) provide useful information for tracking male genealogies. While maternal lineages are well studied in Old World camelids using mtDNA, the lack of Y-chromosomal references hampers the analysis of male driven demographies. Recently it has been shown in horses, that a shotgun assembly generated from short read next generation sequencing (NGS) data revealed sufficient resolution to trace individual male lines in this species. In a similar approach we generated MSY shotgun assemblies for *Camelus dromedarius* and *Camelus bactrianus* by *de novo* assembling NGS data enriched for Y-chromosomal reads followed by a remapping-filtering approach. The resulting assemblies are used as a reference for variant calling using short-read data from multiple individuals and to generate first Y-chromosomal phylogenies for these species.

W136: Cannabis Genomics - Advances and Applications

Cannabis Science 101.

C J Schwartz, Sunrise Genetics, Hempgene Division, Fort Collins, CO

Cannabis has a long history of human use going back millennia. Uses have included fiber for clothing and parchment, seeds for fuel and nourishment, and chemical composition for both medical and recreational purposes. Because of the intoxicating effects of Cannabis, it has long been illegal worldwide. Specifically, in the US, Cannabis is classified as a schedule one narcotic, which contradicts scientific data supporting its medicinal use. For example, a recent New England journal of medicine article provides clinical evidence that chemical components present in certain Cannabis cultivars clearly have effects on treatment for seizures in children.

Research designed to test and improve Cannabis cultivars for specific end uses has been greatly inhibited by severe Cannabis laws, even for cultivars that lack intoxicating effects, such as industrial hemp. Recent changes at the federal level in regulations for hemp and at the state level for marijuana have led to greater public awareness of the facts and misperceptions about this diverse plant. This greater understanding has stimulated Cannabis research with the major focus on the chemical composition and potential therapeutic uses. In addition, genomic data on cultivars for classification has revealed much about the origins of certain Cannabis cultivars and demonstrated a very high level of admixture, presumably due to human actions.

The legal Cannabis industry and scientific research are both currently in their infancies. Much of the initial scientific work on Cannabis now can be used as the groundwork for further studies, and has been provided by a small subset of pioneering researches with minimal funding, and in some cases at considerable risk. As science replaces superstition, the true potential of Cannabis can be realized and exploited to aid humankind.

W137: Cannabis Genomics - Advances and Applications

Chromosome-Scale Pseudomolecules of *Cannabis sativa*

Christopher J. Grassa[1,2], Jonathan P. Wenger[3], Clemon J. Dabney III[3], Todd P. Michael[4], George D Weiblen[5] and C J Schwartz[6], (1)Hempgene, Fort Collins, CO, (2)Economic Herbarium of Oakes Ames at Harvard University, Cambridge, MA, (3)Weiblen Laboratory, Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (4)J. Craig Venter Institute, Carlsbad, CA, (5)Bell Museum, University of Minnesota, Saint Paul, MN, (6)Sunrise Genetics, Hempgene Division, Fort Collins, CO

Knowledge of the linear order and recombination rate of genes facilitates both applied and pure genomic research. Our team is a cooperative of academic and industry scientists with the common goal of building this important resource. Here we present ten high-quality pseudomolecules corresponding to the chromosomes of *Cannabis sativa* and an ultra-high density map of meiotic crossovers in a carefully bred marijuana x hemp hybrid population. We assembled contigs from real-time single-molecule nanopore reads, inferred their linkage, order, and orientation from the genetic map, validated contig order and placed repeats using a library prepared from cross-linked precipitated chromatin, and polished the pseudomolecules with high-coverage short-read libraries.

W138: Cannabis Genomics - Advances and Applications

Curation and Data Management for Cannabis Germplasm and Use of XP-GWAS to Unravel Chemical Diversity

Matthew Welling[1], Lei Liu[1], Kathryn Eales[2], Jos Mieog[1], Omid Ansari[3], Tim Shapter[3], Phil Warner[3] and **Graham J. King**[1], (1)Southern Cross University, Lismore, Australia, (2)Southern Cross Plant Science, Southern Cross University, Lismore, Australia, (3)Ecofibre Industries Operations, Virginia, Australia

Given the ongoing presence of the Single Narcotics Convention and associated legislation, the management and exploitation of *Cannabis* genetic resources for multiple end-uses continues to be constrained. In order to stimulate a belated Green Revolution for this genus, we have proposed the establishment of virtual genetic resource management (Welling et al., 2016, Frontiers Plant Sci. 1113). Combining global efforts will contribute to stimulating further research and market demand in the industrial fibre, hempseed and medicinal cannabis markets. At present we are working closely with Ecofibre (EIO) to manage their unique genetic resource Collection, which includes accessions that represent a wide distribution of the extant genepool. We also are establishing an Australian national germplasm resource to include seed, DNA bank, chemical profiles and herbarium voucher specimens. We have commenced a systematic regeneration and characterization program, and are focusing on comprehensive and flexible data curation from trait to genome, implemented in the generic CropStoreDB data framework. Initial characterization of industrial hemp and recreational drug types within the Collection demonstrated a wider range of THC:CBD cannabinoid

ratios than previously reported (Welling et al., 2016 208:463-75). We are now prospecting for novel functional gene variants using a strategy (XP-GWAS), which involves whole genome sequencing of pooled extreme chemical phenotypes. We expect this approach to accelerate the rate at which uncharacterized natural cannabinoids can be metabolically engineered in order to explore their value as novel therapeutic agents.

W139: Cannabis Genomics - Advances and Applications

Research and Development of Industrial Hemp Genetics for US Production Environments and Emerging Markets

John McKay, Colorado State University, Fort Collins, CO

The 2014 US Farm Bill permits cultivation of Cannabis sativa for research to study the "growth, cultivations, or marketing of industrial hemp" in states that allow such cultivation. I will report on research and development in Colorado, a state that has legal hemp and high THC cannabis. At Colorado State University we conducted a variety trial of the major cultivars from Europe . These trials were conducted across 2 years, at 2 locations spanning the latitude and growing conditions of Colorado. I will report on the performance of these varieties, investigating the degree to which yield, terpenoids and cannabinoids are influenced by genetics and the environment. We are filling this up with genetics in bi-parental crosses to identify the QTL underlying these traits. I will also report on a genomics collaboration, comparing hemp to US marijuana strains in order to understand divergence and diversity across their genomes. I will introduce a federally funded multi-state collaboration, that includes efforts to define regions of the genome that contribute to yield, gender and other important traits. Finally, I will report on work in the private sector to create new varieties of Cannabis that fit into US production environments, emerging markets and regulation.

W140: Cannabis Genomics - Advances and Applications

Cannabis Genomics in Canada

Jason Sawler and Jonathan Page, Anandia Labs, Vancouver, BC, Canada

We are using genomics to elucidate the metabolic pathways leading to the major cannabis metabolites (cannabinoids and terpenes) and to better understand the genetic organization of the genus *Cannabis*. A major experimental approach has been the use of EST and transcriptome data derived from glandular trichomes, the specialized epidermal structures that synthesize cannabinoids. We have successfully applied trichome-focused analysis in combination with classical biochemistry to identify three enzymes of the cannabinoid pathway: hexanoyl-CoA synthetase, olivetolic acid cyclase and an aromatic prenyltransferase. A draft assembly of the ~820 Mbp genome from the marijuana strain Purple Kush, has opened up new avenues for gene discovery as shown by the identification of a novel cannabinoid synthase enzyme, cannabichromenic acid synthase. In addition, the genetic and biochemical basis of terpene production in now under investigation. We have recently used genotyping-by-sequencing (GBS) to analyze the genetic variation in 43 hemp and 81 marijuana accessions. GBS shows that hemp and marijuana are genetically distinct, and provides insight into the differentiation of marijuana into “Indica” and “Sativa” groups. As cannabis emerges from the shadow of prohibition, genomics promises both to clarify its evolutionary history and to accelerate the development of this valuable, multi-use crop.

W141: Cannabis Genomics - Advances and Applications

Cannabinoid Polymorphism in Minnesota Feral *Cannabis* Supports a Simple Genetic Model for the Inheritance of Phenotype

George D Weiblen[1], Jonathan P. Wenger[2] and Clemon J. Dabney III[2], (1)Bell Museum, University of Minnesota, Saint Paul, MN, (2)Weiblen Laboratory, Plant and Microbial Biology, University of Minnesota, St. Paul, MN

Cannabis sativa was cultivated in Minnesota during the Second World War as a source of fiber for U.S. Army canvas and cordage. The industry collapsed after the war but the plant was naturalized and persists to this day in disturbed environments throughout the state. We measured cannabinoid variation (THC and CBD) in feral plants at three locations in the Mississippi River Valley to test a genetic model for the inheritance of cannabinoid phenotype. Major cannabinoid phenotypes (CBD-dominant, intermediate, and THC-dominant) were present in each population with CBD-dominant and intermediate plants outnumbering THC-dominant plants by a factor of ten. Cannabinoid phenotype in 300 plants was perfectly correlated with allelic variation in the CBDA synthase gene. The rarity of THC-dominant plants might be attributed to either impurity of the original introductions or to recent introgression with marijuana-type Cannabis.

W142: Cassava Genomics

The Cassava Virus Action Project

Laura Boykin, University of Western Australia, Crawley, Western Australia, Australia, Joseph Canisius Ndunguru, MIKOCHENI AGRICULTURAL RESEARCH INSTITUTE, DAR ES SALAAM, Tanzania, United Republic of, Titus Alicai, National Agricultural Research Organisation (NARO), Kampala, Uganda, Peter Sseruwagi, NATIONAL CROPS RESOURCES RESEARCH INSTITUTE, DAR ES SALAAM, Uganda, Monica A. Kehoe, Department of Primary Industries and Regional Development, Bentley, Australia, Fred Donati Tairo, Mikocheni Agricultural Research Institute, Dar es Salaam, Tanzania, Elijah M. Ateka, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, Anders Savill, The University of Western Australia, Crawley, Australia and Cassava Virus Action Project Team

CVAP (www.cassavavirusactionproject.com) is a network of researchers, farmers and others, collaborating to use genomic technologies to improve the management of these Cassava viruses. For example, if you can analyze the DNA of the virus, quickly and close to the crop, you could understand what virus it is and decide what action to take. We have empowered local communities to take decisions that maximize their crops while also minimizing the spread of these whitefly-borne viruses. For the first time, farmers struggling with diseased cassava crops can take immediate, positive action to save their livelihoods based on information about the health of their plants, generated using a portable, real-time DNA analysis device. The team now plans to expand the project; 800 million people worldwide depend on the threatened cassava crop. The project aims to reduce the risk of community crop failure and help preserve livelihoods. Oxford Nanopore’s portable MinION DNA sequencer was used to identify which strain of virus was destroying the cassava crops of farmers in Tanzania and Uganda as part of a collaboration of scientists and farmers, known as CVAP. As MinION delivers the information in real time (compared to the usual three months), farmers were able to take action much faster. One was advised to destroy the crop and plant a different variety that is more resistant to

the virus for example. The team's latest work to bring portable DNA sequencing to east African farmers has been featured on CNN, BBC World News, BBC Swahili, BBC Technology News, and the TED Fellows Ideas Blog.

W143: Cassava Genomics
Engineering Cassava Brown Streak Disease Resistance through Genome Editing
Dan Lin, Donald Danforth Plant Science Center, St Louis, MO

W144: Cassava Genomics
Capturing Next-Generation Genome Wide Molecular Markers in Cassava Helps to Untangle the Crop's Genetic Improvement History
L. Augusto Becerra Lopez-Lavalle[1], Fausto Villafrade Rodriguez Zapata[2], Tatiana Ovalle[2], Manuel Ruiz[3], Anestis Gkanogiannis[2] and Joe Tohme[4], (1)International Center for Tropical Agriculture, CIAT, Cali, Valle del Cauca, Colombia, (2)CIAT, Cali, Colombia, (3)CIRAD, UMR AGAP, Montpellier Cedex 5, France, (4)International Center for Tropical Agriculture – CIAT, Cali, Colombia
The release of the cassava reference genome by Prochnik et al. in 2012 has allowed cassava geneticists identifying tens of thousands genome-wide sequence variations across multiple cultivars. These genomic variations have helped to develop a new generation of molecular markers for the crop's genetics research either by re-sequencing using restriction-site associated DNA-sequencing (RAD-seq) or by the SNPtype™ allele-specific PCR assays technology. Today CIAT's cassava program using these technologies has characterized at the DNA level more than 3000 cassava cultivars used by farmers in South East Asia or Latin America; including improved cultivars, LAC landraces, and its potential wild close relatives conserved at the world's largest collections of the crop. These next generation molecular markers along with the analytical methods implemented in the Cassava Genome Hub has allowed to analyze the crop's genetic diversity, performed population and family structure analyses, unravel the crop's phylogenetic and phylogeographic history and confirm its recent introduction histories in Africa and South East Asia. Also, using these genomic and bioinformatics resources, the SNPtype™ technology has allowed us to examine the factors affecting the adoption of improved cassava varieties in the Cauca Department in southwest Colombia, as well as, six regions of Vietnam where most of the cassava is grown. These analytical approaches have showed the power of our next-generation sequencing analytical methods could have in identifying both historical population structure and recent colonization history along with the identification of clones recently adopted including its pedigree. But more importantly, it is guiding our efforts to understand the nature of complex traits in cassava such as whitefly resistance, post-harvest physiological deterioration of the roots, starch stability and content and its resistance to diseases such as frog skin disease.

W145: Cassava Genomics
Cassava Source-Sink Relations
Uwe Sonnewald, Friedrich-Alexander-University of Erlangen-Nürnberg, Erlangen, Germany

W146: Cassava Genomics
TBA
Chiedozie Egesi, National Root Crops Research Institute (NRCRI), Umudike, Umuahia, Nigeria

W147: Cassava Genomics
Introgressed *Manihot glaziovii* Alleles Segregate in Cassava Germplasm and Influence Key Traits
Marnin Wolfe[1], Guillaume J. Bauchet[2], Ariel W. Chan[1], Ramu Punna[3], Chiedozie Egesi[4], Robert Kawuki[5], Peter Kulakow[6], Ismail Rabbi[6] and Jean-Luc Jannink[7], (1)Cornell University, Ithaca, NY, (2)Boyce Thompson Institute, Ithaca, NY, (3)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (4)National Root Crops Research Institute (NRCRI), Umudike, Umuahia, Nigeria, (5)National Root Crops Resources Research Institute, Namulonge, Uganda, (6)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, (7)USDA-ARS, Cornell University, Ithaca, NY
Cassava is a staple food crop that feeds hundreds of millions of people throughout the tropics, especially in sub-Saharan Africa. Hybridization of cassava (*Manihot esculenta*) to wild relatives has been and continues to be an important source of beneficial genetic variation for breeders. Crosses to ceara rubber tree (*Manihot glaziovii*) conducted in Tanzania in the 1930's, saved cassava production from epidemics of cassava mosaic disease (CMD). Those crosses have also reportedly contributed resistance to brown streak disease and bacterial blight. Descendants of these original hybrids have served as key parents to a large portion of today's breeding germplasm. Indeed, a recent study revealed large segments of *M. glaziovii* genome segregating in a sample of African genotypes, suggesting that historical introgressions remain important today. In our study, we leverage genome-wide marker data on a large population of diverse cassava to answer a crucial, outstanding question: Which wild genome segments, if any, contribute significantly and positively to cassava phenotypes? We (1) use ancestry informative markers to detect introgressed genome segments in breeding germplasm. (2) We combine genome-wide association analysis and genetic variance partitioning to assess the phenotypic importance of introgressed *M. glaziovii* genome segments. Results so far indicate significant genetic variance is attributable to introgressed regions for disease resistance and yield related traits. Our study will have direct implications for cassava breeding by enabling genome-wide selection for (or against) wild alleles in breeding populations.

W148: Cattle/Sheep/Goat 1
A High Resolution Atlas of Gene Expression in the Domestic Sheep
Emily L. Clark[1], Stephen J Bush[1], Mary E. B. McCulloch[1], Iseabail L. Farquhar[1], Rachel Young[1], Lucas Lefèvre[1], Clare Pridans[1], Hiu G. Tsang[1], Chunlei Wu[2], Cyrus Afrasiabi[2], Bruce Whitelaw[1], Mick Watson[1], Tom Freeman[1], Kim M. Summers[1], Alan L.

Archibald[1] and David A. Hume[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)The Scripps Research Institute, La Jolla, CA
Sheep are a key source of meat, milk and fibre for the global livestock sector. To support functional annotation of the sheep genome, and increase the available genetic and genomic resources, we have produced a high-resolution transcriptional atlas using Texel x Scottish Blackface individuals. RNA-Seq libraries were generated by Edinburgh Genomics (http://genomics.ed.ac.uk) from tissues and cells representing all major organ systems from adult sheep and from multiple juvenile, neonatal and prenatal developmental time points. The *Ovis aries* reference genome (Oar v3.1) includes 27,504 genes (20,921 protein coding), of which 25,350 (19,921 protein coding) have detectable expression in at least one tissue in the sheep gene expression atlas dataset. Network cluster analysis in Miru (http://kajeka.com) was used to describe the overall transcriptional signatures present in the sheep gene expression atlas and assign those signatures, where possible, to specific cell populations or pathways. Using the 'guilt by association' principle we were able to use the dataset to assign meaningful gene names and putative function to hundreds of previously unannotated genes. We relate the expression profiles to innate immunity by focusing on clusters with an immune signature, and to the advantages of cross-breeding by examining the patterns of genes exhibiting the greatest expression differences between purebred and crossbred animals.The Sheep Gene Expression Atlas provides a model transcriptome for ruminants, has the potential to inform future improvements in livestock productivity, efficiency and health and is a valuable resource for the international Functional Annotation of Animal Genomes (FAANG) initiative.

W149: Cattle/Sheep/Goat 1
Two New Cattle Reference Genomes for the Price of One
Arang Rhie[1], Sergey Koren[1], Sarah B Kingan[2], Timothy P.L. Smith[3], John Williams[4] and Adam M. Phillippy[1], (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)Pacific Biosciences, Menlo Park, CA, (3)USDA, ARS, USMARC, Clay Center, NE, (4)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia
Generating high quality genome assembly is challenging due to the innate heterozygosity and repetitive structure in diploid genomes. In the past, highly inbred individuals were preferably chosen for genome sequencing, to reduce heterozygosity. However, most reference assemblies are still a collapsed representation of multiple haplotypes. Long-read sequencing in combination with other long-range technologies improves the quality of genomes but are still costly. We developed a novel method for resolving haplotypes during the assembly process, which is capable of creating independent, completely phased haplotype assemblies. We demonstrate our method on a taurine x indicine cattle breed (Angus x Brahman) F1 hybrid. The assembly produced two completely resolved haplotype sequences with an initial haplotig NG50 size exceeding 20 Mbp. The improved quality of the haplotype assemblies is shown by their ability to correct thousands of large structural errors in the UMD3.1 reference, identify haplotype-specific markers, and recapitulate known repeat variation in the ruminant lineage. Combining this method with complementary scaffolding and phasing approaches, such as chromatin conformation capture (Hi-C), may soon enable the complete reconstruction of vertebrate haplotypes.

W150: Cattle/Sheep/Goat 1
The Efficient Dairy Genome Project: An International Effort to Increase Feed Efficiency and Reduce Methane Emissions in Dairy Cattle
Paul Stothard, Livestock Gentec, Department of Agricultural, Food & Nutritional Sciences, University of Alberta, Edmonton, AB, Canada

W151: Cattle/Sheep/Goat 1
Impact of the New Reference Genome on Genotype Imputation Accuracy
Robert D. Schnabel, Division of Animal Sciences, Informatics Institute, Columbia, MO

W152: Cattle/Sheep/Goat 1
Rambouillet Sheep Genome and Annotation Resources
Kim C. Worley, Baylor College of Medicine, Houston, TX
We report the high quality Rambouillet sheep reference genome and initial analysis of FAANG sample RNA sequencing from the reference ewe, Benz2616.
We *de novo* assembled 200 Gb of Pacific Biosciences (PacBio) sequence with 12.6 kb N50 sub-read length with Celera Assembler and polished with Arrow. Scaffolding the contigs using Hi-C data and Phase PGA incorporated 98.1% of the assembly into 32 large scaffolds. Scaffold gaps were filled using PBJelly, misassemblies identified with misFinder and additional gap-filling completed. Error correction using Pilon and Illumina data produced the final 2.87 Gb genome. More contiguous, complete and correct than most, the contig N50 is 2.6 Mb, with half the genome in 309 contigs (longest 16.3 Mb). Most ESTs (98% of 338,551) align to the genome, 90% with nearly complete alignments, aligning over >90% of their length. Base quality is high, (error rate <1%).
FAANG assays from over 100 collected reference animal tissues including PacBio IsoSeq, Illumina RNAseq and miRNAseq, ATAC-Seq and other assays are underway, to complement the genomic PacBio, Illumina and Hi-C sequence. RNA sequence analysis of PacBio IsoSeq, Illumina RNAseq and microRNAseq (5, 9 and 20 tissues respectively) is ongoing. IsoSeq matched ~11,000 per tissue (15,888 total) of 20,921 annotated proteins. Illumina RNAseq with ~100x more reads per tissue identified ~15,500 transcripts per tissue. MicroRNA sequences analyzed using miRDeep2 identified a total of 6,523 novel miRNAs and 659 known miRNAs, with most (471) of the known and many of the novel miRNAs similar to annotated bovine miRNAs.

W153: Cattle/Sheep/Goat 1
Using Cattle and Sheep Reference Assemblies to Develop Genomic Resources for North American Moose

Michael P. Heaton, USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE
Moose (*Alces alces*) colonized the North American continent from Asia less than 15,000 years ago, and spread across the boreal forest regions of Canada and the northern United States. Contemporary populations have low genetic diversity, due either to low number of individuals in the original migration (founder effect), and/or subsequent population bottlenecks in North America. Genetic tests based on informative single nucleotide polymorphism (SNP) markers are helpful in forensic and wildlife conservation activities, but have been difficult to develop for moose, due to the lack of a reference genome assembly and whole genome sequence (WGS) data. WGS data were generated for four individual moose from Alaska, Idaho, Wyoming, and Vermont with minimum and average genome coverage depths of 14- and 19-fold, respectively. Cattle and sheep reference genomes were used for aligning sequence reads and identifying moose SNPs. Approximately 11% and 9% of moose WGS reads aligned to cattle and sheep genomes, respectively. The reads clustered at genomic segments where sequence identity between these species was greater than 95%. In these segments, average mapped read depth was approximately 19-fold. Sets of 47,236 and 38,250 high-confidence SNPs were identified from cattle and sheep comparisons, respectively. Among the four moose, heterozygosity and allele sharing of SNP genotypes were consistent with decreasing levels of moose genetic diversity from west to east. A minimum set of 317 SNPs, informative across all four moose, was selected as a resource for future SNP assay design.

W154: Cattle/Sheep/Goat 1

Long-Read Sequencing Enables the Accurate Annotation of Immune Gene Clusters in Goat

Derek M. Bickhart, Dairy Forage Research Center, USDA-ARS, Madison, WI, John C. Schwartz, The PirBright Institute, Woking, Surrey, United Kingdom, John A Hammond, Pirbright Institute, Woking, United Kingdom, Benjamin D. Rosen, ARS, USDA, Beltsville, MD, Juan F. Medrano, University of California, Davis, CA, Adam M. Phillippy, NHGRI, NIH, Frederick, MD, Curtis P. VanTassell, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD and Timothy P.L. Smith, USDA, ARS, USMARC, Clay Center, NE
Repetitive and polymorphic immune gene clusters have remained difficult to characterize and compare among most livestock species. This represents a significant issue within the field of livestock genomics, as alleles of genes related to the adaptive and innate immune system likely influence animal health outcomes in practice yet remain difficult to track with existing genotyping tools. We use the recent long-read assembly (ARS1) of the domestic goat (*Capra hircus*) as an example of how newer sequencing technologies and assembly methods can fully characterize previously recalcitrant immune gene regions in Eukaryotic genomes. The Natural Killer Cell (NKC) immune gene cluster, which consists of a large number of cell-surface receptor encoding genes that are expressed in NKC lymphocytes, is present on a single contig in the current ARS1 version of the goat reference genome. This represents a 53-fold improvement in contiguity over the prior goat reference genome version, and it confirms a previously discovered duplication of the locus that is specific to ruminant species. The full characterization of the structure of these regions has revealed a large gap in current marker-assisted selection efforts, as many immune gene regions lack even a single internal SNP marker in current genotyping tools. In the Caprine 50k SNP chip, the goat *IGHV* region has no internal SNP markers, whereas the *IGKV*, *IGLV* and *NKC* regions have 3, 5 and 3 internal markers respectively. With new, highly contiguous assemblies of these regions, additional SNP markers can be selected to track unique alleles for downstream genomic selection efforts.

W155: Cattle/Sheep/Goat 2

Oxford Nanopore and PacBio Shotgun Sequencing and Analysis of the Cattle Rumen Metagenome

Sergey Koren[1], Derek M. Bickhart[2], Mick Watson[3], Timothy P.L. Smith[4] and Adam M. Phillippy[1], (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)Dairy Forage Research Center, USDA-ARS, Madison, WI, (3)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (4)USDA, ARS, USMARC, Clay Center, NE
Metagenomics, the reconstruction of a population of genomes from an environment, has emerged as an important tool for investigating community structure and function. Despite recent advances in assembly algorithms, metagenomics assemblies are often fragmented and error-prone. Assembly of isolate bacterial genomes has undergone a resurgence in quality with long-read sequencing. However, to date, long-read sequencing has not been widely used to characterize metagenomics datasets. Using the cattle rumen microbial community as a test environment, we used long-read sequencing from both technologies in combination with chromatin conformation capture (Hi-C) to assemble and assign contigs to individual genomes. Long-read sequencing can generate highly contiguous assemblies, in some cases reconstructing complete genomes and chromosomes.

W156: Cattle/Sheep/Goat 2

Sequencing the Australian Brahman to Unravel the Genetic History of the Population.

Stephen Moore[1], Ben J. Hayes[2], Elizabeth Ross[1] and Ross Koufariotis[1], (1)University of Queensland, Brisbane, Australia, (2)The University of Queensland, Brisbane, Australia
Bos indicus cattle are adapted to harsh tropical environments. The *Bos indicus* breed widely used in beef production in Northern Australia are Brahman, developed in the southern USA through cross-breeding 4 types of Zebu (Bos indicus) cattle breeds, Ongole, Guzerat, Gir and Krishna cattle. Australian cattle were “graded up” twice to increase numbers once in USA and then in Australia using Brahman bulls and *Bos taurus* cows with the resulting calves then back crossed to Brahman bulls. The grading up process has led to an introgression of 7-10% *Bos taurus* genome into modern Brahman cattle. Identifying variation in Brahman genomes associated with adaptation, fertility, meat quality and growth rates would facilitate genome selection and therefore accelerate genetic gain for these traits, in both Brahman cattle and composite cattle with Brahman ancestry. With this ultimate aim, 50 Brahman cattle that were key ancestors of the breed were whole genome sequenced. Regions of *Bos taurus* introgression and regions favouring *Bos indicus* alleles were identified.

W157: Cattle/Sheep/Goat 2

Holstein Reference Genome: Long Read Sequencing, *de novo* Assembly and Annotation

Gonzalo Rincon, Zoetis, Kalamazoo, MI

Zoetis has a substantial portfolio devoted to the health and wellness of Holstein dairy cattle; therefore we generated a complete Holstein annotated reference genome to better understand the genetic basis of dairy cattle phenotypes. Semen samples were obtained from a Holstein bull and high molecular weight DNA extracted. This DNA was sequenced using multiple approaches to support a robust genome assembly: 174 SMRT cells of PacBio RSII; Dovetail Chicago libraries; 2kb Illumina Nextera matepair libraries sequenced 2x75 on a NextSeq500; 300 and 500bp Illumina paired end libraries sequenced 2x150 on a NextSeq500. PacBio reads were assembled using the PacBio FALCON assembler and polished using Quiver. Scaffolding was performed using Dovetail data, combined with *Bos taurus* linkage map and Hereford optical map BtOM1.0. The final assembly constitutes 30 scaffolds (BTA1-29, X), 271 contigs assigned to the Y chromosome, and 2,958 unplaced contigs. After addition of the mitochondrion, the complete genome size is 2,772,068,867 bp. The scaffold N50 is 103.87 Mb with an L50 scaffold count of 11. This genome is 94.2% complete when assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO. A comparative alignment of the ARS-UCD1.1, and this assembly, while generally parsimonious, identified numerous rearrangements. Genome annotation was performed using Iso-seq and RNA-seq sequence reads from multiple tissues to capture all gene isoforms. This is the first Holstein assembly derived from long read data, and will provide a useful tool for understanding economically relevant traits in dairy cattle.

W158: Cattle/Sheep/Goat 2
Cattle CNVs - the Long and Short (sequence) of It All
Christine Couldrey, LIC, Hamilton, New Zealand
Copy number variants (CNVs) have eluded easy detection and characterisation, particularly in non-human species. However, there is increasing evidence that CNVs not only contribute a substantial proportion of genetic variation but have significant influence on phenotypes. Here we present discovery of CNVs in seven New Zealand dairy bulls using synthetic long read sequencing technology and one sequenced using PacBio technology. Validation of CNVs was undertaken utilising whole genome Illumina sequencing of 557 cattle representing the wider New Zealand dairy cattle population. Comparison of CNV identification and validation using the synthetic and true long reads showed that use of synthetic long reads allows discovery of CNVs not identified in previously reported true long read sequencing but is unable to detect mid-sized duplications. Ultimately no single sequencing is able to identify all CNVs in a population, and a combinatorial approach to generate comprehensive lists of CNVs is required. A comparison of CNV detection using the UMB3.1 and new (ARS) bovine reference genome will be presented to highlight the improvement when using a more complete and accurate genome assembly.

W159: Cattle/Sheep/Goat 2
Genetic Signatures of Caribbean Hair Sheep
Gordon Spangler, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD
Hair sheep of Caribbean origin have become an important part of the U.S. sheep industry. Lack of wool eliminates a number of health concerns and drastically reduces the cost of production. More importantly, Caribbean hair sheep demonstrate robust performance even in the presence of drug resistant gastrointestinal nematodes. Despite the growing economic importance of hair sheep in the Americas their genetic origins have remained speculative. Prior to this report no genetic studies were able to identify a unique geographical origin of hair sheep in the New World. Our study clarifies the African and European ancestry of Caribbean hair sheep. Whole genome structural analysis was conducted on four established breeds of hair sheep from the Caribbean region. Using breeds representing Africa and Europe we establish an objective measure indicating Caribbean hair sheep are derived from Iberian and West African origins. Caribbean hair sheep result from West African introgression into established ecotypes of Iberian descent. Genotypes from 47,750 autosomal single nucleotide polymorphism markers scored in 290 animals were used to characterize the population structure of the St Croix, Barbados Blackbelly, Morada Nova, and Santa Ines. Principal components, admixture, and phylogenetic analyses results correlate with historical patterns of colonization and trade, and support co-migration of these sheep with humans. Subsequent HD haplotype association and IBS analyses based on these findings reveal QTLs relating to hair coat, thermal tolerance, pelt quality, and resistance to the effects of nematode infection.

W160: Cattle/Sheep/Goat 2
Diversity of Copy Number Variation in the Worldwide Goat Population
Mei Liu[1,2], Yang Zhou[3], Benjamin D. Rosen[1], Curtis P. VanTassell[1], Alessandra Stella[4], Gwenola Tosser-Klopp[5], Rachel Rupp[5], Isabelle Palhière[5], Licia Colli[6], Brian L. Sayre[7], Paola Crepaldi[8], Gábor Mészáros[9], Hong Chen[2] and George E. Liu[1], (1)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (2)College of Animal Science and Technology, Northwest A&F University, Yangling, China, (3)Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Education Ministry of China, Huazhong Agricultural University, Wuhan, China, (4)PTP Science Park, Lodi, Italy, (5)INRA-GenPhySE, Castanet Tolosan Cedex, France, (6)Inst. of Zootechnics, Università Cattolica del S. Cuore, Piacenza, Italy, (7)Virginia State University, Petersburg, VA, (8)Department of Veterinary Science and Public Health - University of Milan, Milan, Italy, (9)Division of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria
Goats (*Capra hircus*) represent one of the most important farm animal species. Copy number variation (CNV) is a major source of genomic structural variation. The object of this research is to investigate the diversity of CNV distribution in goats using CaprineSNP50 genotyping data generated by the ADAPTmap Project. We identified 6286 putative CNVs in 1023 samples from 50 goat breeds using PennCNV. These CNVs were merged into 978 CNV regions, spanning ~262 Mb of total length and corresponding to ~8.96% of the goat genome. We then divided the samples into six subgroups per geographic distribution and conducted a comparison. Our results revealed a population differentiation in CNV across different geographical areas, including Western Asia, Eastern Mediterranean, Alpine & Northern Europe, Madagascar, Northwestern Africa and Southeastern Africa groups. The results of a cluster heatmap analysis based on the CNV frequency across different groups was generally consistent with the one generated from the SNP data, likely reflecting the population history of different goat breeds. We observed that the *DGAT1*, a metabolic processes related gene, was harbored in a common and lineage-differential CNV regions in goats for the first time. We also found some other important genes (e.g. *EDNRA*, *ADAMTS2*, *CHRNB1*, *CLCN7*, and *EXOSC4*) overlapping with CNVs, which are involved in local adaptations such as coat color, osteopetrosis and embryonic development. Our study generated an extensive CNV map in the worldwide goat population of goat, which offers novel insight into the goat genome and its functional annotation.

W161: Cattle/Sheep/Goat 2

Examining Conserved DNA Methylation in the Bovine 5' AMPK Gene Family

Fernando Betancourt, Darla M Quijada, Sydney Friedman, Sarah Perlee, Hannah Lachance and **Stephanie McKay**, University of Vermont, Burlington, VT

The AMPK gene family is responsible for cellular metabolism and energy regulation. Activation of these genes is caused by stress induced environmental and nutritional factors. The primary function of these genes is to provide energy-conserving measures when ATP levels are being depleted. Epigenetic modifications of DNA methylation have been associated with the regulation of this gene family. DNA methylation involves the addition of a methyl group onto the 5 carbon position of cytosines within CpG dinucleotides. The presence of methylation in the promoter region of a gene can prevent transcription and thus alter phenotypic effects. Of the seven genes that comprise the AMPK gene family, five genes were selected for analysis. Primers pairs for each of these five genes were generated with MethPrimer and used to amplify DNA extracted from the liver and muscle tissues of two breeds of cattle, Angus (n=6) and Charolais (n=6). The presence of DNA methylation was initially determined by combined bisulfite restriction analysis and positive results were confirmed with Sanger sequencing. DNA methylation was found to be conserved across breed and tissue within genic regions of two genes comprising the AMPK gene family. Further insights into the function and evolution of DNA methylation will incorporate a genome-wide approach towards identifying conservation of DNA Methylation within and across cattle breeds and tissues.

W162: Cattle/Sheep/Goat 2

Genome-Wide Sequencing and Comparative Profiling of Cattle Sperm DNA Methylome Reveals its Hypomethylated Patterns

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

Compared to somatic cells, sperm cells undergo nearly complete reprogramming of DNA methylation and exchange histones by protamine. Although sperm DNA methylation has been studied in humans and other model species, as the most beneficiary species of artificial insemination, its status in cattle is largely unknown. Using whole-genome bisulfite sequencing (WGBS), we profiled the DNA methylome of cattle sperm through comparison with somatic cells from three bovine tissues (mammary gland, brain, and blood). Large differences between cattle sperm and somatic cells were observed in the methylation patterns of global CpGs, pericentromeric satellites, partially methylated domains (PMDs), hypomethylated regions (HMRs), and common repeats. As expected and like other species, we observed low methylation in the promoter regions and high methylation in the bodies of active genes. We detected selective hypomethylation of megabase domains of centromeric satellite clusters, which may be related to chromosome segregation during meiosis and their rapid transcriptional activation upon fertilization. We found more PMDs in sperm cells than in the somatic cells and identified meiosis-related genes like *KIF2B* and *REPIN1*, which are hypomethylated in sperm but hypermethylated in somatic cells. Besides the common HMRs around gene promoters which showed substantial differences between sperm and somatic cells, the sperm-specific HMRs also targeted to distinct spermatogenesis-related genes, including *BOLL*, *MAEL*, *ASZ1*, *SYCP3*, *CTCFL*, *MND1*, *SPATA22*, *PLD6*, *DDX4*, *RBBP8*, *FKBP6*, and *SYCE1*. Although common repeats were heavily methylated in both sperm and somatic cells, some hypomethylated repeats were enriched in gene promoters with large variations among tissues. For example, some young Bov-A2 repeats, which belong to the SINE family, were hypomethylated and could affect the promoter structures by introducing new regulatory elements. Our study provides a comprehensive resource for bovine sperm epigenomic research and enables new discoveries about DNA methylation and its role in male fertility.

W163: Cattle/Sheep/Goat 2

Genotyping without Breaking the Bank: Using Low-Depth Genotyping-by-Sequencing for Genomic Analyses in Ruminants

Andrew S Hess and Shannon Clarke, AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

Genotyping-by-sequencing (GBS) obtains medium to high density genotypes without the need to develop commercial assays or to have a sequenced genome. Therefore, GBS obtains higher density genotypes at a reduced cost compared to standard SNP genotype panels. These genotypes offer new opportunities to harness genomic information ranging from parentage to genome-wide association studies (GWAS) and genomic prediction. Because GBS data does not suffer from the same ascertainment bias as SNP chips, it provides a higher chance of obtaining a causal mutation in the SNP set. To reduce cost and increase throughput, our target sequencing depth (i.e. the expected number of reads for an individual at any locus) is between 2x and 4x, which leads to some level of uncertainty in SNP genotype calling. We have focused on developing methods for genomic analyses that account for this uncertainty in GBS genotype calls and applied these methods to a range of ruminant species. GBS data has been compared to SNP chip data for genomic selection (sheep), GWAS (goats and deer), and using closely related species in the same genetic evaluations (deer). We have shown that GBS data produces similar relationship matrices as SNP chip data; and is able to identify known and strong candidate QTL in a GWAS. GBS data was successfully used to generate genomic relationship matrices combining two species of deer using species-specific allele frequencies. These results show the utility and flexibility of GBS as a low-cost tool for genomic analyses.

W164: Cattle/Sheep/Goat 2

Update on Bovine Genome Annotation

Christine G. Elsik[1], Deepak R. Unni[1], Darren Hagen[2] and Timothy P.L. Smith[3], (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Oklahoma State University, Stillwater, OK, (3)USDA, ARS, USMARC, Clay Center, NE

We are seeking bovine research community members who wish to provide input to manual annotation of the forthcoming improved bovine genome assembly. The Bovine Genome Database (http://BovineGenome.org) provides Apollo2.0, a web-based gene annotation system that allows users to view and edit gene annotations. Changes to genes are updated in real time so that they are immediately visible in the browser, allowing for an unprecedented level of collaboration. Bovine gene evidence tracks will include automated gene predictions, genes transferred from the UMD3.1 assembly using UCSC liftOver, and alignments of RNA-seq reads and full-length transcript sequences. New data available

for genome annotation includes paired-end stranded RNA-seq data from 24 tissues and full-length Iso-seq transcripts from 22 tissues of Dominette, the individual used to generate the Hereford reference genome assembly. We will post announcements on AnGenMap about annotation training webinars. The planned community bovine genome annotation project will provide a great opportunity for students to participate in an international collaboration.

W165: Cattle/Swine
Genetic Architecture of Complex Traits within and across Cattle Breeds: Implication for Genomic Prediction from Sequence Data
Ben J. Hayes, The University of Queensland, Brisbane, Australia
The advantage of using whole genome sequence data for genomic prediction in livestock, over and above the accuracies of genomic prediction that can be achieved with SNP arrays, is dependent on the number of loci affecting the target complex traits and the distribution of their effects. If genomic predictions are made across breeds, then the accuracy of these predictions also depends on the number of loci co-segregating in the breeds, and the joint distribution of their effects. Here we use large data sets of cattle with imputed whole genome sequence data to investigate the number of loci affecting traits of different complexity, ranging from gene expression, a relatively simple trait, to fatty acid profiles, to stature. We demonstrate that even for very complex traits, such as stature, there is still an advantage of using sequence data, particularly if predictions are made across breeds.

W166: Cattle/Swine
Pig FAANG Pilot Project
Pablo J. Ross, Animal Science, University of California, Davis, CA

W167: Cattle/Swine
Genomic Resources for Cattle and Swine at NCBI
Terence D. Murphy[1], Francoise Thibaud-Nissen[1] and NCBI's Eukaryotic Genome Annotation Team, (1)National Center for Biotechnology Information (NCBI), Bethesda, MD
The National Center for Biotechnology Information (NCBI) provides a diverse set of genomic resources for a wide variety of organisms, with a focus on organisms of medical or agricultural relevance. NCBI's assembly resource provides systematic access to over 100,000 genome assemblies with stable sequence content to help promote future data exchange and compatibility. NCBI further provides robust automated gene predictions supplemented with manual expert curation for a subset of genomes, including pig and cattle, as part of NCBI's RefSeq project. The latest pig assembly, Sscrofa11.1 from the Swine Genome Sequencing Consortium, provides the foundation for a substantially improved RefSeq genome annotation that integrates diverse set of short and long-read RNA-seq data available in the public archives. Annotations from NCBI's pipeline for pig, cattle, sheep, goat, and over 400 other eukaryotes are available in NCBI's Gene resource, BLAST databases, and Genome Data Viewer (GDV). Gene and GDV also provide access to other genomic information including RNA-seq expression data, and whole genome alignments to previous assembly versions or assemblies from different breeds, strains, or subspecies. Further information about NCBI's annotation resources and GDV is available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/ and https://www.ncbi.nlm.nih.gov/genome/gdv/.

W168: Cattle/Swine
Can We Apply Lessons Learned from Manual Gene Annotation in Human and Mouse to Livestock?
Adam Frankish, EMBL-EBI, Hinxton, United Kingdom
The Ensembl-HAVANA team have significant expertise in manual genome annotation and over the last 15 years have been providing reference gene annotation for whole genomes (human, mouse and zebrafish), individual chromosomes (Pig chr X and Y), genes (Rat, Pig) and regions (MHC of Gorilla, Pig, Dog, Wallaby, Tasmanian devil) of community interest. Comprehensive manual annotation of high quality genomes is labour intensive and as such is not practical for very many genomes, however, automated gene annotation methods such as the Ensembl genebuild pipeline, can do a good job of a capturing the geneset, particularly protein coding genes. It is clear that experts in individual communities will want to improve the baseline automated annotation, for example to adequately capture their knowledge of functionally important genes or resolve annotation errors in complex regions such as gene clusters that present particular challenges for automated pipelines. We have a history of successful annotation workshops that have been co-ordinated by our team, namely for cow, pig and rat, where we provided training and annotation expertise to particular communities. As individual groups and communities create their own gene annotation, there is a danger that any divergence in their approach could hinder accurate downstream analysis both within and between species. For example, the CCDS collaboration between ourselves and RefSeq was established to agree common annotation for at least one CDS in every protein-coding gene in the already well annotated human and mouse genomes. Despite the technical expertise in both groups and the wealth of available experimental data in these species, small differences in starting annotation guidelines led to significant differences in the annotated genes, requiring the resolution of many hundreds of annotation differences. We will present our guidelines and practices for annotation, based on our accumulated knowledge from producing reference gene annotation as framework that could be used to inform the approach of the livestock community towards manual annotation, for example, by providing guidelines that can be used in a platform agnostic way to help inform decisions on annotating structural and functional information for genes and transcripts.

W169: Cattle/Swine
Third Time's a Charm: Into "the Genomics Era" Once More
Timothy P.L. Smith, USDA, ARS, USMARC, Clay Center, NE
The first references to entering "the genomics era" were as early as the mid-1990s, referring (in livestock species) to the development of genetic maps for use in detection of QTL. The development of capillary-based parallel sequencing technology supported the release of the first draft of the human genome in 2000, and very expensive draft genome assemblies for cattle and chickens by 2008. These achievements spawned

a second wave of reports on entering “the genomics era”, which soon morphed into discussions of whether we had entered “the post-genome era”, referring to the development of genome-guided selection products and methods. The proclamations and discussions of being in a “post-genome era” had effect on the expectations and funding prospects of continued genome work. Fortunately, technological developments reduced the price of genome science so efforts could continue, albeit largely on an ad hoc basis. At present, we have a third wave of “the genomics era” emerging, supported by technologies that provide improved genome assembly and annotation. Will this be the final time we enter the genomics era? A retrospective and prospective discussion will be presented that suggests that we are, in truth, on the verge of entering a genome era. It is time to look towards the future on how to harness genome technology in all its branches, to advance agriculture.

W170: Cattle/Swine

A Natural Disease Challenge Model to Investigate Swine Disease Resilience

Graham S. Plastow, University of Alberta, Edmonton, AB, Canada

Disease resilience is distinct from disease resistance and can be defined as the ability of pigs to recover and continue to perform despite infection challenges. Genetic variation in susceptibility has been identified for several different pig diseases. In some cases we also know the genes and even causative mutations involved in this variation. However, there has been little use of this information to date. There are several reasons for this state of play. A major one is the lack of information available from which to determine the overall value of selection for reduced susceptibility. This is in part because any value may only materialize when the disease is present. A further complication is that in many cases the impact of disease is due to a number of pathogens rather than a single agent. For example, porcine respiratory disease complex can include several viruses such as porcine reproductive and respiratory syndrome virus, and swine influenza, as well as bacterial agents such as *Mycoplasma hyopneumoniae,* and *Actinobacillus pleuropneumoniae*. The final outcome may also be determined by the presence of opportunistic agents such as *Haemophilus parasuis,* and *Streptococcus suis*. In order to try to address these issues we set out to develop a challenge model that could be used to investigate factors impacting the resilience of pigs. The model which involves continuous flow of naïve pigs was established with “seeder” pigs in the first few batches entering the test station. Details of the approach and the first results will be described.

W171: Challenges and Opportunities in Plant Science Data Management - an International Workshop

Fshaw Abstract Placeholder

Felix Shaw, Earlham Institute, Norwich, United Kingdom

Replace with "real" abstract

W172: Challenges and Opportunities in Plant Science Data Management - an International Workshop

Improving the FAIRness of INRA’s Data for Plant Biology and Breeding

Michael Alaux[1], Cyril Pommier[1], Esther Dzale-Yeumo[2], Sophie Durand[1], Raphael Flores[1], Erik Kimmel[1], Thomas Letellier[1], Célia Michotey[1], Nacer Mohellibi[1], Hadi Quesneville[1] and **Anne-Francoise Adam-Blondon**[1], (1)URGI, INRA, Université Paris-Saclay, Versailles, France, (2)Department of Scientific Information (DIST), INRA, Université Paris-Saclay, VERSAILLES, France

INRA is involved in several projects (e.g. EU H2020 ELIXIR-Excelerate, n°676559) or global initiatives (e.g. Wheat Initiative, Research Data Alliance) contributing to the development of : (i) community recommandations for data standardisation (e.g. wheatis.org ; doi:10.1038/hortres.2016.56), (ii) data standards for phenotyping data (www.miappe.org), (iii) crop specific ontologies in the frame of the CropOntology (http://www.cropontology.org/) and (iv) standard web services (www.brapi.org). These global resources are used to capture the data produced by INRA and its partners in large scientific projects with a standard and structured vocabulary and to store them into INRA’s central repository for plant genomic, phenomic and genetic data, GnpIS (https://urgi.versailles.inra.fr/gnpis/) under the FAIR principles (https://www.force11.org/group/fairgroup/fairprinciples). For this purpose, standardization good practices are actively promoted in particular in the french community using as levers large french scientific projects centered around crop species (Wheat, Maize, Rapeseed, SunFlower, Pea and Sugar Beet) or infrastructures such as the french node of the European infrastructure for phenotyping (EMPHASIS) or the french infrastructure for biological resources for research in agriculture (AgroBRC). Recently, the standards for phenotyping data have been extended to support forest tree data in collaboration with the french node of the European infrastructure for Analysis and Experimentation on Ecosystems (AnaEE Services). Finally, these progresses in the FAIRness of our data are used to develop or contribute to federations of interoperable information systems (see for instance the Wheat community use case: doi: 10.12688/f1000research.12234.1).

W173: Challenges and Opportunities in Plant Science Data Management - an International Workshop

Elyons Place Holder

Eric Lyons, University of Arizona; BIO5 Institute; CyVerse, Tucson, AZ

CoGe was first publicly launched 9 years ago and contained four genomes from two species – two version of *Arabidopsis thaliana* and two version of *Oryza sativa* with a few tools for analyzing and comparing genomes. Today it has grown to manage over 33,000 genomes from nearly 18,000 organisms. Its data management system stores over 1.5 billion structural annotations, 2.5 billion feature names, and over 3.1 billion genomic locations. Recently, CoGe has deployed several data processing pipelines to let users easily use fastq to do transcriptomics, identify variants, quantify epigenomic marks, and a variety of other functional and diversity data. These pipelines automatically add these data to CoGe for more detailed functional analyses of genomes. Of particular note, CoGe’s development team has always been small, averaging one full time employee (max of two). To achieve this growth and sustainability, several design decisions and choices were made during the development of CoGe. This talk will discuss those choices and the lessons learn in developing and maintaining one of the world’s most popular and open platforms for comparative genomics.

W174: Challenges and Opportunities in Plant Science Data Management - an International Workshop

FAIR Plant Data Management Challenges

Richard Finkers, Wageningen UR Plant Breeding, Wageningen, Netherlands
Efficient utilisation of Genetic diversity within Plant Breeding is one of the elements in the ambition to produce twice as much food with twice less input. Technological advances currently drive initiatives to discover genetic and phenotypic diversity in many plant crop species. However, accessibility of this data is difficult, especially for computers, limiting its potential for re-use. Data standards such as the minimal information about a plant phenotype experiment (MIAPPI) and exchange mechanisms such as the plant breeding API (BrAPI) emerging to solve part of the challenge, there are steps to be made to push these technological developments to be compliant with the FAIR data principles. Possibilities and challenges in describing crop passport data, utilizing the multi crop passport descriptor standard, according to a FAIR semantic data model will be discussed.

W175: Challenges and Opportunities in Plant Science Data Management - an International Workshop
Community Based Challenges for Improving Genomic Annotation
Iddo Friedberg, Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA
A biological experiment is the most reliable way of assigning function to a protein. However, in the era of high-throughput sequencing, scientists are unable to carry out experiments to determine the function of every single gene product. Therefore, to gain insights into the activity of these molecules and guide experiments, we must rely on computational means to functionally annotate the majority of sequence data. To understand how well these algorithms perform, we have established a challenge involving a broad scientific community in which we evaluate different annotation methods according to their ability to predict the associations between previously unannotated protein sequences and Gene Ontology terms. I will discuss the rationale, benefits and issues associated with evaluating computational methods in an ongoing community-wide challenge.

W176: Challenges and Opportunities in Plant Science Data Management - an International Workshop
How the Maize Genome Database Helps Guide Data Management Best Practices in the Maize Community
Margaret Woodhouse, ISU, Ames, IA
The Maize Genome Database (MaizeGDB) is one of the primary repositories for genetic and genomic data for maize researchers. We act as a clearinghouse for nomenclature and metadata guidelines, and have a close relationship with the members of the maize community. Because of this, we are able to help guide best practices for maize genetic and genomic metadata, provenance, and data quality control. This talk outlines our methods and pipelines for ensuring proper data management among the maize community, including outreach, metadata standards, and GenBank submission requirements.

W178: Citrus Genome
Functional Genomics of Citrus Pathogens Focusing on Effectors Biology
Marcos Antonio Machado[1], Ronaldo JD Dalio Dalio[2], Eduardo Henrique Goulin[2], Carolina Munari Rodrigues[2], Paulo José Camargo dos Santos[2] and Marco Aurélio Takita[3], (1)Centro de Citricultura/IAC, Cordeiropolis, Brazil, (2)Centro de Citricultura/IAC, Cordeirópolis - SP, Brazil, (3)Centro de Citricultura/IAC, Cordeiropolis - SP, Brazil
Diseases caused by fungi and oomycetes causes serious problems in the production of citrus worldwide. Pathogens that affect mainly the rootstock, such as root rot and gummosis caused by Phytophthora spp, or the production or quality of the fruits as post-bloom fruit drop (PFD) caused by Colletotrichum abscissum, black spot (Phyllosticta citricarpa), and brown spot (Alternaria alternata) represent a significant increase in production costs. The molecular mechanisms of interactions of these pathogens with citrus are still unknown. With the sequencing of their genomes, an approach to understanding these interactions is the search for candidate genes to act as effectors, i.e., to produce proteins that are secreted out of the cell and capable of altering host structure and physiology. Since they are fundamental virulence factors of the pathogens, management strategies for disease control, specially to be used as probes to screen for genome and transcriptome data from these pathogens, a bioinformatic pipeline was developed for selection of candidate effectors genes based on criterious, such as presence of signal peptides for secretion, presence or absent of transmembrane alfa-helices, secondary and tertiary structure. All candidate genes are re-sequenced and some of them have been functionally evaluated in model plants for induction of hypersensitivity responde (HR). The most promising are being used in the transformation of Arabidopsis to confirm or not their ability to alter host physiology. Financial support: INCT Citros (Fapesp 2014/50880-0 and CNPq 465440/2014-2).

W179: Citrus Genome
Identification of *Citrus* Species and the Genetic Heterogeneity of Mandarins
Guohong Albert Wu, DOE Joint Genome Institute, Walnut Creek, CA
Whole genome shotgun sequencing has made it possible to revisit citrus phylogeny, genealogy and taxonomy and to provide an unambiguous operational definition of the genetic species of the genus Citrus. Our recent re-sequencing effort revealed over ten ancestral citrus species and subspecies, and showed Poncirus in a separate genus than Citrus. Pummelo introgression is found to be widespread among mandarins, and the admixture pattern suggests two phases of pummelo introgression (early and late). This resulted in the high genetic heterogeneity of mandarins that are absent in other citrus taxa. We also observe an extensive relatedness network among the mandarins and suggest that this is a direct consequence of the domestication process.

W180: Citrus Genome
Resolving the Diploid Genome of Fairchild Mandarin with Single Molecule Sequencing Technologies
Jinfeng Chen, Ruidong Li, Sergio Pietro Ferrante, Travis R. Wrightsman, Zhenyu Jia, James Burnette, Mikeal L. Roose, Susan Wessler and Jason Stajich, University of California, Riverside, CA
The citrus genomes are highly heterozygous, while the haplotype information in the heterozygous genome is essential to fully understand the genome structure and genetic diversity of citrus. Yet, the published reference genomes of citrus, including mandarin, sweet orange, and pummelo, were sequenced and assembled from haploid varieties. The haplotype structure of the citrus genomes is largely unexplored. Here, we

assembled the Fairchild mandarin genome from a diploid heterozygous tree by using single molecule sequencing technologies (Pacific Biosciences, Bionano, and 10x Genomics). We have built a reference quality genome with these advanced sequencing technologies directly from a diploid, saving the effort of producing a haploid variety. Our reference quality assembled Fairchild genome was 366 Mb (contig N50=10Mb) in size as compared to assembly sizes of 300-325 Mb (contig N50=50-120kb) for other citrus varieties using only Illumina short reads or Sanger sequencing. The Fairchild assembly was scaffolded using Bionano genome map, 10x Genomics linked reads, and genetic linkage map to achieve a nine chromosome assembly. We further phased the heterozygous SNPs using both 10x Genomics linked reads and a pollen-genotyping method, which generated chromosomal-level haplotype blocks. These phased heterozygous SNPs were used to facilitate the assembly of haplotypes and resolving the structural differences between haplotypes. The strategy developed here is not only useful for studying citrus genomes but also valuable for other tree species that have a high heterozygous genome.

W182: Citrus Genome

A Practical Genetic Mapping Strategy for Citrus Polyembryony using 'Omics Technology

Xia Wang[1], Yuantao Xu[1], Siqi Zhang[1], Li Cao[2], Xiuxin Deng[1] and Qiang Xu[1], (1)Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China, (2)Southwest University, Chonqing, China

In citrus, nucellar polyembryony is a unique apomixis phenomenon, in which the embryos develop from somatic nucellar cells. Polyembryony is widely employed in citrus nurseries and propagation programs to generate large numbers of uniform rootstocks from seeds and to permanently fix valuable traits and hybrid vigor, while it has also caused problems for breeding that required sexual crosses, which make it of great value to study the citrus polyembryony. Because of long juvenile phase of woody citrus plants and citrus polyembryony, conventional genetic mapping methods by constructing hybrid population after several generations are not applicable for citrus. With the rapid development of sequencing methods, a new strategy combing omics technology with simple genetic population construction was designed for genetic mapping in citrus. For the loci mapping for polyembryony in citrus, bulk segregant analysis was conducted and a region of 1.96Mb was mapped. To fine map the polyembryony locus, a local association analysis within the 1.96-Mb region among natural population, which consisted of 108 citrus accessions individually sequenced with an average depth of 30, was performed. The final genetic locus responsible for citrus polyembryony was narrowed to an 80-kb region containing 11 genes. Among this region, a codominant InDel and MITE insertion showed cosegregation with polyembryony and were developed as molecular makers for identification of polyembryony in citrus. This study provides new insights into citrus polyembryony and provides practicable references for genetic mapping of agriculturally important genes in citrus.

W183: Citrus Genome

A New Evolutionary Framework for the Genus *Citrus*: Its Origin, Evolution and Dispersal

G. Albert Wu[1], Javier Terol[2], Victoria Ibañez[3], Antonio López-García[4,5], Estela Pérez-Román[5], Carles Borredá[5], Concha Domingo Carrasco[5], Francisco R Tadeo[5], Jose Carbonell[6], Roberto Alonso[6], Franck Curk[7], Dongliang Du[8], Patrick Ollitrault[9], Mikeal L. Roose[10], Joaquin Dopazo[6], Fred G. Gmitter[8], Daniel S. Rokhsar[1] and **Manuel Talon**[5], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)Centro de Genomica, Instituto Valenciano de Investigaciones Agrarias, IVIA, Moncada, Spain, (3)Centro de Genomica IVIA, Moncada, Spain, (4)Instituto Valenciano de Investigaciones Agrarias, IVIA, Moncada, Valencia, Spain, (5)Centro de Genomica, Instituto Valenciano de Investigaciones Agrarias, IVIA, Moncada, Valencia, Spain, (6)Institute for Genomcis and Computational Medicine, CIPF, Valencia, Spain, (7)UMR Agap Corse, San Giuliano, France, (8)University of Florida, IFAS-CREC, Lake Alfred, FL, (9)CIRAD, UMR AGAP, Montpellier, France, (10)University of California, Riverside, CA

We present first solid insights on the origin, evolution and dispersal of citrus and elucidate the genealogy of the most important wild and cultivated varieties. These findings draw a new evolutionary framework for these fruit crops, a scenario that challenges current taxonomic and phylogenetic thoughts and points towards a reformulation of the genus Citrus. Based on genomic, phylogenetic and biogeographical analyses of the genus Citrus we propose that the center of origin of citrus was the Southeast foothills of the Himalayas, in a region including the eastern area of Assam, northern Myanmar and western Yunnan. Our analyses suggest that the ancestral citrus species underwent a sudden speciation during late Miocene and that the new species dispersed from there to surrounding regions, coinciding with a drastic transition from wetter monsoonal conditions to a drier climate. The Australian citrus and Tachibana mandarin split later from mainland citrus during the early Pliocene and Pleistocene, respectively.

W184: Climate Change and ICRCGC 1

Towards Combating Climate Change for Food and Nutrition Security

Chittaranjan Kole, Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad, India; Department of Atomic Energy, Government. of India, Kalyani, India

International Climate Resilient Crop Genomics Consortium (ICRCGC) was founded in 2011 to formulate the concepts, strategies, tools and techniques to develop climate-smart crop varieties. Its missions include utilization of the available genomic resources, particularly in the wild crop relatives, and the advanced tools of genomics, specifically next-generation sequencing. ICRCGC organized eleven workshops during the PAG conference in 2012, 2013, 2014, 2015, 2016 and 2017; one workshop in collaboration with Bill and Melinda Gates Foundation during the PAG conference in 2012, and one conference in collaboration with International Society of Crop Science held at Brazil in 2012. During these eleven workshops, the deliberations were focused on the basic, 'translational' and participatory research as well as deliberations on future priorities. Meantime, a team of ICRCGC members published a review paper entitled 'Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects' (Front. Plant Sci. 6:563. doi: 10.3389/fpls.2015.00563) as an outcome of the white paper drafted by the ICRCGC members (http://www.icrcgc.org/white_paper.html). In parallel, a number of other platforms with crop or geographic focus have been constituted with similar goals as the ICRCGC. Improved synergy, efficiency and outcomes seem likely if the work of these separate groups can be coordinated. The three workshops of Climate Change and ICRCGC I, II and III during PAG 2018 will aim at forging

and fostering collaboration of the crop-wise, trait-wise and region-wise collaboration and interaction among academia, funding agencies and policy makers to architect a global platform.

W186: Climate Change and ICRCGC 1

Climate-Change Ready Rice for South America

Antonio Costa De Oliveira, Universidade Federal de Pelotas, Pelotas-RS, Brazil; Universidade Federal de Pelotas, Capão do Leão, Brazil and Vivian Ebeling Viana, Railson Schreinert dos Santos, Camila Pegoraro, Luciano Carlos da Maia

Food production in the XXI Cenctury has additional challenges associated to Climate changes. Among the major abiotic stresses in rice, submergence and iron excess tolerance play an important role in lowland cultivation systems. The gene machinery behind the tolerance to these stresses is complex and needs further investigation. Following our microarray studies, we have found several transcription factors (TFs) involved in iron responses. Transcriptomic studies regarding responses to salinity, iron and cold indicate unique and common genes regulated under different stresses. The recent advances in breeding strategies and constrains for rice production in South America are discussed.

W187: Climate Change and ICRCGC 1

Leveraging the Root Angle QTLome to Enhance Climate Resilience in Wheat

Marco Maccaferri, DipSA, Department of Agricultural Science, University of Bologna, Bologna, Italy, Danara Ormanbekova, Department of Agricultural Sciences, Bologna, Italy; DipSA - University of Bologna, Bologna, Italy, Ghasemali Nazemi, Department of Agriculture, , Haji abad, Iran (Islamic Republic of) and **Roberto Tuberosa**, Department of Agricultural Sciences, University of Bologna, Bologna, Italy

Optimisation of root system architecture (RSA) is an important objective for the sustainability of durum wheat grown under drought-stressed conditions. In the present study, linkage and association mapping (AM) for RSA evaluated at the seedling stage evidenced 20 clusters of quantitative trait loci (QTLs) for root length and number as well as 30 QTLs for root growth angle (RGA). The most divergent RGA phenotypes observed by seminal root screening were validated by root phenotyping of field-grown adult plants. QTL analysis of RSA and grain yield data indicates RGA as a valuable target to enhance grain yield and yield stability across different soil moisture regimes (Maccaferri et al. 2016). Based on their relative additive effects, allelic distribution in the AM panel and co-location with QTLs for yield, eight RGA QTLs have been prioritised in terms of breeding interest and value. These QTLs were investigated for gene content based on the chromosomal pseudomolecules of Chinese Spring *T. aestivum* and the TriAnnot v4.3 gene prediction and annotation pipeline and the Zavitan *T. dicoccoides* genome assembly (Avni et al. 2017). The chromosome regions contained 25 to 242 predicted genes (123 on average). In six RGA QTLs, from one to four gene annotations were involved in auxin pathways. The comparison between the *T. aestivum* and *T. dicoccoides* gene content indicates the high quality of the *T. dicoccoides* assembly and its usefulness to identify candidates to explore the polymorphism and the structural variation of drought-related genes present in the A and B wheat genomes.

W188: Climate Change and ICRCGC 1

Diversity of Responses of Wheat to Heat Stress as Revealed by Grain Transcriptome Profiling

Parimalan Rangan[1], Agnelo Furtado[2] and **Robert J. Henry**[2], (1)ICAR, New Delhi, India, (2)University of Queensland/QAAFI, Brisbane, Australia

Adapting major crops to climate change requires an understanding of the influence of temperature on plant performance. Wheat is a crop that often matures in a warming environment and is prone to heat stress especially late in the growth of the crop. We have examined the impact of heat stress at mid and late seed development using the transcriptome of the developing wheat grain. Genotypes displayed remarkable diversity in response at the transcript level and in the associated impact on grain size and yield. Genotypes also displayed differences in the timing of susceptibility to heat stress. Short periods of heat stress and longer periods of continuous high temperatures may require different genetic adaptation. Higher temperatures are likely to impact not only on wheat productivity but also on the composition of the wheat grain. This has implications for both the functional (especially end use quality) and nutritional quality of the grain. However, these studies suggest significant potential to select genotypes that are better adapted to heat stress and provide a better understanding of the genetic basis of heat stress tolerance.

W189: Climate Change and ICRCGC 1

Emerging Threats to Global Development: Combating Plant Disease in a Changing Climate

Angela Records, USAID Bureau for Food Security, Washington, DC

W191: Climate Change and ICRCGC 2

Population Genomic Analyses of Stress Adaptation in Wild Sunflowers

Loren Rieseberg, University of British Columbia, Vancouver, BC, Canada

Crop domestication is frequently accompanied by a reduction in resistance to biotic and abiotic stress. This decline in resistance represents a major impediment to global efforts to increase crop productivity, especially in the context of climate change and heightened competition for land and water. To reduce stress-induced yield losses, attention has increasingly turned to crop wild relatives, which often have evolved mechanisms to cope with environmental stress. Here I will discuss a series of projects underway to identify alleles underlying stress resistance in extremophile wild sunflower species and to move these alleles into cultivated germplasm. At the center of this effort is a genome wide association study involving whole genome sequencing data for circa 1800 genotypes representing 175 accessions of three wild species. Searches for associations between genotypic variation, phenotypic traits, and climate and soil characteristics are underway. We also have created pre-bred lines containing introgressions from the same wild species to evaluate the effects of candidate resistance alleles in a cultivated genetic background. Lastly, I will report on the development of modified Multiparent Advanced Generation Inter-Cross (MAGIC) mapping populations that include both cultivated and wild sunflower donors. These populations are expected to facilitate the efficient characterization of resistance traits in sunflower and to produce materials containing exotic alleles that can be readily deployed in breeding programs.

W192: Climate Change and ICRCGC 2
Potential Role of Underutilized Food Legumes in Facing the Challenge of Climate Change in West Africa
Michael Abberton, International Institute of Tropical Agriculture, Ibadan, Nigeria

W193: Climate Change and ICRCGC 2
Soybean Sudden Death Syndrome: A Harbinger of Doom?
David A. Lightfoot, Southern Illinois University, Carbondale, IL
Climates change and bring on new diseases. Novel tools to improve resistance to sudden death syndrome (SDS) and the underlying Fusarium root rot (FRR) caused by *Fusarium virguliforme* (Aoki) have been developed for soybean [*Glycine max* (L.) Merr.]. Mainly because this Rebel disease has marched North opposed by a Rebel General named cultivar. Between eighteen and thirty resistance loci have been identified due to clustering the real number is unclear. Many were confirmed over the past twenty one years (named *Rfs1* to *Rfs18*). To select the beneficial alleles of 8 to 10 loci per cross needed for optimal resistance is a difficult task for plant breeders. Crops contend with many Fusaria, a group with a wide host range and flexible hemibiotrophic lifestyle. Full resistance is absent. SDS is a combination of two diseases. FRR, includes rotted roots and toxin-restricted root development. Leaf scorch, supra-petiolar abscission, pod abortion and early plant maturity are consequences of many toxin to target interactions. Breeding for combined FRR and SDS resistance using a set of exciting new tools for pathogen quantification in roots. Resistance genes were proven, including GmRLK18-1 (Glyma_18_02680) *Rfs2* through trangenics, and also MIPs1a (EC 5.5.1.4; *Rfs3*) by mutation. NILs and SNPs were used to confirm two more genes on chromosome 18 (*Rfs* and *Rfs1*). Some candidate genes were identified. The new tools provide an opportunity for new breeding initiatives. Core discoveries from the past 21yrs aim to incorporate best practices from old and new initiatives to deal with global warming

W194: Climate Change and ICRCGC 2
Cowpea: Navigating the Genome of a Climate Resilient Legume
Timothy J. Close, Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA
Cowpea, *Vigna unguiculata* L. *Walp*, is a diploid warm-season legume (tribe Phaseoleae) with a genome size of ~620 Mb. Cowpea, known as blackeyed pea among other common names, is relevant as a grain legume in the USA and Europe, and as a fresh vegetable in China and elsewhere, but is of major importance as food and fodder in sub-Saharan Africa. An annotated reference genome sequence of an elite African variety, IT97K-499-35, is now available through Phytozome (www.phytozome.net). The v1.0 cowpea pseudomolecules contain 519 Mb of sequence, derived from superscaffold sequences with N50 = 16.4 Mb and L50 = 12. A total of 29,773 gene models were annotated using a combination of *ab initio* and transcript (RNA-Seq and Sanger EST) evidence, providing a measure of 95.9% plant completeness using BUSCO v2. Synteny between cowpea and other warm-season legumes has been clarified, including common bean (*Phaseolus vulgaris* L.), which provided the basis of new cowpea chromosome numbering. The genome assembly is based on single molecule real-time sequencing (91x coverage; Pacific Biosciences) together with two optical maps (BioNano Genomics) and ten genetic linkage maps containing a total of 44,003 SNPs. Transition to the use of the reference genome has improved the resolution of QTL mapping in biparental and multiparent populations, and using a minicore of diverse cowpea germplasm. An accumulating list of QTLs and candidate genes for traits now includes flowering time, pod shattering, seed coat patterns and texture, leaf shape, seed size, resistance to several pests and pathogens, yield in several environments, and others. Cowpea and its genome information constitute an important set of resources to understand the biology of this and related species, and to apply such knowledge to agricultural needs and germplasm conservation. This work was conducted mainly under the NSF BREAD project “Advancing the Cowpea Genome for Food Security” with partial support from the Feed the Future Innovation Lab for Climate Resilient Cowpea.

W195: Climate Change and ICRCGC 2
Dissection of Pea Responses to Water Stress during Seed Filling Identifies Candidate Genes for Drought Tolerance
Vanessa Vernoud[1], Nadia Rossin[1], Marion Prudent[1], Christine Le Signor[1], Myriam Sanchez[1], Sandrine Pateyron[2], Sandrine Balzergue[2], Catherine Rameau[3], Grégoire Aubert[1], Judith Burstin[1], Karine Gallardo[1] and Richard Thompson[1], (1)INRA UMR1347 Agroécologie, Dijon, France, (2)POPS transcriptomic plateform Paris-Saclay, Gif-sur-Yvette, France, (3)INRA Institut Jean-Pierre Bourgin, Versailles, France
Given their ability to fix atmospheric nitrogen, legumes are pivotal to the development of sustainable agriculture in Europe as a source of protein for food and feed. Pea (*Pisum sativum*) is currently the leading grain legume crop in France and major efforts are being made to reintroduce legumes as protein crops in Europe. However, instability of seed yield and quality due to environmental fluctuations still represent a real barrier for the development of these cultures, and breeding for stable yields is needed. In pea, drought stress occurring during the reproductive phase can greatly affect seed yield and quality. We investigated the response of pea plants (var. *Caméor*) subjected to water stress during the seed filling period, a phase associated with massive remobilization of nutrients from the vegetative organs (including leaves) to sustain the seed’s high-nitrogen demand. Transcriptomic profiling of leaf response to water stress by hybridization of a 40k pea micro-array revealed metabolic and regulatory pathways affected by drought and enabled the selection of candidate genes for drought resistance. One of these genes, named *RAMOSUS1*, encodes a carotenoid cleavage dioxygenase involved in strigolactone biosynthesis. Interestingly, phenotyping of the corresponding mutant showed it to have increased sensitivity to drought compared to the wild-type, suggesting that strigolactones could act as positive regulators for crop abiotic stress resistance, as already shown for Arabidopsis.

W196: Climate Change and ICRCGC 2
Translational Genomics for Developing Climate-Smart and High Yielding Pulse Crops
Rajeev Varshney, ICRISAT, Hyderabad, India
Pulses because of their higher nutrient contents play an important role in providing nutritional food security in developing countries. Average productivity of pulse crops like chickpea and pigeonpea is very low as these pulse crops are grown in semi-arid regions and exposed to a number of biotic and abiotic stresses. Breeding efforts for enhancing productivity could not meet the desired target that are critical to feed the

vastly growing global population especially in context of climate change. Genomics and molecular breeding, therefore, have huge potential to enhance the crop productivity and accelerate the rate of genetic gains. With an objective to understand the genome architecture, draft genomes of several pulses have been assembled. Large scale re-sequencing efforts have been initiated in these crops for identifying new sources of genetic variation and allelic variants of candidate gene(s) associated with beneficial traits which can be targetted for molecular breeding and genome editing. Molecular breeding approach has been used to improve tolerance to abiotic stress and resistance to biotic stresses. Several improved lines have shown higher yield as compared to the best checks and are expected to be released as improved varieties in due course. Such lines may be ready for coping drought, heat and biotic stresses and providing higher produce to farmers in changing climate.

W197: Climate Change and ICRCGC 1
Infusing Climate Resilience in High Yielding Crop Varieties by Genomics-Assisted Breeding
Nagendra K. Singh, ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

W198: Climate Change and ICRCGC 3
Limiting Greenhouse-Gas Emissions by Enhanced Breeding Progress for Nitrogen Use Efficiency – an Example from an N-Hungry Crop
Andreas Stahl, Mara Pfeifer, Paul Vollrath, Benjamin Wittkop and Rod Snowdon, Justus Liebig University, Giessen, Germany
Nitrogen (N) is a key driver of global production of agricultural commodities and essential to enhance food, feed and fuel production in order to match the demand of a world population expected to reach 10 billion people by 2050. However, conversion of atmospheric N_2 into reactive forms is associated with massive environmental impact. For example, unused nitrogen can escape from the agricultural production system by run-off, nitrate leaching or volatile NOx or ammonia. Furthermore, CO_2 emissions associated with mineral N fertilizer production cause additional greenhouse gas emissions. Together, more than 75% of greenhouse gas emissions are associated to N fertilization.
Oilseed rape is one of the most important oilseed crops in the world and of high relevance in crop rotations in Canada, Northern Europe, China and Australia. With its relatively high acquisition of N during vegetative growth stages, but a comparatively low N seed yield, oilseed rape cultivation is often associated with an N-balance surplus. In this study we analyzed the impact of breeding on N use efficiency over the last two decades. A total of 30 elite oilseed rape varieties, registered in Germany between 1989 and 2014, were tested under divergent N fertilization levels in multi-location field trials. The results revealed that enhancement of N use efficiency though intense selection for improved seed yield is a powerful lever for mitigation of greenhouse gas emissions. On average, recent oilseed rape varieties were found to require 13% less N fertilizer than older varieties to achieve the same oil production.

W199: Climate Change and ICRCGC 3
Development of Climate Resilient Soybeans
Henry T. Nguyen, University of Missouri, Columbia, MO

W200: Climate Change and ICRCGC 3
GplusE: A Large Scale Genomic Selection Experiment for Crop Improvement
Robert Jackson[1], Jaap Buntjer[2], Stefan Hoj-Edwards[2], R. Chris Gaynor[2], Alison R Bentley[1], Eric Ober[1], John Hickey[2] and Ian J Mackay[1], (1)NIAB, Cambridge, United Kingdom, (2)The Roslin Institute, Edinburgh, United Kingdom
Globally, and particularly within Europe, plant breeding will play an important role in preparing our crops for the effects of climate change. Especially when you consider increases in yield over the past 30 years have almost solely been down to genetic gains, with limited impact from environmental factors. One potential technique that could improve the breeding method is genomic selection (GS), a form of marker assisted selection where genetic markers are selected for the whole genome so all quantitive trait loci are in linkage disequilibrium with one gene. The GplusE project, collaboration between NIAB, Roslin and four commercial UK wheat breeders, aims to develop a GS strategy that enhances the accuracy of a GS wheat training population for a trait (e.g. yield) by utilising genetic relationships between traits whilst accounting for environmental effects and incorporating data from high-throughput field phenotyping. As part of this project 3,000 lines produced from elite UK winter wheat crosses will be grown over two years, at two sites, resulting in a total of 12,000 plots requiring assessment. Along with environmental analysis, remote sensing will be a key component of the high-throughput phenotyping of these lines. By utilising remote sensing on a breeding scale we will identify methods that will enhance the accuracy of selection. Coupled with the increased the benefits of GS this could provide wheat breeding with another tool to deal with the challenges of agriculture in a rapidly changing climate.

W201: Climate Change and ICRCGC 3
Strategies for Resolving and Data Mining of Crop GxE Interactions
Graham J. King[1,2], Liliana Andres[1,2], Kathryn Eales[1], Sadaf Naz[1], Sean Mayes[2] and Abdul Baten[1], (1)Southern Cross Plant Science, Southern Cross University, Lismore, Australia, (2)Crops For the Future, Semenyih, Malaysia
Cultivar development for crops better adapted to climate variability requires an understanding of the molecular and genetic basis of genotype by environment ('G x E') interactions. Low-cost re-sequencing and dense genotyping now focuses attention on **trait components**. Lack of **generic approaches** to data description, curation and interchange hinders progress for global crops, and is even more critical for regionally-adapted minor crops that will increasingly play a role in global nutritional security. Limited funding to achieve this is a tragedy, with research groups and international consortia often 're-inventing the data wheel'. More consistent, open and inter-operable platforms will drive demonstrable gains in trait resolution, especially where close taxo-evolutionary proximity can be exploited. Challenges remain in establishing a **common world view** and i**ncreasing communication** within the network of those responsible for generating, curating and interpreting data, including bioinformaticians, field and laboratory researchers, breeders and research managers.
We are currently integrating subsets of genetic, genomic and trait data for genetic resources, pre-breeding and breeding of Bambara groundnut, brassicas, coffee, hemp, macadamia, passionfruit and tea tree. This experience highlights the benefits of generic platforms and workflows from field and lab to web-enabled browser. Our FAIR (**F**indable_**A**ccessible_**I**nteroperable_**R**e-usable) compliant approach includes i. development

and agreement of **nomenclature standard**s; ii. establishment of **data entity registries**; iii, use of generic schema such as **CropStoreDB**; iv. formal approaches to **knowledge representation** via ontology inference workflows; and v. graphically-enabled interfaces to ease **data discovery** and **navigation**, especially for trait data.

W202: Climate Change and ICRCGC 3

Developing Climate Resilient Agriculture through the Development of Minor Crops - an Example from Bambara Groundnut

Sean Mayes, University of Nittingham, Leicestershire, United Kingdom

W203: Climate Change and ICRCGC 3

The Role of Alternative Crops in Addressing Climate Change

Bill Payne, University of Nevada-Reno, Reno, NV

W204: Coffee Genomics

Chromosome Scale Scaffolding of the High-Quality Genome Assemblies of the Allotetraploid *Coffea arabica* and Its Maternal Ancestor *C. eugenioides* and Validation Using Genetic and Physical Mapping Data

Aleksey Zimin[1,2], Carlos Ernesto Maldonado[3], Marcela Yepes[4], Keithanne Mockaitis[5,6], Pilar Moncada[3], Carrie Ganote[7], Sheri A. Sanders[5], Carmenza E. Góngora[3], Claudia Flórez[3], James A Yorke[1], Alvaro Gaitán[3] and Herb Aldwinckle[8], (1)University of Maryland, College Park, MD, (2)Johns Hopkins University, Department of Computer Science, Baltimore, MD, (3)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (4)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (5)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (6)Department of Biology, Indiana University, Bloomington, IN, (7)National Center for Genome Analysis Support, Pervasive Technology Institute/ Indiana University, Bloomington, IN, (8)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY

Cost effective strategies for sequencing and assembly of complex polyploid and highly heterozygous diploid genomes are a major need for plants, in particular because polyploidy frequently occurs in flowering plants (~70%) providing an important pathway for plant evolution and specialization. *De novo* assembly of such genomes remains a critical unsolved technical problem resulting in incomplete and fragmented assemblies. In this project, we produced high quality assemblies of the allopolyploid *C. arabica* genome and one of its ancestors, the diploid *C. eugenioides* genome. The genome of the other ancestor, *C. canephora* was published previously (Denoeud *et al*. 2014) and is publicly available. Our success in the assembly project stemmed from using ~80X coverage long PacBio reads, error corrected with ~120X coverage of PCR-free Illumina 2x250bp paired end reads for the initial genome assembly using the MaSuRCA assembler (Zimin *et al*. 2017), combined with assembly of just PacBio reads with Falcon-unzip assembler (Chin *et al*. 2016), followed by mid- and long-range scaffolding using 10X Genomics (Gemcode and Chromium), and Dovetail Hi-C data. The availability of both ancestral genomes for *C. arabica* combined with genetic (Moncada *et al*. 2016) and physical mapping data allowed us for the first time to gain a panoramic view of all the homologous chromosomes of *C. arabica* corresponding to each sub-genome. For *C. arabica*, we were able to achieve a contig N50 of 3.91 Mb, and were able to construct the 22 pseudo-chromosomes split by sub-species. Our assembly is the most contiguous and complete so far generated for this species with 91% of the genome anchored to chromosomes. An accurate chromosome scaffolded high quality genome reference assembly is crucial for advancing coffee genomics and climate change adaptation studies. The assemblies *of C. arabica* and *C. eugenioides* are being used to improve downstream analyses, including gene annotation, synteny, comparative genomics and population genetics using natural and breeding populations being phenotyped for climate change adaptation.

This research is co-funded by the US National Science Foundation (Award 1444893), the Inter-American Development Bank, and the Colombian National Coffee Growers Federation (FNC).

This abstract will be presented by coauthors A. Zimin (genome assembly and chromosome scaffolding); C. Maldonado (physical mapping) and M. Yepes (project introduction and outreach).

W205: Coffee Genomics

Targeted Sequencing in Allopolyploids: Comparison of BAC-By-BAC and Whole Genome Approaches Using Third Generation Sequencing

Carlos Ernesto Maldonado[1], Beatriz Padilla[2], Alvaro Gaitán[1], Marcela Yepes[3], Aleksey Zimin[4,5], Keithanne Mockaitis[6,7], Carrie Ganote[8], Sheri A. Sanders[6], Dave Kudrna[9], Rod A. Wing[9] and Herb Aldwinckle[10], (1)Centro Nacional de Investigaciones de Café, CENICAFE, Chinchiná, Colombia, (2)Universidad Católica de Manizales, Manizales, Colombia, (3)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (4)Johns Hopkins University, Department of Computer Science, Baltimore, MD, (5)University of Maryland, College Park, MD, (6)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (7)Department of Biology, Indiana University, Bloomington, IN, (8)National Center for Genome Analysis Support, Pervasive Technology Institute/ Indiana University, Bloomington, IN, (9)Arizona Genomics Institute, University of Arizona, Tucson, AZ, (10)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY

A targeted sequencing approach was implemented to characterize genomic regions containig QTLs associated with important agronomic traits in the allotetraploid *Coffea arabica*. The physical map of *C. arabica* var. Caturra was integrated with the *C. canephora* genome using the programs FPC and SyMap, and the Minimum Tilling Path (MTP) that covers the target region was selected, sequenced by single molecule real time sequencing (SMRT-Seq) and assembled by HGAP and postHGAP software from PACBio and the Arizona Genomics Institute respectively. A second dataset included the whole genome obtained by PACBio long read sequencing of 20 Kb libraries (WGS-SMRT ~80X

coverage) and assembled using Falcon/Falcon Unzip (Chin *et al*. 2016) and MaSuRCA (Zimin *et al*. 2017) and the transcriptome from pooled and MID labeled tissues (meristem, leaves and flowers), whole length transcrips were obtained by cDNA sequencing (Iso-Seq PacBio) of *C. arabica* var. Caturra. For both approaches a region covering 30 cM upstream and downstream from the markers (SNV, DArT and SSR) associated with the QTL was determined by the integration of genetic and physical maps of *C. arabica* and the *C. canephora* genome, used as a reference, by alignment of the specific marker or BAC end sequence (BES) using BLAT or e-PCR as mapping tools. The *C. arabica* WGS was aligned to *C. canephora* using SyMap and the contigs covering the target region were selected. As was expected for an allotetraploid species several WGS contigs mapped over the same reference region; contigs with higher identity to *C. canephora* were assigned to that subgenome and those with lower identity to the *C. eugenioides* subgenome. Contigs of non-overlapping regions were assigned to both subgenomes in the scaffolding process. In the case of the BAC-by-BAC sequencing the assembly was performed using Minimus. Gene prediction was done using the program MAKER and functional annotation was done uysing Blast2GO-pro, and repeats annotation using REPET.
An assembled sequence of 7.38 Mb with 1,188 predicted genes was obtained from the BAC-by-BAC approach. WGS derived scaffolds of 10.2 Mb with 1,675 predicted genes for the *C. canephora* subgenome and 10.7 Mb with 1,991 for *C. eugenioides* subgenome. Dot-Plot analysis showed evidence of chimeric assembly between subgenomes in the BAC-by-BAC approach. By functional annotation the enzyme category with greater representation was transferases. The results of Interproscan showed 11 NBS domains of disease resistance genes in the *C. eugenioides* subgenome, and 9 in the *C. canephora* subgenome. These results coupled with annotation of P-loop domains and leucine-rich regions domains suggest that this region could be associated with disease resistance that could be contributing to coffee yield. Overall, the best sequencing strategy to obtain high quality data from the tetraploid genome was WGS sequencing, demonstrated by no apparent presence of chimeric regions and a clear differentiation of subregions between subgenomes.

W206: Coffee Genomics

Resistant, Resilient, Highly Productive, High Cup-Quality *Coffea arabica* Varieties for the Colombian Coffee Farmers on the 80th Anniversary of the Colombian National Coffee Research Center (CENICAFE)

Carmenza E. Góngora[1], Claudia Flórez[1], Carlos Ernesto Maldonado[1], Marcela Yepes[2], Aleksey Zimin[3,4], Keithanne Mockaitis[5,6], James A Yorke[3], 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, (3)University of Maryland, College Park, MD, (4)Johns Hopkins University, Baltimore, MD, (5)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (6)Department of Biology, Indiana University, Bloomington, IN

CENICAFE was founded by the Colombian National Coffee Growers Federation in 1938 with the mission to generate scientific knowledge and technologies for sustainable coffee production for the Colombian coffee growers. Research covers a wide agronomical spectrum, from breeding of coffee varieties to improved planting and harvesting/post-harvesting practices.
Coffee genomics and transcriptomics research is on-going in collaboration with Cornell University, University of Maryland/ Johns Hopkins University, and Indiana University to identify the location of genes of interest (resistance to diseases and pests, physiological traits and yield) on the chromosomes, their sequences (structural genomics) and function (functional genomics). Sequencing of the coffee genome has been one of our major milestones to develop advanced genomics tools to accelerate the development of varieties resilient to climate change with enhanced use of the diversity present in non-cultivated *Coffea* germplasm. The project also targets the genome of the coffee berry borer, *Hypothenemus hampei* the major insect pest of coffee. Genomic and transcriptomic studies are on-going on the biological mechanisms involved on insect/coffee interactions using host genotypes with differential response for analysis of pathways and novel candidate genes that could enhance pest management strategies.
In addition, CENICAFE maintains one of the largest *ex situ* collections of *Coffea arabica* outside its center of origin Ethiopia, that includes >1,000 accessions. The collection is maintained in the field and is being genotyped and phenotyped for traits of interest including resilience to climate change (biotic and abiotic stresses), plant architecture, high yield and high quality. Microsatellites were used to select 190 accessions of *C. arabica* that were characterized by genotype by sequencing (GBS). Genomic variants were mapped on our recently generated *C. arabica* reference genome assembly to identify 1.15 million di-allelic SNPs. This analysis allowed determination of population structure with seven ancestral populations (K) with diversity coefficients (Fst) between populations varying from 0.179 to 0.487.
A mini core collection was selected based on this study and planted under contrasting environmental conditions throughout Colombia, and is being phenotyped for agronomically important traits. Integration of genotypic and phenotypic information has been crucial to reshaping the current composition of the rust resistant Castillo variety and for the release of the Cenicafe1 variety and Regional Castillo variants, as well as to accelerate the development of new varieties that are highly productive, durably resistant and resilient, in addition to yielding excellent cup quality.
This abstract will be co-presented by co-authors C. Góngora (CENICAFE sustainable coffee production, coffee berry borer/coffee plant interaction), C. Flórez (phenotyping, abiotic stress and climate change), and C. Maldonado (Population Structure and Diversity studies in non-cultivated *Coffea* germplasm).
This multi-component presentation will have an extended time (30 min).

W207: Coffee Genomics

Coffea Centromeric Retrotransposons Play the Central Role in Centromere Organization and Function

Romain Guyot, Institut de Recherche pour le Developpement, Montpellier cedex 5, France

Centromeric regions of plants are generally composed of large array of satellites organized with a specific lineage of *Gypsy* LTR-retrotransposons, called Centromeric Retrotransposons (CR). Repeated sequences interact with a specific H3 histone, playing a crucial function in the kinetochore formation. To study the structure and composition of centromeric regions in *Coffea*, we annotated and classified into ten distinct families Centromeric Retrotransposons sequences (called hereafter CRC) from *C. arabica* genome and its two diploid ancestors: *Coffea canephora* and *C. eugenioides*. The sequence mapping and FISH experiments of CRC Reverse Transcriptase domains in *C. canephora, C. eugenioides* and *C. arabica* clearly indicate a strong and specific targeting mainly onto centromeric regions, which can be associated also with

heterochromatin. Sequence analysis of putative centromeric regions on *C. arabica* and *C. canephora* chromosomes showed an exceptional density of one family of CRC elements, and the complete absence of satellite arrays, contrasting with usual structure of plant centromeres. Altogether, our data demonstrated a specific centromere organization in *Coffea*, suggesting that one CRC family alone might play the central role for the continuation of the centromere function.

W208: Coffee Genomics
Better Coffee Quality from the Lower Canopy?
Bing Cheng, QAAFI, The University of Queensland, St Lucia, Queensland, Australia
From an evolutionary viewpoint, seeds have developed to store nutrients to support plant reproduction. They also have a defence system to survive environmental treats to survival of the seed. With the threat of climate change, a comprehensive understanding of how the growth environment influences seed maturation and composition is required. The Arabica coffee bean, a tropical dicotyledonous albuminous seed, was used in this research to study how seed ripening was influenced by the micro-environment as determined by canopy position. The transcriptome of the developing coffee beans (covered by green, yellow and red pericarp) was analysed for both the upper and the lower canopy (above and below 170 cm). A long read coffee transcriptome obtained from the same samples was used as a reference. Comparative transcriptome analysed was used to investigate the influence of canopy position at different developmental stages. Phenotypic variations of the bean were analysed, including the key physical and chemical attributes influencing coffee quality. Additional sensory testing was also conducted to investigate the aromatic differences in the beans. This comprehensive study will facilitate an improved understanding of the molecular basis coffee quality and provide insights to the variation of seed ripening in different canopy position that produce different phenotypic traits.

W209: Coffee Genomics
Using the *Coffea arabica* Genome to Expedite Coffee Breeding
Dominique Crouzillat, Centre R&D Nestlé Tours, Tours, France and Arabica Coffee Genome Consortium
The Arabica Coffee Genome Consortium (ACGC) undertook the sequencing of the *Coffea arabica* genome and that of its two parents: *C. eugenioides* and *C. canephora*. The resulting genomic resources were used to better assess the *C. arabica* genetic diversity and characterize the neo-diversification induced by man through selection and/or introgression. For this purpose, the consortium re-sequenced 21 wild accessions and 14 varieties of *C. arabica*. The *C. arabica* lectotype conserved at the Natural History Museum in London was also re-sequenced using ancient DNA sequencing techniques. This re-sequencing confirmed the drastic genetic bottleneck that occurred with the cultivation of *C. arabica* varieties and the genetic reservoir still available in the wild pool. It also allowed the development of modern genomic tools such as DNA chip. Furthermore, it brings a large amount of genetic markers covering the entire genome allowing new breeding approaches (GWAS) for this economically important crop species.
In parallel, ultra high-density genetic maps were obtained for *C. arabica* and *C. canephora* by sequencing segregating populations, and thus allowing to anchor genomic to genetic data. This will allow a more fine definition of QTLs for a large number of agronomic and sensory traits and precise their relation to candidate genes.
All the obtained data will be made available through a public database giving access to performing tools allowing breeders and agronomists to answer present and future challenges such as higher tolerance to pests, quality of the coffee beverage and resilience to climate change

W210: Comparative Genomics
Regulatory Networks Underlying Diverse Inflorescence Architectures in Panicoid Cereals
Andrea L. Eveland, Donald Danforth Plant Science Center, St. Louis, MO

W211: Comparative Genomics
Subgenome Dominance across Time and Ploidy in *Mimulus*
Joshua R. Puzey, College of William and Mary, Williamsburg, VA

W212: Comparative Genomics
Long-Distance Regulatory Elements in Plant Genomes
Xiaoyu Zhang, University of Georgia, Athens, GA

W213: Comparative Genomics
From Cactus to Pitcher Plants: Dissecting Molecular Evolution in Diverse Non-Model Plants using Comparative Genomics and Transcriptomics
Ya Yang, University of Minnesota, Saint Paul, MN

W214: Comparative Genomics
"All Genomes Great and Small" Illuminate the Causes of Genome Evolution
Jan Dvorak, Department of Plant Sciences, University of California, Davis, Davis, CA
The recent publication of genome sequences of wild emmer (http://science.sciencemag.org/content/357/6346/93), the source of the bread wheat A and B genomes, and *Aegilops tauschii* (https://www.nature.com/articles/nature24486), the source of the bread wheat D genome, made it possible for the first time to compare the structure and evolution of grass genomes greatly differing in size and content of transposable elements (TEs). Gene collinearity was quantified along the pseudomolecules of large genomes of wild emmer and *Ae. tauschii* and compared with that of the small genomes of *B. distachyon,* rice, and sorghum. The quantification revealed that rates of genomic change have been slow in the rice and sorghum phylogenetic lineages, faster in the phylogenetic branches of subfamily Pooideae, including that of *B. distachyon*, but greatly

accelerated in the lineages of the annual Triticeae species, wild emmer and *Ae. tauschii*. The *Ae. tauschii* genome was shown to have a greater number of duplicated genes than other sequenced genomes. The *Ae. tauschii* and wheat genomes were also shown to have greater amount of homogeneous TEs than other sequenced plant genomes. Gene duplications and gene collinearity along the *Ae. tauschii* pseudomolecules correlated with recombination rates. These relationships suggest that in the annual Triticeae species, the vast amounts of homogeneous TEs cause frequent errors in recombination and lead to frequent gene duplications and structural chromosome changes, the hallmarks of fast genomic change.

W215: Comparative Genomics

The Pan-Genome of the Diploid Grass *Brachypodium distachyon* and its Implications for Polyploid Genome Evolution

Sean Gordon[1], Bruno Contreras-Moreira[2], Daniel Woods[3], David L. Des Marais[4], Diane Burgess[5], Shengqiang Shu[1], Christoph Stritt[6], Anne Roulin[6], Wendy Schackwitz[1], Ludmila Tyler[7], Joel Martin[1], Anna Lipzen[1], Niklas Dochy[8], Jeremy L. Phillips[1], Kerrie W. Barry[1], Koen Geuten[8], Hikmet Budak[9], Thomas Juenger[10], Ana Caicedo[11], Richard Amasino[3], David M. Goodstein[1], Patrick Davidson[1], Luis Mur[12], Melania Figueroa[13], Vinh Ha Dinh Thi[14], Boulos Chalhoub[15], Virginia Markham[16], Michael R Freeling[5], Pilar Catalan[17], **John Vogel**[1,16], Jeremy Schmutz[18] and Jerry Jenkins[18], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)Fundación ARAID, Zaragoza, Spain, (3)University of Wisconsin-Madison, Madison, WI, (4)The University of Texas at Austin, Austin, TX, (5)University of California, Berkeley, CA, (6)University of Zürich, Zürich, Switzerland, (7)Biochemistry and Molecular Biology Department, Amherst, MA, (8)University of Leuven, Leuven, Belgium, (9)Montana State University, Bozeman, MT, (10)University of Texas at Austin, Austin, TX, (11)Biology Department, University of Massachusetts, Amherst, MA, (12)Aberystwyth University, Aberystwyth, United Kingdom, (13)University of Minnesota, St. Paul, MN, (14)Université d'Evry Val d'Essonne, Evry, France, (15)UEVE-INRA, Evry, France, (16)University of California Berkeley, Walnut Creek, CA, (17)Department of Agriculture (Botany), High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain, (18)HudsonAlpha Institute for Biotechnology, Huntsville, AL

While prokaryotic pan-genomes have been shown to contain many more genes than any individual organism, the prevalence and functional significance of differentially present genes in eukaryotes remains poorly understood. Whole-genome *de novo* assembly and annotation of 54 lines of the grass *Brachypodium distachyon* yield a pan-genome containing nearly twice the number of genes found in any individual genome. Genes present in all lines are enriched for essential biological functions, while genes present in only some lines are enriched for conditionally beneficial functions (e.g. defense and development), display faster evolutionary rates, lie closer to transposable elements and are less likely to be syntenic with orthologous genes in other grasses. Our data suggest that differentially present genes contribute substantially to phenotypic variation within a eukaryote species, these genes have a major influence in population genetics, and transposable elements play a key role in pan-genome evolution. In addition, the pan-genome provides a new lens though which we can examine genome evolution in polyploid species by enabling us to differentiate between polymorphisms that happened after polyploidization and those that were part of the standing variation in the diploid progenitors. To explore this, we sequenced the allopolyploid *B. hybridum* and compared the sub-genome that was derived from the diploid *B. distachyon* to our pan-genome. Interestingly, we noted multiple origins for *B. hybridum* spread over time creating a natural time course to study polyploid genome evolution.

W216: Components of Apomixis

Diploidization Processes may Produce Genomically Unique Sexual Species from Alloploid *Boechera* Apomicts

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

Boechera (Brassicaceae) contains ca. 83 inbreeding sexual species and many thousands of alloploid hybrids. Reticulate evolution (speciation involving interspecies hybridization) readily explains the proliferation of alloploid apomicts, but how the many sexual diploid species originated is less obvious. The traditional view is that they arose by divergent speciation, i.e., range expansion of sexual diploids followed by selection and speciation along ecological gradients. However, the large number of sexual *Boechera* species, many of which have geographically restricted ranges, is not consistent with such an origin. Evidence will be presented that many of these sexual diploid *Boechera* evolved like their apomictic counterparts, by reticulation, and that an intermediate phase of facultative apomixis facilitated their speciation. Newly formed allodiploid apomicts of *Boechera* are immortal in that they can clone themselves through their seed indefinitely. But they also are facultatively sexual. With each infrequent generation of sexual inbreeding, segregation and assortment reduces homoeologous chromosome heterozygosity by 50 %. After several sexual generations, perhaps interspersed by multiple apomictic generations, allodiploid lineages become diploidized, i.e., their chromosomes become chiasmata-generated composites of alternating homozygous sections of the homoeologous chromosomes of their original parents. During this recombinational diploidization process, genetic information responsible for apomixis may be lost, and facultative sex may become obligate sex. Most alleles of species originating by divergent evolution originate from a single ancestor. In contrast, alleles of species originating by reticulation reflect a diverse ancestry. In the simplest case, near-equal contributions of alleles from two divergent ancestors are evident in a species that evolved by reticulation. We will present microsatellite data from the *Boechera* Microsatellite Website as evidence for the evolution of at least several sexual diploids from apomictic alloploids by apomixis-facilitated reticulation and diploidization.

W217: Components of Apomixis

A Genomic Approach to Study Apomixis using *Eragrostis curvula a*s a Model Species

José Carballo[1], Ingrid Garbus[1,2], Juan P. Selva[1], Bruno A. Santos[3], Emidio Albertini[4], Mario J Caccamo[3] and **Viviana Echenique**[1,5], (1)CERZOS CCT_CONICET Bahía Blanca, Bahía Blanca, Argentina, (2)Departamento de Ciencias de la Salud - Universidad Nacional del Sur - Argentina, Bahía Blanca, Argentina, (3)NIAB, Cambridge, United Kingdom, (4)University of Perugia, Perugia, Italy, (5)Dpto. Agronomía and CERZOS/CONICET, Bahía Blanca, Argentina

Apomixis is defined as asexual reproduction by seeds, avoiding meiotic reduction and fertilization, being generally present in polyploid plant species. *Eragrostis curvula* is a perennial grass native to Southern Africa. This species can be taken as a model for the discovery of genes that

govern pseudogamous diplosporic apomixis since its polyploid cytotypes (4x to 8x) may undergo sexual reproduction, facultative apomixis, or obligate apomixis whereas diploids are always sexual.
Here we present the first draft of a diploid version of the *E. curvula* genome. The cultivar selected was Victoria (~1200 Mb) originated from *in vitro* culture of inflorescences of the apomictic cv. Tanganyika (2n=4x=40). Two libraries were prepared with fragment lengths of 20 kb and 10 kb to get longer reads and to increase the coverage, respectively. The sequencing through PACBIO technology resulted in 6.223.627 and 3.309.811 reads respectively with 90X coverage. The assembly was performed using the software Falcon. The N50 was 380.026 bp with 3.118 contigs representing 95% of the haplotype length. The software BUSCO was used to find single copy orthologous genes, being represented 97% of the BUSCO genes. Dovetail Hi-rise software revealed the architecture of the complete genome chromosome-by-chromosome.
The first draft of *E. curvula* genome showed high level of contiguity with a coverage of 95% of the diploid genome. The high proportion of annotated genes would allow the identification of those related to the reproductive mode. This draft represents the start point to obtain more complex tetraploid genomes, harboring the region/s involved in apomixis.

W218: Components of Apomixis
Efficient Hybrid Breeding Based on the Modification of Meiosis Using Virus- Induced-Gene-Silencing
Vanesa Calvo, WUR Wageningen University, Wageningen, Netherlands
Vanessa Calvo Baltanas[1], Cris Wijnen, Nina Lukhovitskaya, Basiaan de Snoo, Linus, Hans de jong[1], Arp Schnittger and Erik Wijnker[1].
Hybrid production in traditional breeding programmes is usually a lengthy process. Obtaining new inbred lines or finding favourable parental combinations requires a great investment of resources. To overcome these issues, we have developed a new breeding approach to produce hybrids more efficiently. First, it allows to select and fix any unknown hybrid genotype. Secondly, we can also obtain near-full hybrids that can perform either as the initial hybrid or potentially better. Our new breeding technique is based on the reduction of 80% of meiotic crossovers, by silencing MSH5 directly in the hybrid, using Virus-Induced-Gene-Silencing (VIGS). The major advantages of this new application are: i) absence of stable transgenes to modify gene expression ii) rapid generation of hybrids (only 3 generations from the initial hybrid) and iii) the production of new parental lines that are either chromosome substitution lines or low-recombinant lines (1 or 2 recombination events). This efficient hybrid breeding approach based on the modification of meiosis brings a new insight to hybrid production and evaluates hybrid performance by comparing full hybrids with rapidly obtained near-full hybrids.

W219: Components of Apomixis
Apomixis and Heterochrony in Boechera: Stressful Connections Revisited
Amal Johnston, University of Heidelberg, Heidelberg, Germany

W220: Components of Apomixis
Laser Assisted Micro-Dissection of Egg Apparatus from (a)Sexual Dandelion (*Taraxcum*) to Resolve the Genetic Basis of Parthenogenesis
Kitty Vijverberg, Carla Oplaat and M. Eric Schranz, Wageningen University & Research, Biosystematics Group, Wageningen, Netherlands
We aim to unravel the molecular basis of the initiation of plant embryo development without fertilization (*Parthenogenesis*) and to apply this in protocols for breeding line production and maintenance, e.g., for the induction of embryogenesis in (doubled) haploid production or in (un)reduced gametes obtained after modified steps in meiosis, to produce (homozygous) breeding lines of interest. Combining parthenogenesis with the omission of meiosis is also of great interest in breeding, to maintain vigorous F1-hybrids via clonal seed production. Parthenogenesis occurs naturally in a number of plant species in combination with apomeiosis (unreduced gametes), a reproduction mode known as *Apomixis*. Within the common dandelions (*Taraxacum*), both sexual as well as apomictic plants occur, making it an ideal model system for comparative analysis of their reproductive pathway. We isolated triplicates of egg apparatus (egg cells and synergids) and central cells from the young-mature embryo sac of an apomictic and sexual accession and compared their amplified, Illumina sequenced transcriptomes in order to find clues for the genetic basis of Parthenogenesis.

W221: Components of Apomixis
***Boechera* Species: *de novo* Assembly of Genomes of Sexual and Apomictic Accessions and Apomixis Associated Genes Analysis**
Vladimir Brukhin[1], Sergei Kliver[1], Thomas Mitchell-Olds[2], Ueli Grossniklaus[3], Catherine Rushworth[4], Jeremy Schmutz[5], Mikhail Rayko[6] and Daniel S. Rokhsar[7], (1)Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russian Federation, (2)Duke University, Durham, NC, (3)University of Zurich, Department of Plant and Microbal Biology, Zurich, Switzerland, (4)University and Jepson Herbaria, University of California, Berkeley, CA, (5)Hudson Alpha, Huntsville, AL, (6)Saint Petersburg State University, Saint Petersburg, Russia, (7)DOE Joint Genome Institute, Walnut Creek, CA
The genus *Boechera* (formerly *Arabis*) is known to contain both sexual and apomictic species or accessions within the species. *Boechera retrofracta* is a diploid sexually reproducing species and is thought to be an ancestral parent species of the apomictic *Boechera divaricarpa*. We performed the *de novo* assembly of the *B. retrofracta* genome using short Illumina and Roche reads from 1 paired-end and 3 mate pair libraries. The distribution of 23-mers from the paired end library has indicated a low level of heterozygosity compared to apomictic *B. divaricarpa* and the presence of detectable duplications and triplications. The genome size was estimated to be equal 227 Mb. N50 of the assembled scaffolds was 2.3 Mb. 27048 protein-coding genes were predicted using a hybrid approach that combines homology-based and *de novo* methods. Also repeats, tRNA, and rRNA genes were annotated. Genes of *B. retrofracta* and 6 other Brassicaceae species were used for phylogenetic tree reconstruction. In this regard we also identified the alignment coverage of a genome by other Brassicacea genomes and the alignment identity was calculated as fraction of matches in the alignment. We checked and compared similarity of the several apomixis-associated genes, including *APOLLO,* in genomes of the sexual and apomictic accessions of *Boechera*. *B. retrofracta* is a plant of particular interest as an ancestor of various hybrids including apomictic species. An assembled genome of *B. retrofracta* will help in the challenging assembly of the

highly heterozygous genomes of hybrid apomictic species such as *B. divaricarpa*, and to decipher the hybridogenesis events that took place in the formation of apomictic *Boechera* accessions.

W222: Compositae

Draft *Tragopogon dubius* (Asteraceae) Genome Assembly using Linked-Read Sequencing and a Survey of Alternative Splicing

Xiaoxian Liu, Brad Barbazuk, Pamela S. Soltis and Douglas E. Soltis, University of Florida, Gainesville, FL

Allopolyploid species and their diploid parents in *Tragopogon* (Asteraceae) represent a great model for examining the genomic consequences of recent and recurring allopolyploidy. However, reference genome sequences of species of *Tragopogon* are not available, which limits investigations of the impacts of polyploidizaiton in this evolutionary model system. Here we used the linked-read sequencing approach (10x Genomics) to construct a draft genome assembly of *T. dubius* (2.8 Gb), the shared diploid parent of the recently formed allotetrapoloids *T. mirus* and *T. miscellus*. *De novo* assembly was based on 59.8x coverage of linked-reads and resulted in a genome with an N50 scaffold size of 0.11 Mb and N50 contig size of 17.04 kb. Using long genome scaffolds (>100 kb) as a reference, a preliminary annotation including alternatively spliced transcript isoforms was conducted based on *T. dubius* leaf Iso-Seq™ reads. In all, 11,685 isoforms from 7,572 genes were confirmed and annotated; 28.1% (2,126) of the annotated genes have evidence of producing alternatively spliced transcripts. We are examining patterns of alternative splicing among *T. dubius*, *T. pratensis*, and their tetraploid *T. miscellus*. This draft genome assembly provides extensive genomic resources for transcriptome analysis, especially the genome-wide detection of alternative splicing. It is the foundation for ongoing and future studies of the evolutionary model system *Tragopogon*; the genome will lead to a better understanding of the genetic and genomic consequences of WGD, especially during the early stages after WGD.

W223: Compositae

Building a Reference Genome for *Centrapalus pauciflorus* (Compositae), an African Oilseed Crop

Vanessa Liz Gonzalez, Global Genome Initiative, National Museum of Natural History, Smithsonian Institution, Washington, DC, DC

W224: Compositae

Genetic Analysis of Important Traits in Lettuce

Hanhui Kuang, Huazhong Agricultural University, Wuhan, China

Lettuce is one of the most important vegetable crops worldwide. As a leafy vegetable, the color and shape of leaves are important horticultural traits of lettuce. We used biparental segregating population and natural population to genetically dissect these traits in lettuce. GWAS analysis identified at least five loci controlling leaf color of lettuce, which were verified using biparental segregating populations. Furthermore, dozens of eQTLs for genes in the anthocyanin biosynthesis pathway were identified using GWAS. Pyramiding of these loci a cultivar is expected to increase anthocyanin concentration in leaves. Leaf color is also affected by the concentration of chlorophyll, and natural mutations in the *GLK* and *PhyB* genes contribute considerably variation in green intensity among different lettuce cultivars. The natural variation of the *PhyB* gene also changes leaf angle and flowering time in lettuce. As many as five loci controlling heading formation in crisphead were identified using a combination of bulk segregant analysis (BSA) and second generation sequencing. Fine mapping showed that knockout mutation of the *STM* gene is necessary but not sufficient for the heading in lettuce. The *STM* gene is also involved in undulation of leaf margin. The underlying molecular mechanisms of these genes are being studied. Our results provide critical information for molecular design breeding of lettuce cultivars.

W225: Compositae

Evolution of Invasiveness by Genetic Accommodation in a Perennial Sunflower

Dan G. Bock[1], Michael Kantar[2], Celine Caseys[3], Rémi Matthey-Doret[4] and Loren H. Rieseberg[4], (1)Harvard University, Cambridge, MA, (2)University of Hawaii, Honolulu, HI, (3)University of California - Davis, Davis, CA, (4)University of British Columbia, Vancouver, BC, Canada

W226: Compositae

Towards More Durable Resistance to Lettuce Downy Mildew

Kyle Fletcher[1], Lorena Parra[1], Sebastian Reyes-Chin-Wo[1], Rosa Juliana Gil[1], Lin Zhang[1], Kelsey Wood[1], Aubrey Kenefick[1], Cayla Tsuchida[1], Maria Jose Truco[1], Keri Cavanaugh[1], Frank Martin[2] and Richard Michelmore[1], (1)Genome Center, University of California, Davis, Davis, CA, (2)USDA-ARS, Salinas, CA

Lettuce is a high value vegetable crop that is often grown as extensive monocultures. Downy mildew, caused by *Bremia lactucae* is the most important disease of lettuce worldwide. Numerous resistance genes exist; however, current deployment strategies have failed to provide resistance for prolonged periods of time due variability in the pathogen. In order to maximize the evolutionary hurdle for the pathogen to become virulent we are using genomics approaches to characterize variation for resistance in lettuce and virulence in *B. lactucae*. Illumina, 10x Genomics, PacBio sequencing and Hi-C scaffolding enabled the production of chromosome-scale and near-chromosome-scale assemblies of lettuce and *B. lactucae* respectively. Screening a lettuce diversity panel with resistance gene enrichment sequencing (RenSeq) identified repertoires of nucleotide-binding domain and leucine rich repeat encoding genes associated with resistance phenotypes. In parallel, whole genome sequencing of over 100 isolates of *B. lactucae* identified virulence associated genes with initial functional validation of candidate genes. Deep sequencing revealed that the many isolates of *B. lactucae* are heterokaryotic with component nuclei contributing different virulence phenotypes. This indicates that selection may act on populations of nuclei existing in a coenocytic mycelium. Defining resistance genes and their interactions will enable the insertion of adjustable stacks of resistance genes targeted against *B. lactucae* resulting in a large evolutionary hurdle for the pathogen to overcome. Such gene stacks should provide more durable resistance in lettuce against *B. lactucae*.

W227: Computational Gene Discovery

Flye: Accurate Assembly of Long Error-Prone Reads using Repeat Graphs

Mikhail Kolmogorov[1], Jeffrey Yuan[1], Yu Lin[2] and Pavel Pevzner[1], (1)University of California San Diego, La Jolla, CA, (2)Australian National University, Canberra, Australia

The problem of genome assembly is ultimately linked to the problem of *repeat classification*, the compact characterization of all repeat families in a genome as a *repeat graph*. The key reason the *de Bruijn graph* emerged as a popular short read assembly approach is because it offered an elegant representation of all repeats in a genome. However, most algorithms for assembling long error-prone reads use an alternative *overlap-layout-consensus (OLC)* approach that does not provide a repeat classification. We present the Flye algorithm for classifying repeats and constructing the assembly graph from long error-prone reads. Benchmarking Flye against several state-of-the-art long read assemblers demonstrated that it generates accurate assemblies of eukaryotic genomes. The Flye assembly graph provides a useful framework for planning genome finishing experiments, analysis of the repeat content, and gene prediction.

W228: Computational Gene Discovery

Guided Assembler – Tool for Finding Antimicrobial Resistance Genes

Alexander Souvorov, NCBI/NIH, Bethesda, MD

In many situations, it is desirable to find a particular type of sequence in a set of Illumina short reads. One of important applications is quick detection of antimicrobial resistance (AMR) genes in a sample. The same approach could be used for finding other types of sequences – 16S rrna for taxonomy applications, transcripts using RNA-seq, viral genomes and others.

The input for the guided assembler is a run of Illumina reads and a set of target sequences. The guided assembler builds a de Bruijn graph from the reads. Then it analyses the target sequences and finds kmers which are reasonably close to some kmers in the de Bruijn graph. The kmers found in the graph are used as the seed kmers. The program builds all extension contigs of the seed kmers as long as the contigs and the target sequence have reasonably close alignments. At the end of the process the assembled sequences are cleaned using alignments of the full length reads and mate pair information.

We built a database of known AMR genes and use the guided assembler in our pathogen analysis pipeline for fast and precise detection of the AMR genes.

Guided assembler is publicly available in https://github.com/ncbi/ngs-tools/tree/skesa/tools/skesa

W229: Computational Gene Discovery

Employing Proximity-Ligation Data from Hi-C to Enable Genomic Discovery

Ivan Liachko, Zev Kronenberg, Andrew Wiser, Kaylee Mueller, Maximilian Press and Shawn Sullivan, Phase Genomics, Seattle, WA

The loss of long-range sequence contiguity in the process of NGS sequencing is an obstacle to understanding genes and genetic pathways. This obstacle negatively affects both plant/animal research efforts as well as microbiome-centric gene mining. Fragmented genome assemblies break the large genes in plants and animals across multiple contigs, making them difficult to define and study. This problem is compounded in microbiome datasets, as most contigs cannot be linked dependably to their species of origin.

The chromosome conformation capture method, Hi-C, is able to restore chromosome-scale contiguity to large genome assemblies and enables the deconvolution of numerous genomes from mixed samples such as complex microbial communities. Hi-C captures genomic proximity interactions through *in vivo* crosslinking, followed by proximity-ligation and sequencing. Because the crosslinks are created within intact cells with intact chromosomes, whole-chromosome contiguity is preserved allowing the scaffolding of genomes of virtually any size. Since the crosslinks occur inside intact cells, any two loci that interact by Hi-C must have originated in the same cell, and this data can be used to deconvolute high quality genomes directly from mixed populations.

We have developed two platforms that exploit Hi-C proximity-ligation data: Proximo Hi-C is a genome scaffolding platform that allows chromosome-scale genome assembly, and ProxiMeta Hi-C is a metagenomic deconvolution platform that extracts high numbers of genomes form microbiome samples. Here we will discuss the application of these methods to diverse genome and metagenome projects, highlighting their effect on gene discovery efforts.

W230: Computational Gene Discovery

BUSCO v3: Expanded Plant Coverage and Utilities Beyond Benchmarking.

Felipe A. Simão[1], Mathieu Seppey[1], Robert M. Waterhouse[1], Mose Manni[1], Panagiotis Ioannidis[1], Guennadi Klioutchnikov[2], Nicolas Sierro[3], Nikolai V. Ivanov[3], Evgenia V. Kriventseva[1] and Evgeny M. Zdobnov[1], (1)University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland, (2)Swiss Institute of Bioinformatics, Geneve, Switzerland, (3)PMI R&D, PMP S.A., Neuchatel, Switzerland

The advent of high-throughput genomics has brought about a veritable paradigm shift in biological research. Due to rising demands and increasing volumes of data, technologies and downstream analysis tools have been rapidly evolving. This makes thorough quality control of the 'products' of sequencing data, e.g. genomes, genes, or transcriptomes, essential. Addressing this need, the Benchmarking Universal Single-Copy Orthologues (BUSCO) assessment tool provides intuitive quantitative measures of genomic data completeness in terms of expected gene content (Simão et al, 2015, PMID:26059717, http://busco.ezlab.org). BUSCO assessments identify complete, duplicated, fragmented, and missing genes and enable like-for-like quality comparisons of different datasets. These features mean that BUSCO has rapidly become established as an essential genomics tool, using up-to-date data from many species and with broader utilities than the popular but now discontinued Core Eukaryotic Genes Mapping Approach (Parra et al, 2007, PMID:17332020). Selected from major species clades of the OrthoDB catalog of orthologs, more than forty clade-specific datasets can be used with BUSCO v3, permitting analysis using a large number of highly specific single-copy genes across all domains of life. Here we present a summary of the latest BUSCO features including an increased coverage of plants, as well as a variety of scenarios highlighting the wide range of uses of BUSCO assessments, such as: (i) performing

genomics data quality control, but also applicable for (ii) building robust training sets for gene predictors, (iii) selecting high-quality reference strains or species for comparative analyses, (iv) identifying reliable markers for large-scale phylogenomics studies, and (v) separating haplotypes in highly heterozygous assemblies. DOI: 10.1093/molbev/msx319

W231: Computational Gene Discovery

Rare Event Modeling in Computational Gene Prediction

Alexandre Lomsadze, Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering,Georgia Institute of Technology, Atlanta, GA

U12 introns, introns with "GC" donor sites or species with small number of introns, are examples of rare events in genomes. We present a heuristic approach for predicting such introns in the framework of GeneMark-EX algorithm.

W232: Computational Gene Discovery

BRAKER2: Incorporating Protein Homology Information into Gene Prediction with Genemark-EP and Augustus

Katharina Hoff[1], Alexandre Lomsadze[2], Mario Stanke[1] and Mark Borodovsky[3], (1)University of Greifswald, Greifswald, Germany, (2)Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering,Georgia Institute of Technology, Atlanta, GA, (3)Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA

The rapidly growing number of sequenced genomes requires fully automated methods for accurate gene structure annotation. With this goal in mind, we have developed BRAKER1, a combination of GeneMark-ET and AUGUSTUS, that uses genomic and RNA-Seq data to automatically generate full gene structure annotation in novel genomes. BRAKER2 is an extension of this earlier developed tool which allows for the integration of protein homology information.

In presence of RNA-Seq data, BRAKER2 runs self-training GeneMark-ET supported by RNA-Seq data, trains AUGUSTUS on the basis of GeneMark-ET predictions, maps protein sequences of related species to the genome in question and incorporates both the RNA-Seq and protein information into gene prediction with AUGUSTUS.

In absence of RNA-Seq data, BRAKER2 executes self-training GeneMark-EP supported by protein data, trains AUGUSTUS on the basis of GeneMark-EP predictions and predicts genes with protein homology information with AUGUSTUS.

BRAKER2 is available for download at http://bioinf.uni-greifswald.de/bioinf/braker/ and http://exon.gatech.edu.

W233: Connecting Crop Phenotype and Genotype Data

Challenges and Opportunities of Connecting Phenotype with Genotype; Perspectives from Seeds of Discovery and Excellence in Breeding.

Sarah Hearne, CIMMYT International Maize and Wheat Improvement Center, Texcoco, Mexico, Kate A. Dreher, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico and Kelly Robbins, Cornell University, Ithaca, NY

Management of data along complex and logistically challenged crop research and breeding processes is often the last thing to be considered in planning processes and is consequently an Achilles heel, limiting the impact of many projects and initiatives. Here we present a review of some of the challenges, interventions and opportunities in data management from the perspectives of an established initiative and a new cross commodity platform.

The Seeds of Discovery initiative (SeeD) aims to explore and leverage high value novel diversity for maize and wheat breeding application. In the 6.5 years since inception, SeeD has generated vast quantities of genotypic and phenotypic data from extensive evaluation of germplasm bank accessions. The effective management of this data along the collection, curation, analysis and dissemination continuum has evolved, resulting in multidisciplinary and multi-institutional development of systems, standard operating procedures and business rules. We present some of the experiences of SeeD in developing effective practices to connect genotype with phenotype.

The CGIAR Excellence in Breeding Platform (EiB), established in 2017, is developing a resource and support structure to modernize breeding programs targeting the developing world. EiB draws from innovations in the public and private sector to provide access to cutting-edge tools, services and best practices, training and practical advice for breeding programs. Data accuracy, integrity and interconnectivity are fundamental to breeding gain and EiB is placing strong emphasis on the sharing of best practices and resources in this area. Current plans and initial platform activities in this area will be presented.

W234: Connecting Crop Phenotype and Genotype Data

G.E.M.S™: Enabling Agricultural Innovation and the International Agroinformatics Alliance.

Kevin A. T. Silverstein, Supercomputing Institute, University of Minnesota, Minneapolis, MN

In a strategic partnership, the College of Food, Agricultural and Natural Resources Sciences at the University of Minnesota has teamed up with the Minnesota Supercomputing Institute to create a novel data sharing and analysis platform called G.E.M.S™. This platform, once launched in early 2018, will enable public-private research collaborations for innovation in agricultural production and other domain areas. The platform allows data contributors to truly make their data interoperable, via structured ontologies and data cleaning modules, and allows them to share only those fields within their data sets that they choose to make available.

Within this framework, communities that otherwise might not have taken shape are forming, such as the International Agroinformatics Alliance (IAA). The IAA is in the process of finalizing agreements with a number of major public and private research agencies based in the United States and elsewhere in the world, and is in active discussion with other potential partners.

W235: Connecting Crop Phenotype and Genotype Data

James Hutton Institute Informatics Visualization Software

Iain Milne, The James Hutton Institute, Dundee, United Kingdom

Modern day genomics, genetics and plant breeding relies on high-throughput sequencing and phenotyping technologies, generating data sets that have been increasing over the last few years in terms of both size and complexity. Storage and analysis of these diverse data sets is only possible via a combination of utilizing data warehousing, HPC clusters (often via the parallelization of existing code sets), information visualization, and visual analytics technologies.
At the James Hutton Institute we develop novel software tools, web applications, databases and information resources (https://ics.hutton.ac.uk/software) which allow users to explore and query their data in logical and intuitive ways – ultimately leading to improved solutions for scientific data management and information dissemination.
Germinate 3 stores various diverse data types and acts as a hub for other tools: Flapjack is built around graphical genotyping enabling users to sort and manipulate lines based on their genotype or on observed or predicted phenotypes; CurlyWhirly is a 3D viewer for PCA/PCo data; Tablet provides 2nd-generation sequence assembly and alignment visualization; and Helium utilizes plant pedigrees as a visualization framework. These resources are used by many institutes, companies, and large international projects, including the Genomic & Open-source Breeding Informatics Initiative (GOBII), Seeds of Discovery (including UK Seed), Crop Wild Relatives, and the International Wheat Yield Partnership.
We will describe our tools, and the benefits they offer, and show how they're evolving to embrace new technologies such as the Plant Breeding API (BrAPI) for connectivity with external tools and resources, and Galaxy for the pipelining of analyses/HPC utilization.

W236: Connecting Crop Phenotype and Genotype Data
Integrating Cassavabase with the GOBII System
Lukas Mueller, Boyce Thompson Institute, Ithaca, NY
Cassavabase (https://cassavabase.org/) is a large and comprehensive database for cassava breeding for the NextGen Cassava project (https://nextgencassava.org/). Cassavabase is designed to enable the genomic selection breeding paradigm, where extensive phenotypic datasets are correlated with genotypic information to predict the traits of lines that have only been genotyped. As genotyping is becoming faster and cheaper than phenotyping, this can result in faster breeding cycles, increasing the overall rate of gain. Cassavabase is based on a large relational database, with the major database schemas adapted from the Chado database system. Until now, the relational database stored all the information, including the phenotypic values and the genotype information. The phenotypic values are stored in conjunction with field trial metadata and are based on trait ontologies developed in collaboration with the CropOntology project. Other data stored in the database include pedigree and crossing data. Genotyping data is stored as indexed (jsonb). Recently, a new system, GOBII (http://gobiiproject.org/), has been available for the efficient storage and retrieval of genotyping data. We are in the process of adopting the GOBII system as our central genotyping backend and integrating it with Cassavabase and other databases, such as Musabase, Yambase and Sweetpotatobase. An important ingredient in this integration is the standard breeding application programming interface (BrAPI, https://brapi.org/), which allows information exchange between breeding databases using a standardized calls and data formats.

W237: Connecting Crop Phenotype and Genotype Data
Using the Tripal Breeding Information Management System (BIMS) to Enable Efficient Management of Phenotypic and Genotypic Data
Sook Jung[1], Taein Lee[1], 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 volumes of data that require efficient management systems to keep track of performance, pedigree, geographical and image-based data. With the development of DNA-based screening technologies, more breeding programs perform genotyping in addition to phenotyping for performance prediction or evaluation. The integration of private breeding data with publicly available genomic and genetic data in a database can help enhance genetic understanding of important crop traits and maximize marker-assisted breeding utility by crop breeders and allied scientists. We report progress on the Tripal Breeding Information Management System (BIMS) which we have implemented in in GDR (Genome Database for Rosaceae) and CottonGEN. 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 functionality. BIMS incorporates the use of an Android App called Field Book, an open-source software for phones and tablets, which allows 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 collection. The current functionality includes manage breeding, data import, search and download and basic statistical analysis components.

W238: Connecting Crop Phenotype and Genotype Data
IRRI's Pipeline for Integrating Genotype and Phenotype Data for Enhanced Product Development
Joshua N. Cobb, Juan D. Arbelaez, John Carlos I. Ignacio, Ramil Mauleon, Marko Karkkainen and Tobias Kretzschmar, International Rice Research Institute, Los Baños, Philippines
As part of a concerted effort to modernize the rice breeding efforts at the International Rice Research Center, significant resources have been deployed to ensure that breeders' decisions are supported by the best available information.The effective organization and storage of genotype data and phenotype data has been central to this effort, has have innovations applied to the analysis of that data to turn it into actionable information. This includes the development and deployment of B4R as a trial management tool for phenotype data, GOBII as a mechanism for storing and organizing genotype data, and the Galaxy analysis portal for helping bring the two data sets together. IRRI's pipeline for collecting, storing, and appropriately associating genotype and phenotype data and the entry points for that information into the breeding strategy will be discussed.

W239: Connecting Crop Phenotype and Genotype Data

CIMMYT's Phenotype and Genotype Integration and MABC and Genomic Selection Use Cases

Michael S. Olsen, CIMMYT, MINNEAPOLIS, MN, Kenya

The CIMMYT Global Maize Program initiated introgression of putative QTLs for maize lethal necrosis (MLN) tolerance into several elite African susceptible lines in response to the rapid emergence of MLN as a major threat to maize productivity in eastern Africa. The goal of the effort was to improve hybrid yield by greater than 10% under MLN pressure while maintaining yield equivalency in the absence of MLN. Twenty-six African recurrent parent (RP) inbred lines were involved targeted using six MLN donor lines, three each for heterotic group A and heterotic group B RPs. Five MLN tolerant yellow grain lines from Latin America were converted to white grain and a major effect QTL for maize streak virus (MSV) was introgressed into the MLN tolerant lines. During the final stages of line conversion, BC4F3:4 ears were prioritized for advancement using %RP and favorable target allele count information displayed through an MABC feature added to FlapJack by the James Hutton Institute in conjunction with the Genomics and Open-source Breeding Informatics Initiative (GOBII) project. For 23 of the 26 recurrent parents we introgressed MLN QTL leads into, we surpassed our stated target of increasing yield under MLN pressure by 10% or greater. Most of the RP lines were improved by 25-50% for grain yield under severe MLN pressure, typically on the order of 2 to 3 t/ha advantage compared with the RP control hybrid under conditions where susceptible commercial checks have near zero grain yield. Although hybrid equivalency data is still fairly preliminary, in general two to four promising MLN-improved BC4F4 families which show similar testcross yield to the RP control hybrid in the absence of MLN have been identified. Similarly, four of the five Latin America yellow MLN tolerant lines were successfully converted to white grain versions containing a favorable MSV1 allele and demonstrate similar testcross yield levels to RP control hybrids under MLN and non-MLN conditions. The conversion program considerably expands the scope of utility of some of the most important maize lines in commercial use in east Africa and brings new MLN tolerant lines into use for new hybrid identification.

W240: Connecting Crop Phenotype and Genotype Data

Identifier Services for Distributed Data Management

Ramona Walls, CyVerse, Tucson, AZ, Maria Esteva, University of Texas, Austin, TX, Andrew B. Magill, Texas Advanced Computing Center, Austin, TX, Ming Chen, University of Arizona, Tucson, AZ and James Carson, Texas Advances Computing Center, University of Texas, Austin, TX

The integration and management of data for crop phenotype studies presents multiple challenges. Many phenotype datasets are big, have multiple contributors, contain components at different stages of completion, and are stored across different platforms. Data often have multiple identifiers (local or global) that need to be managed pre- and post-publication, and data identifiers need to link to identifiers and metadata for other objects such as specimens, accessions, projects, and publications. Manually performing data management actions for datasets containing hundreds or even thousands of files is tedious at best, impossible at worst. Solving the genotype-to-phenotype challenge requires that data be discoverable, trustworthy, interpretable, and accessible, no matter where they are located or what stage of completion they are in. Therefore, a new generation of data management tools is needed for the kind of big data being generated by crop phenotyping studies. Identifier Services (IDS) is an Early-concept Grants for Exploratory Research (EAGER) project that is exploring technical solutions to managing large, distributed data. IDS developed a number of proof-of-concept micro-services for scientists to register their data, organize and describe them with metadata, and run checks for identity and location. IDS records the relations among data components, including those stored across repositories and storage platforms and in different stages of completion. These relations provide a representation of the dataset, based on a simple generic data model that can be adjusted to represent different types of research. The data model in turn supports management of large datasets, through tasks like bulk metadata upload and dataset creation based on metadata queries. Together with community data standards, the micro-services provided by IDS trace data provenance and establish authenticity over time, supporting reproducible science.

W241: Cool Season Legumes

The RNA-Seq Based Gene Expression Atlas of a Major Food Legume Chickpea (*Cicer arietinum* L.)

Himabindu Kudapa, Vanika Garg, Annapurna Chitikineni and Rajeev K Varshney, ICRISAT, Hyderabad, India

Chickpea (*Cicer arietinum* L.) is an important food legume and is an excellent source of protein with wide range of essential amino acids to human diet. In addition, chickpea plants have root nodules with tremendous nitrogen-fixing ability. Plant growth/development are controlled by programmed expression of a suit of genes at the given time, stage and tissue. To understand how underlying genome sequence results in specific plant phenotypes at key developmental stages, information on gene expression patterns and their functions representing multiple tissues at important growth stages of plant is crucial. In this context, a comprehensive *Cicer arietinum* Gene Expression Atlas (CaGEA) was generated that provides a global view of gene expression in all major organs across the plant developmental stages covering entire life cycle of chickpea. The most drought tolerant and widely used chickpea cultivar, ICC 4958 has been used to generate RNA-Seq data from 27 samples at five important developmental stages (germination, seedling, vegetative, reproductive and senescence) of the plant. From these samples, 27 cDNA libraries were generated and sequenced, resulting in a total of 816 million raw reads. Of these, 794 million filtered reads after QC were subjected for analysis. Gene expression patterns were analyzed to better understand changes during different developmental stages. A total of 25,784 genes were identified to be transcriptionally active in one or more than one tissues representing 91% of 28,269 predicted genes in chickpea genome. CaGEA revealed 15,947 differentially expressed genes (≥2 folds) and among these 4,829 transcription factor genes were identified. In addition, 1,837 novel genes were found to be differentially expressed. In summary, CaGEA is a valuable resource for gene discovery and functional characterization to understand the systematic process of growth/development of chickpea.

W242: Cool Season Legumes

Expanding the Adaptive Range of Narrow-Leafed Lupin Using an Allelic Series at the *LanFTc1* Gene

Candy M. Taylor[1], Lars G. Kamphuis[2,3], Weilu Zhang[1], Jens D. Berger[3], Karam Singh[2,3], Wallace A. Cowling[2] and **Matthew Nelson**[2,4], (1)School of Biological Sciences, The University of Western Australia, Perth, Australia, (2)UWA Institute of Agriculture, The University of Western Australia, Perth, Australia, (3)Agriculture Flagship, CSIRO, Floreat, Australia, (4)Natural Capital and Plant Health, Royal Botanic Gardens, Kew, Ardingly, United Kingdom

Precise control of the time of flowering is critical for adapting plants to their natural or agricultural environments. The selection of a vernalisation-insensitive, early flowering mutant (*Ku*) was key to successfully adapting narrow-leafed lupin (*Lupinus angustifolius* L.) to short-season cultivation in mild winter environments in Australia and mild summer environments in northern Europe.
Recently, *Ku* was identified as *LanFTc1*, an *FT* homologue. A 1.4 kb deletion in the 5' upstream regulatory region of *LanFTc1* was associated with the loss of vernalisation requirement and constitutively high gene expression in *Ku* cultivars, leading to early flowering. Using a combination of targeted amplicon sequencing and comprehensive whole-genome re-sequencing for 42 wild and domestic narrow-leafed lupin accessions, we identified two additional deletions within the 5' upstream regulatory region of *LanFTc1*: a 1.2 kb deletion in an Isreali wild type and a 5 kb deletion in a Belarussian breeding line. The coding sequences were identical in all accessions tested.
We set out to discover if these newly discovered alleles impact on flowering and *LanFTc1* expression. We found that the accession with the 5kb deletion was vernalisation insensitive and expressed *LanFTc1* in the same manner as the *Ku* reference line. In contrast, the accession with the 1.2kb deletion was intermediate between the *Ku* and *ku* (wild type) lines, both in their vernalisation responsiveness and in their *LanFTc1* gene expression profiles. We discuss how this the growing allelic series of *LanFTc1* can be used to expand the adaptive range of narrow-leafed lupin cultivars beyond their current distrubution.

W243: Cool Season Legumes
Developing Breeders' Tools: Identifying and Visualising Genotype to Phenotype Associations in the Common Pea and its Wild Relatives
Kirstie Hetherington, Earlham Institute, Norwich, United Kingdom
Pea (*Pisum sativum*) is an important crop plant for food security. Legumes are an important family of crops and are high in protein, making them important for food and animal fodder. Furthermore, peas are useful in crop rotation as they can symbiotically fix nitrogen, reducing the need for fertiliser.
The John Innes *Pisum* Collection is a well-characterised collection of peas, holding approximately 3,600 accessions. Previous work with retrotransposon-based insertion polymorphism markers (RBIP) characterised the collection into 3 distinct groups: cultivar, landrace and wild material (Jing *et al*., 2010). We are undertaking a GWAS analysis of a number of important traits using this collection.
We generated a core collection of representatives across all 3 groups consisting of 350 accessions. Biological replicates were sown in field and glasshouse environments for phenotypic and sequence analysis. To quantify phenotypic traits, we developed a novel automated image analysis tool, to measure simple and morphological shape descriptors (SMSDs) from images of pea leaflet, pod and seed. We also collected data on seed weight and plant height
The Pea genome is large and highly repetitive and currently has no reference. To tackle this issue, we have used Genotyping-by-Sequencing (GbS) to generate a high-density marker panel of SNPs at low cost. Here we present an update on the GWAS analysis.

W244: Cool Season Legumes
Faba Genomics - TBA
Donal M. O'Sullivan, University of Reading, Reading, United Kingdom

W245: Cool Season Legumes
NorFab: Devoloping Genomic Resources for Faba Bean
Stig Uggerhøj Andersen[1], Alan H. Schulman[2], Luc L. Janss[1], Donal M. O'Sullivan[3], Bert Vandenberg[4], Frederick L. Stoddard[5], Jihad Orabi[6], Sabine Ravnskov[7], Fernando Geu Flores[8], Svend Christensen[9], Henrik Skovgård[7], Jens C. N. Knudsen[10], Birger Eriksen[11], Ahmed Jahoor[6] and Jens Stougaard[1], (1)Aarhus University, Aarhus, Denmark, (2)LUKE & University of Helsinki, Helsinki, Finland, (3)University of Reading, Reading, United Kingdom, (4)University of Saskatchewan, Saskatoon, SK, Canada, (5)University of Helsinki, Helsinki, Finland, (6)Nordic Seed, Galten, Denmark, (7)Aarhus University, Slagelse, Denmark, (8)University of Copenhagen, Frederiksberg C., Denmark, (9)University of Copenhagen, Frederiksberg C, Denmark, (10)Nordic Seed, Odder, Denmark, (11)Sejet Plant Breeding, Horsens, Denmark
Like other EU countries Denmark is a net importer of protein, mainly soybean-meal from the US and South America. The imported protein is crucial for sustaining a large livestock production and also represents an important food ingredient. The challenge is to increase domestic protein production and maintain global competitiveness while improving agricultural diversity and sustainability. We are addressing this challenge by establishing a genomics-based platform for breeding of improved *Vicia faba* (faba bean) cultivars. These will be introduced as competitive protein crops, thereby creating growth and jobs through the establishment of a major new agricultural product line.
We have initiated this work by establishing a reference gene set and gene expression atlas based on the Hedin /2 inbred line. In parallel, the same set of samples have been subjected to metabolite profiling to facilitate dissection of secondary metabolite biosynthesis pathways with a specific focus on vicine/convicine. To allow rapid progress in genomics-based breeding, we are now carrying out geno- and phenotypic characterization of 200 diverse faba lines as well as RIL and MAGIC populations segregating for agronomic traits of interest.

W246: Cool Season Legumes
Genomic Regions of *Lens* spp. associated with Flowering Time Sensitivity to Light Quality
Hai Ying Yuan[1], Larissa Ramsay[1], Chu Shin Koh[2], Ezgi Ogutcen[1], Bert Vandenberg[1] and Kirstin Bett[1], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)Global Institute for Food Security, Saskatoon, SK, Canada
Light is necessary for the adaptation of plants to specific environments. Information based on light quality induces a collective photomorphogenetic response in plants. In general, light quality with high far-red light (low Red/Far-red ratio), which is enriched in high density cultivation or weedy environments, causes early flowering in many species. However, our previous published study found that wild *Lens* species tend to be less sensitive to the light quality change compared to cultivated lentil. In the current study, we used an interspecific RIL population of lentil (*Lens culinaris* X *Lens orientalis*) with parents showing contrasting sensitivity to light quality change to gain an

understanding of the genetic basis of light responses in lentil. A high density linkage map, constructed by genotyping-by-sequencing generated SNP data, was used to discover the QTLs and potential candidate genes controlling flowering time sensitivity to light quality change in lentil. In addition, sequencing data generated by exome capture were used to compare the various genotypes from *Lens* species on these genes and their relative genomic regions were also identified on the assembled genomes of *Lens lamottei, Lens odemensis* and *Lens ervoides*. Combining these different approaches, we seek to understand the adaptation of lentil, especially that of crop wild relatives, which are now being more widely used to introduce genetic variability into the crop species. Understanding environmental responses of this staple grain legume crop will play an important role in developing genetic strategies for crop improvement in response to changes in environment.

W247: Cool Season Legumes
The Wild Side of Lens Genomes
Chu Shin Koh, Global Institute for Food Security, Saskatoon, SK, Canada

W248: Cool Season Legumes
Kasp Assays for Powdery Mildew Resistance Breeding in Pea
Yu Ma[1], **Clarice J Coyne**[2], Jamin Smitchger[1], Rebecca McGee[3] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)USDA-ARS Western Regional Plant Introduction Station, Pullman, WA, (3)USDA-ARS, Pullman, WA
Powdery mildew of pea, caused by *Erysiphe pisi* DC, is a serious production constraint to pea (*Pisum sativum* L.) production in the U.S. and elsewhere. Utilization of genetic resistance to powdery mildew using *er1* has been an effective strategy to manage this disease. This gene, *er1*, conferring powdery mildew resistance was previously cloned and sequenced, and eight kompetitive allele-specific PCR (KASP) functional markers for each resistance allele were developed. In order to identify additional pea germplasm with powdery mildew resistance, these KASP markers were used to genotype a pea collection derived from the USDA pea single-plant (PSP) collection, USDA advanced breeding lines and a set of U.S. pea cultivars. The efficacy of the KASP markers and the previously reported resistance alleles will be discussed.

W249: Crop Evolution Genomics & Future Agricultural Productivity
***Tragopogon* (Goatsbeard; Asteraceae) as a Developing Model for the Study of Polyploidy**
Douglas E. Soltis and Pamela S. Soltis, University of Florida, Gainesville, FL
Tragopogon mirus and *T. miscellus* are allotetraploids that originated ~80 years ago from the diploids *T. dubius* and *T. porrifolius* and *T. dubius* and *T. pratensis*, respectively. Given this recent ancestry and the availability of synthetic lines, these species provide excellent models for studying the extent to which fractionation and gene silencing occur shortly following polyploid formation. The allotetraploid *Tragopogon castellanus* (0.5 - 1.5 my old; diploid parents *T. crocifolius* and *T. lamottei*) is a resource for looking at the effects of polyploidization deeper in time. With this range of ages, we are able to address diverse questions of genome evolution. For example, fractionation is greater in *T. castellanus* than in naturally formed *T. mirus* and *T. miscellus* and greater in the latter two than in synthetic lines of these species. Using a custom-designed pipeline, we explored the extent of homeolog-specific expression in *T. mirus* and *T. miscellus*. In both species, approximately half of the scorable loci exhibit additive expression of the parental species, with the remaining half showing homeolog-specific expression. Of the latter, half are biased in favor of *T. dubius* and half in favor of either *T. porrifolius* (in *T. mirus*) or *T. pratensis* (in *T. miscellus*), with hypothesized gene loss less than 1% of the level of expression shifts. To help make *Tragopogon* a genetic model, we have developed new resources for this unique evolutionary model system. A draft genome sequence for the diploid, *T. dubius*, provides coverage of gene space and allows us to investigate alternative splicing and gene loss in greater detail in the polyploids. We have also made successful first steps in the use of CRISPR and in regenerating CRISPR-modified plants.

W250: Crop Evolution Genomics & Future Agricultural Productivity
The Cultural Distinction between Plant Domestication and Crop Evolution
Avi Gopher, Tel Aviv University, Tel Aviv, Israel
The pace of domestication is a "well-known" disagreement in plant domestication research in the Near East. Its long history notwithstanding, the two debated views are: 1). A protracted (millennia long) unconscious process and 2). A short event within the resolution of Neolithic chronology in the Near East, i.e., ± 50 years. The distinction between plant domestication and crop evolution which we consider major in contributing parsimony to the core area-one event domestication model was presented recently (Abbo et al. 2014, TIPS) as a means enabling a better distinction between domestication syndrome traits and in the service of a higher resolution in plant domestication research. It was based on biological considerations. Yet, a major reservoir of direct data on plant domestication originates in archaeological sites. Archaeology, has developed in the last century to turn into a high resolution discipline both by using higher resolution archaeological analyses (of sites and finds) and by employing radiometric absolute dating (e.g. C-14). These developments contributed options for the accurate dating of finds in sites relevant to plant domestication. It also contributed a potential of reconstructing how archaeological finds (materials, ideas) spread through the geography and in the case of plant domestication research in the Near East this was accompanied by studies of DNA polymorphism of relevant plant populations. Surprisingly, archaeologists (and to a certain extent archaeobotanists too) studying plant domestication of the Near East tend to undermine these achievements by lowering their resolution and blurring the quite evident cultural distinction and processes. This presentation will discuss these trends in plant domestication research in the Near East and attempt offering some explanations.

W251: Crop Evolution Genomics & Future Agricultural Productivity
The Acquisition of Rachis Brittleness in *Triticeae*
Takao Komatsuda[1], Xiaoxue Zeng[1] and Mohammad Pourkheirandish[2], (1)Institute of Crop Science, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan, (2)The University of Sydney, Faculty of Science, Plant Breeding Institute, Cobbitty, Australia
About 12,000 years ago in the Near East, humans began the transition from hunter-gathering to agriculture-based societies. Barley was a founder crop in this process, and the most important steps in its domestication were mutations in two adjacent, dominant and complementary

genes, through which grains were retained on the inflorescence at maturity, enabling effective harvesting. Independent recessive mutations in each of these genes *Btr1* and *Btr2* caused cell wall thickening in a highly specific grain 'disarticulation zone', converting the brittle floral axis (the rachis) of the wild type into a tough, non-brittle form that promoted grain retention. By tracing the evolutionary history of allelic variation in both genes, we conclude that spatially and temporally independent selections of germplasm with a non-brittle rachis were made during the domestication of barley by farmers in the southern and northern regions of the Levant, actions that made a major contribution to the emergence of early agrarian societies. However, acquisition of *Btr1* and *Btr2* has been unknown. I will present an inference on the acquisition of *Btr1* and *Btr2* in the molecular evolution of Triticeae.

W252: Crop Evolution Genomics & Future Agricultural Productivity
The *Pisum* Genus: Getting out of Pea Soup!
Jonathan Kreplak[1], Mohammed-Amin Madoui[2], Karine Labadie[2], Grégoire Aubert[3], Petr Capal[4], Philipp E. Bayer[5], Petr Novak[6], Anthony Klein[1], Krishna Kishore Gali[7], Cyril Fournier[1], Léo d'Agata[2], Ayite Kougbeadjo[1], Morgane Terezol[1], Bunyamin Taran[7], Caroline Belser[2], Françoise Jacquin[1], Marianne Chabert-Martinello[1], Marie-Christine Le Paslier[8], A. Bendahmane[9], Valerie Barbe[2], Matthieu Falque[10], Pavel Neumann[6], Jacqueline Batley[5], Clarice J Coyne[11], Tom Warkentin[7], David Edwards[5], Judith Lichtenzveig[5], Jiri Macas[6], Jaroslav Dolezel[4], Patrick Wincker[2] and **Judith Burstin**[3], (1)INRA, UMR1347 Agroécologie, Dijon, France, (2)CEA - Genoscope, Evry, France, (3)INRA UMR1347 Agroécologie, Dijon, France, (4)Institute of Experimental Botany, Olomouc, Czech Republic, (5)University of Western Australia, Perth, Australia, (6)Biology Centre CAS, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic, (7)University of Saskatchewan, Saskatoon, SK, Canada, (8)INRA, US1279 Etude du Polymorphisme des Génomes Végétaux, CEA-IG / Centre National de Génotypage, Evry, France, (9)IPS2-Paris-Sud University, Orsay, France, (10)GQE– Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France, (11)USDA ARS, Pullman, WA
Pea (*Pisum sativum* L.) has long been a model for plant genetics and is a widely grown pulse crop producing protein-rich seeds in a sustainable manner. However, many questions remain open about (sub)species relationships in the *Pisum* genus. The ongoing pea genome sequencing project and the recent genomic resources now available for pea allow exploring the factors that shaped the genetic diversity of pea and dissecting traits of interest.

W253: Crop Evolution Genomics & Future Agricultural Productivity
Harnessing 'Left behind' Drought-Adaptive Alleles for Modern Wheat Improvement
Yehoshua Saranga, R. H. Smith Institute of Plant Science & Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel
The genetic diversity in wild ancestors of crop plants has been considerably eroded throughout plant domestication, evolution under cultivation and recent plant breeding, thereby making modern crop germplasm vulnerable to various biotic and abiotic stresses. Therefore, an important task of modern breeding is to identify and reintroduce valuable 'left behind' alleles into the modern domesticated gene pool. Introduction of such alleles has been mostly employed for improving biotic stress resistances, while abiotic stress adaptations received only minor attention. Water deficit is the major environmental factor limiting crop productivity, hence developing drought-resistant crop cultivars is essential to ensure a sustainable agricultural production under the ongoing climatic change.

Wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides* [Körn.] Thell.), the progenitor of the domesticated durum (*T. turgidum* spp. *durum* (Desf.) MacKey) and bread (*T. aestivum* L.) wheats, harbors a rich allelic diversity that is valuable for future improvement. Selected drought-adaptive QTL alleles were introgressed from wild emmer wheat into modern durum and bread wheat cultivars and the resultant near isogenic lines (NILs) were tested for their drought responses. NILs introgressed with wild emmer QTL on chromosome 7A, exhibited under water-limited conditions greater grain yield, osmotic adjustment, photosynthetic capacity and root development compared with their recurrent parent. Selected NILs, carrying the 7A QTL in two genetic backgrounds, tested across 3 years under commercial field conditions, exhibited a significant advantage over their parental cultivars, particularly under drought, thus confirming the potential of this left behind QTL allele for improving drought resistance in modern wheat.

W254: Crop Evolution Genomics & Future Agricultural Productivity
Adaptation in Plant Genomes: Bigger is Different.
Jeffrey Ross-Ibarra, University of California, Davis, CA; Universitiy of California, Berkeley, CA
In this paper, we propose the functional space hypothesis, positing that mutational target size scales with genome size, impacting the number, source, and genomic location of beneficial mutations that contribute to adaptation. Preliminary evidence, mostly from *Arabidopsis* and maize, appears to bear out at least some of the predictions of our hypothesis, but clearly more data are needed before any rigorous assessment can be made. If correct, the functional space hypothesis suggests that we should expect plants with large genomes to exhibit more functional mutations outside of genes, more regulatory variation, and likely less signal of strong selective sweeps reducing diversity. These differences have implications for how we study the evolution and development of plant genomes, from where we should look for signals of adaptation to what patterns we expect adaptation to leave in genetic diversity or gene expression data. While flowering plant genomes vary across more than three orders of magnitude in size, most studies of both functional and evolutionary genomics have focused on species in the extreme small scale. Our hypothesis predicts that methods and results from these small genomes may not replicate well as we begin explore large plant genomes.

W255: Crop Genomics for Global Food Security
Taiwan Oil Millet: An Oil-Rich Orphan Cereal
Thomas Roscoe[1], Yuan-ching Tsai[2], Hshin-Ping Wu[2] and **Yue-ie C. Hsing**[2], (1)UMR DIADE, CNRS-IRD, Montpellier, France, (2)Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

Eccoilopus formosanus (Taiwan Oil Millet, TOM) is an orphan cereal endemic to Taiwan that was domesticated by the aboriginal population. TOM is a perennial C4 species, with the genome size of 2.6 Gb and chromosome number 2n=40. This orphan crop is still cultivated in mountain communities at altitudes up to 2000m. The seed of Taiwan oil millet is characterised by an exceptionally large embryo and consequently possesses substantial quantities of triacylglycerol and storage protein in addition to the starchy endosperm. This species secretes oil in the form of monoacylglycerols from the panicle and reproductive structures and exudes copious amounts of a pure fatty acid wax from pore-like structures on the leaf sheath. Transcriptomic analyses in vegetative tissues have identified genes associated with the regulation and biosynthesis of cuticular lipids and waxes. Taiwan Oil Millet is unique in protein-calorie rich seed and as an energy-rich biomass for forage or a source of unusual cuticular waxes of uniform composition for use as industrial feedstocks.

W256: Crop Genomics for Global Food Security

Widening the Genepool and Identifying Genes Controlling Key Traits in Rice and Wheat

Robert J. Henry, University of Queensland/QAAFI, Brisbane, Australia

Food security can be advanced by capturing more useful diversity in crop improvement and by better understanding the molecular basis of key traits that limit the rate of genetic gain in breeding. Recent genome sequencing research has identified wild populations representing significant new diversity in the primary gene pool of rice. This provides new sources of resistance to pests and disease, diversity to allow adaptation to climate change and novel grain qualities with potential consumer appeal and health benefits. Analysis of the transcriptome of the developing grain of diverse wheat germplasm has identified the genetic control of traits such as milling performance, hardness, and end-use quality (bread and chapatti). Genetic improvements can generate more grain per hectare more flour per tonne of wheat and more end-product per tonne of flour. These studies have demonstrated the great diversity of response to heat stress in the wheat gene pool. When combined these developments offer large improvements in food security.

W257: Crop Genomics for Global Food Security

Developmental and Physiological Dissection of QTLs in Wheat NAM Populations

Stefano Bencivenga, Luzie U. Wingen, Scott Boden and Simon Griffiths, John Innes Centre, Norwich, United Kingdom

The control of timing and duration for phases of wheat inflorescence development, and its interaction with the environment, is an important tool to maximize harvest index. For this purpose, we have selected and crossed wheat varieties with different origin and traits to generate populations with a high degree of diversity. In particular, varieties characterized by earliness, including CIMCOG15, MISR1, Super 152, Pfau, Waxing, Baj, and high yielding lines, such as CIMCOG49, CIMCOG47 and GARCIA, were crossed with the UK variety Paragon. F4 and F5 generations were grown in the UK , Mexico and Spain where they have been scored for several phenotypic traits, namely days until GS31, heading, bolting, anthesis, stem height, grain yield, grain dimension and thousand grain weight. Plants were genotyped using the 35k breeders array and the results elaborated through MST mapping. With the obtained phenotypic and genotypic data, we performed QTL analysis to associate genomic regions to specific traits.

As a validation of reliability for data collection and analysis methods, genes known to be involved in the regulation of the flowering time pathway, such as VRN1, Rht1, Vrn-A1, Vrn-B1, Vrn-D1, Ppd-D1, showed strong correlation with traits of time to heading, bolting and anthesis in different parental crosses.

We then searched for genomic regions, not comprising any known gene that correlated with a specific developmental stage. Among all the candidate locus identified, a genomic regions located on chromosome 7D of plants coming from a Paragon X Baj cross was selected because of its strong correlation with time of heading in both Mexico and UK dataset. This genomic region comprises the wheat homologue of the important plant florigen FT, and we discovered that in Baj the FT region is duplicated. The copy number variation can be linked with the booting and heading time variation between Paragon and Baj in UK and Mexico. As consequence plants carrying the Baj 7D region have a higher level of FT that make them more prone to flower.

Moreover the wheat populations generated can be used to improve yield. Lines more suitable for the different environment were selected and grown for another season in the three different locations. We measured flowering time and biomass, as well as harvest index. We discovered new QTLs related to yield for the three different environments.

Finally we have generated the new populations WeebilxCIMCOG03, WeebilxCIMCOG26, WeebilxCIMCOG32, WeebilxCIMCOG49 and WeebilxCIMCOG53 and created genetic maps.

This work not only open new lines of research for genes that can influence flowering time, but also makes available populations with a high degree of variability as powerful tool to be used around the world for genetic phenotypic association.

W258: Crop Genomics for Global Food Security

Genome-Wide Analysis of NBS-LRR Genes in the Brassicaceae and Applications for Breeding

Fabian C. Inturrisi[1], Soodeh Tirnaz[1], Philipp E. Bayer[1], Ting Xiang Neik[1], Hua Yang[1,2], Aria Dolatabadian[1], Fangning Zhang[1], Anita Severn-Ellis[1], Dhwani A. Patel[1], Aneeta Pradhan[1], HueyTyng Lee[1], David Edwards[1] and Jacqueline Batley[1], (1)University of Western Australia, Perth, Australia, (2)University of Queensland, Brisbane, Australia

The Brassicaceae family contains some of the world's most important economic and agronomic crops, which are utilised as edible and industrial oilseeds (e.g *Brassica napus, B. juncea*) and vegetables (e.g *Brassica oleracea, Raphanus raphanistrum*), along with the scientific model plant *Arabidopsis thaliana* and highly diverse wild species. Pathogens, such as *Leptosphaeria maculans* (causal agent of Blackleg) and *Sclerotinia sclerotiorum* (causal agent of Sclerotinia stem rot), severely affect production of important crop species from the Brassicaceae. Plant genomes harbour resistance (R) genes, which play an important role in plant immunity, where nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes are the most common type of R gene. In this study, NBS-LRR genes were identified and classified in 35 wild and cultivated Brassicaceae species to determine the phylogenetic relationship and better understand the long-term evolutionary history of these R genes. The variation of classes of NBS-LRR genes was observed to vary greatly within and between species, and among different genome assemblies of the same species. The expansion and loss of NSB-LRR genes was also observed among the Brassicaceae. The change in climate will have an,

as yet, unknown impact on pathogen populations, with the potential for disease pressure to increase. This analysis provides a valuable resource for the identification of R genes for enhanced crop protection by developing elite resistant cultivars where by the functionality of R genes is studied against particular diseases and bred into commercial cultivar by marker-assisted breeding.

W259: Crop Genomics for Global Food Security
High Throughput Phenotyping of Crop Plants in Field Trials
Mitchell Tuinstra, Purdue University, WEST LAFAYETTE, IN

Bottlenecks in our ability to collect accurate, high-resolution, phenotype data on crops like sorghum limit how efficiently we can combine plant characteristics with genomic information in developing superior cultivars. Field-based phenotyping of crops for above-ground traits occurs at the plant and canopy scale of biological organization. These traits are typically measured or expressed at the plot, management zone, or field level. New sensors and sensor platforms, novel georeferencing techniques, and sophisticated image and data analysis methods (e.g., feature extraction, image segmentation) are being implemented to quantify variation in subplot or plant-level traits. These measurements provide insights into research plot and field quality, field equipment performance, genotype productivity, physiological plasticity, and spatial variability. Such information contributes to field crop breeding and management, precision agriculture, and equipment manufacturing communities. Challenges from phenotyping at the individual plant level using traditional and emerging techniques will be discussed. Agronomic performance and remote sensing data matched with genotypic data enables trait dissection and optimization of sorghum for biomass and energy yield for transportation fuel.

W260: Crop Genomics for Global Food Security
TBA
Maria Fátima Grossi de Sa, Embrapa, Brasilia, Brazil; Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

W261: CSSA: Translational Genomics
Integration of High-Throughput Phenotyping to Genomic Prediction Models to Improve Yield Prediction in Wheat
Jesse Poland, Department of Plant Pathology, Kansas State University, Manhattan, KS

W262: CSSA: Translational Genomics
Establishing a Translational Genomics Infrastructure for Crop Plants
Liya Wang[1], Zhenyuan Lu[1], Peter Van Buren[1], Xioafei Wang[1], Kapeel Chougule[1], Joshua Stein[1] and **Doreen Ware**[2], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)USDA/ARS - Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

With the advance of next-generation sequencing, it is feasible to develop genomic knowledge for any crops and translate this information into genetic improvement. To achieve this a starting point is an accurate sequence representation of the a genome or even a population of genomes. Once the reference is generated it is followed by functional annotation efforts (e.g. ENCODE) that aims to distinguish 'real' functional regions from random DNA. To support the storage, transfer, analysis, and management of a large amount of data, maintain data integrity, make data Findable, Accessible, Interoperable and Reproducible (FAIR), scalable requires cyber infrastructure, including robust hardware, software and standards are needed. In this talk we will describe our recent work motivated by the MaizeCode (an initial ENCODE effort for maize) datasets, and FAIR, to build a scalable data management and analysis platform, called SciApps.org, to ensure the processing of over 400 sequencing assays through a uniform and automated bioinformatics pipelines. Utilizing standardized pipelines supports consistent and high-quality datasets, which can be utilized to support integration across different assays, application of statistical methods and machine learning algorithms, and to support interpretation of the functional regions. In addition to the Encode assays, we have also developed processes to support gene structural annotations of several crops using XSEDE Jetstream.

W263: CSSA: Translational Genomics
Soybean Genomics Resources for Sustainable Crop Production
Babu Valliyodan, University of Missouri, Columbia, MO

W264: CSSA: Translational Genomics
Characterization of the *Gmsacpd* Gene Family in Soybean uncovers a New Role in Plant Development and Structure
Khalid Meksem, Southern Illinois University Carbondale, Carbondale, IL and Naoufal Lakhssassi, Vincent Colantonio, Flowers ND, Zhou Zhou, Henry J, Liu Shiming, arbottam Piya, Gunvant Patil , Azam Baharlouei , My Abdelmajid Kassem , Henry Nguyen , Tarek Hewezi and Khalid Meksem

Stearoyl-acyl carrier protein desaturase (SACPD) has been reported to control the accumulation of seed stearic acid in soybean. Using a forward genetics approach, one nonsense and four missense *Gmsacpd-c* mutants were identified to have high levels of seed, nodule, and leaf stearic acid content. Soybeans carrying *Gmsacpd-c* mutations at conserved residues showed the highest stearic acid content, and these mutations were found to have deleterious effects on nodule development and function. Interestingly, mutations at nonconserved residues show an increase in stearic acid content yet retain healthy nodules. Thus, random mutagenesis and mutational analysis allows for the achievement of high seed stearic acid content with no associated negative agronomic characteristics. Finally, we report that *Gmsacpd-c* mutations cause an increase in leaf stearic acid content and an alteration of leaf structure and morphology in addition to differences in nitrogen-fixing nodule structure. *GmSACPD-D* has been reported to be a pseudo-gene, however, three missense Gmsacpd-d mutants were identified by TILLING, yet all mutant lines contained the same level of seed stearic acid as the wild-type Forrest. Moreover, GmSACPD-C showed cytoplasmic localization, whereas GmSACPD-D showed a mitochondrial-like localization patterns contradicting previous bioinformatics predictions. Surprisingly, mutations on GmSACPD-D were lethal for the soybean plants. Our results demonstrate that GmSACPD-D member has been

neofunctionalized to acquire a possible role in cyst nematode infection, in seed germination, to be temporary regulated in embryo development, and to exhibit an expression confined to ovule.

W265: CSSA: Translational Genomics

Implications of Homeologous Gene Interactions for Breeding Allopolyploid Crops.

Nicholas Santantonio[1], Jean-Luc Jannink[2] and Mark E Sorrells[1], (1)Cornell University, Ithaca, NY, (2)USDA-ARS / Cornell University, Ithaca, NY

The sub-genomes of an allopolyploid will each contain complete, yet evolutionary divergent, sets of genes. With the availability of affordable genome-wide markers, breeders of allopolyploids now have the opportunity to manipulate individual sub-genomes and investigate interactions of homeoalleles across sub-genomes. We present theory and a statistical framework for partitioning genetic variance and predicting breeding values for each sub-genome and their inter-genomic interactions. Using an allohexaploid wheat breeding population for demonstration, sub-genome main effects and interactions were fit using multi-kernel mixed models for variance component estimation and genomic prediction. Strictly modeling inter-genomic interactions resulted in equivalent increases in genomic prediction accuracy as modeling all pairwise marker interactions. Using the IWGSC RefSeq v1.0 wheat genome sequence, 18,184 triplicate and 5,612 duplicate homeoallelic gene sets were identified and anchored to the nearest GBS marker, forming 10,172 unique sets of homeologous markers. Functional homeologous marker interactions for each homeoallelic marker set were used to predict whole genome breeding values, as well as estimate homeologous main and interaction effects. Using gain in genomic prediction accuracy as a proxy for importance of marker interactions, we show that homeologous marker interactions can explain up to 66% of the additional genetic signal from the additive model. Negative relationships observed between homeologous marker main effects and interaction effects points to a pattern indicative of homeoallelic subfunctionalization. Thus, we provide new tools for breeders of allopolyploid crops to characterize the genetic architecture of existing populations, determine breeding goals, and develop strategies for selection of sub-genome additive effects and inter-genomic epistasis.

W266: CSSA: Translational Genomics

Delivering Improved Genetic Gain to Small Holders in Sub-Saharan Africa

Jeffrey D. Ehlers, Bill & Melinda Gates Foundation, Seattle, WA

In sub-Saharan Africa food production has just kept up with the rapid population growth in the region over the past four decades. However, in contrast with other parts of the world, food production increases have been accomplished by bringing more land under cultivation, with very limited productivity gains (Ethiopia and Rwanda are notable exceptions). It will be difficult and undesirable to continue such an extensification approach to accommodate the food requirements given predicted sustained high rates of population growth coming in the next few decades. Thus, it is imperative for African countries to boost productivity of key staple crops. Based on experience in Asia and other developing countries, boosting small farm productivity is likely to help alleviate extreme poverty in rural areas and kick start agricultural transformation. Improving crop improvement systems, and the delivery of the improved varieties, preferably combined with greater use of inputs and production knowledge, can play a key role in boosting small farm productivity and reducing production risk. At the current time, there is a substantial opportunity to more than double the genetic gains of breeding programs serving small holders through the adoption of modern tools, best practice and new technologies. Seed systems serving smallholder farmers need to be designed and implemented to promote rapid varietal turnover so that the genetic gains generated by the breeding programs can be realized by farmers. This presentation will describe the work of several key initiatives supported by the Bill and Melinda Gates Foundation designed to increase the rate of genetic gain within the crop breeding programs of several CGIAR institutes and key national programs in sub-Saharan Africa. These include development and deployment of information management systems Breeding Management System, electronic data capture, systematic program evaluations using the Breeding Program Assessment Tool BPAT Assessment Tool, centralized high-throughput genotyping facilities

W267: Cucurbit Genomics

Indian Germplasm is a Center of Melon Diversification According Genotyping-by-Sequencing Analysis

Maria José Gonzalo, IBMCP (CSIC-UPV), Valencia, Spain, Aurora Díaz Bermúdez, CITA, Zaragoza, Spain, Narinder Dhillon, AVRDC - Regional Office for East and Southeast Asia, Nakhon Pathom, Taiwan, Umesh K. Reddy, Department of Biology, West Virginia State University, Institute, WV, Maria Belen Pico, COMAV-Universidad Politecnica Valencia, VALENCIA, Spain and **Antonio J. Monforte**, IBMCP, CSIC-UPV, Valencia, Spain

Previous SSR-based genetic diversity analysis supported that Oriental and Occidental melon varieties arose by divergent selection from Indian ancient varieties. Samples from interesting Indian germplasm collections, a broad range of Occidental, Oriental cultivars and some African accessions were analyzed by Genotyping-by-Sequencing. A total of 6169 informative SNPs were obtained, showing a high level of genetic diversity (π=0.32, He=0.33) in the worldwide melon germplasm. Tajima's *D* was 2.63 which may reflect balancing selection due to the fixation of alternative alleles in different world regions. Linkage disequilibrium among SNPs was very low, within 1 kb, even in centromeric regions. PCA displayed a high genetic structure; groups including African accessions, conomon, dudaim, cantaloup, inodorus and Indian cultivars were clearly defined, only a few accessions did not clustered within these groups. Genetic variability among populations was 32% and 68 % within populations. Indian germplasm cluster was located in the center of the PCA plot, confirming the original hypothesis. Genomic regions involved in the genetic differentiation between Indian, conomon, inodorus and cantaloups groups were identified through the genome by F_{st} and H_e analysis. Interestingly, a few large haplotypes blocks were found to be specific of some groups, being also candidate regions for harboring genes selected by farmers to develop the present cultivated types.

W268: Cucurbit Genomics

Comparative Population Genomics Reveals the Evolution of Fruit Quality Traits during Watermelon Domestication

Shaogui Guo[1], Honghe Sun[1], Xin Wang[2], Yi Ren[1], Jie Zhang[1], Guoyi Gong[1], Haiying Zhang[1], Zhangjun Fei[2] and Yong Xu[1], (1)National Engineering Research Center for Vegetables, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, (2)Boyce Thompson Institute, Cornell University, Ithaca, NY

Watermelon fruit quality evolved dramatically from the ancestor to the cultivated variety. 323 wild, semi-wild and cultivated watermelon accessions were resequenced and analyzed. Watermelon taxa were clarified and the process of watermelon evolution was recurred based on the whole genome comparative analysis. Three steps were identified during the process from the ancestor to the cultivated variety. In first, the fruit sink strength and fruit size were evolved from wild watermelon *Citrullus colocynthis* to semi-wild watermelon *Citrullus mucosospermus* due to the selection of the genome regions containing photosynthate unloading and fruit expansion related genes. During the second step from the recent progenitor *Citrullus mucosospermus* to watermelon landrace of *Citrullus lanatus*, the edible watermelon fruit were domesticated due to the selection of the genome regions containing the taste related transcription factor and sugar transporters. During the third step from watermelon landrace to the dessert watermelon *Citrullus lanatus*, the desirable fruit were improved further due to the selection of the genome regions containing different sugar transporters. Comprehensive analysis of population differentiation, GWAS and QTL mapping indicated that the fruit sweetness, fruit bitterness, fruit flesh color related genes were involved in the evolution, domestication and improvement of fruit quality traits. The data provide a valuable resource for the further analysis of watermelon fruit development and ripening, and to facilitate future breeding of *Citrullus* genus.

W269: Cucurbit Genomics

Genetic Architecture of Downy Mildew Resistance in Cucumber

Yuhui Wang[1], Kyle VandenLangenberg[2], Todd Wehner[2] and **Yiqun Weng**[3], (1)Horticulture Department, University of Wisconsin, Madison, WI, (2)Horticultural Science, North Carolina State University, Raleigh, NC, (3)USDA ARS/ Universtiy of Wisconsin, Madison, WI

Downy mildew (DM) caused by the obligate oomycete *Pseudoperonospora cubensis* is the most devastating fungal disease to cucumber production. The molecular mechanism of DM resistance in cucumber is not well understood. We conducted QTL mapping for DM resistances in four cucumber lines including WI7120 and PI 197088 that are highly resistant to post-2004 DM strain(s) as well as Gy14 and WI2757 that offered effective protection to DM infection for over 50 years until 2004. Four QTL, *dm2.1*, *dm4.1*, *dm5.2* and *dm6.1* were detected for WI7120-derived resistance which together could explain 62-76% phenotypic variance; *dm4.1* and *dm5.2* were major-effect QTL. In PI 197088, 11 QTL were identified accounting for 73.5% total phenotypic variance, among which, three (*dm5.1*, *dm5.2* and *dm5.3*) and five (*dm2.1*, *dm3.1*, *dm3.2*, *dm4.1* and *dm6.1*) were major- and minor-effect contributing QTL, respectively whereas *dm1.1*, *dm2.1*, and *dm6.2* conferred susceptibility. DM resistance to the post-2004 strain in Gy14 and WI2757 was controlled by one (*dm1*) and two (*dm1*, *dm5.2*) QTL on chromosome 5, respectively. Genome-wide association analysis identified consistent or novel DM resistance loci in natural cucumber populations. Fine genetic mapping of *dm4.1*, *dm5.2*, and *dm1* further identified potential candidate genes for *dm4.1*, *dm5.2* and *dm1*. The work highlighted the power of reliable phenotypic data collection, next-gen high throughout sequencing based genotyping and marker-assisted NIL development in improving the power of QTL detection. The genetic architecture of DM resistance in different resistance sources and their potential in cucumber breeding for DM resistance will be discussed.

W270: Cucurbit Genomics

Cucurbita Genome Sequences Provide Insights into Polyploid Genome Evolution and Heterosis in Interspecific Hybrid

Shan Wu[1], Honghe Sun[2], Guoyu Zhang[2], Chen Jiao[1], Shaogui Guo[2], Yi Ren[2], Jie Zhang[2], Haiying Zhang[2], Guoyi Gong[2], Zhangcai Jia[2], Fan Zhang[2], Jiaxing Tian[2], William Lucas[3], Jeffrey J. Doyle[4], Haizhen Li[2], Zhangjun Fei[1,5] and Yong Xu[2], (1)Boyce Thompson Institute, Cornell University, Ithaca, NY, (2)National Engineering Research Center for Vegetables, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, (3)Deptartment of Plant Biology, Davis, CA, (4)Plant Biology Department, Cornell University, Ithaca, NY, (5)USDA-ARS Robert W. Holley Center for Agriculture and Health, Ithaca, NY

The *Cucurbita* genus in the Cucurbitaceae family contains several economically important species and interspecific hybrids between *C. maxima* and *C. moschata* are widely used as rootstocks for other cucurbit crops. We *de novo* assembled the genomes of *C. maxima* and *C. moschata*, which provided evidence supporting an allotetraploidization event in the *Cucurbita* lineage. We partitioned the genome into two homoeologous subgenomes based on different genetic distances to the species in the Benincaseae tribe, including melon, cucumber and watermelon. We estimate that the two diploid progenitors of *Cucurbita* successively diverged from Benincaseae around 31 and 26 million years ago (Mya), and the allotetraploidization happened earlier than 3 Mya, when *C. maxima* and *C. moschata* diverged. The subgenomes have largely maintained the chromosome structures of their diploid progenitors. During evolution, the two subgenomes have retained similar numbers of genes, and neither subgenome is globally dominant in gene expression. Such long-term karyotype stability and unbiased fractionation has not been commonly observed in other allopolyploid plants. These two high-quality genome sequences allowed us to detect transgressive gene expression changes in the *C. maxima* × *C. moschata* interspecific F_1 hybrid 'Shintosa' correlated with heterosis in important agronomic traits such as carotenoid content in fruits.

W271: Cucurbit Genomics

Characterization and QTL Mapping of Ethylene Production and Climacteric Ripening in a Melon RIL Population

Lara Pereira, Valentino Ruggieri, Jason M. Argyris, Marta Pujol and Jordi Garcia-Mas, IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Bellaterra, Spain

Fruit ripening is a complex process that includes mainly metabolic changes in the fruit to be attractive for the consumer, allowing seed dispersal. Climacteric ripening is defined by a peak of the plant hormone ethylene at the onset of ripening, triggering physiological responses as aroma production, abscission and chlorophyll degradation. Melon (*Cucumis melo* L.) has been proposed as an alternative model to understand fruit ripening due to the co-existence of climacteric and non-climacteric varieties, allowing the use of genetic studies. A recombinant inbred line population of 91 individuals was developed from a cross between the climacteric cantaloupensis type "Védrantais" and the non-climacteric inodorus type "Piel de Sapo". To our knowledge, this is the first population created using two commercial cultivars that belong to different botanical groups. Five replicates of the population were phenotyped for climacteric ripening, consisting of the characterization and quantification of the ethylene peak during the ripening process and the effects associated, such as aroma production, abscission, color change

and flesh softening. We used a genotyping-by-sequencing strategy to construct a genetic map with 4,888 markers and performed a QTL mapping analysis. A major QTL for maximum ethylene production and most of the downstream effects, explaining 30% of the variance, was detected in chromosome VIII, in a 500-kb interval. The combination of genetic and genomic tools will allow to identify a candidate gene underlying this QTL. Our work will contribute to understand the genetic basis of climacteric ripening in melon and to apply this knowledge in breeding programs.

W272: Cucurbit Genomics

Genetic Mapping of a Major Co-Dominant QTL Associated with β-Carotene Accumulation in Watermelon

Lea Vexler[1,2], Sandra E. Branham[3], Ayala Meir[1], Patrick Wechter[3], Galil Tzuri[1], Zohar Frieman[1], Amnon Levi[3], Yaakov (Kobi) Tadmor[1] and **Amit Gur**[1], (1)Newe Yaar Research Center, ARO, Ramat Yishay, Israel, (2)The Hebrew University of Jerusalem, Rehovot, Israel, (3)USDA-ARS, Charleston, SC

The common flesh color of commercially grown watermelon is red due to the accumulation of lycopene. However, natural variation in carotenoid composition that exists among heirloom and exotic accessions, results in a wide spectrum of flesh colors. We previously identified a unique orange-flesh watermelon accession (NY0016) that accumulates mainly β-carotene and no lycopene. We hypothesized this unique accession could serve as a viable source for increasing provitamin A content in watermelon. Here we characterize the mode of inheritance and genetic architecture of this trait. Analysis of test crosses of NY0016 with yellow and red fruited lines indicated a co-dominant mode of action as F_1 fruits exhibited a combination of carotenoid profiles from both parents. We combined visual color phenotyping with genotyping-by-sequencing of an $F_{2:3}$ population from a cross of NY0016 by a yellow fruited line, to map a major locus on chromosome 1, associated with β-carotene accumulation in watermelon fruit. The QTL position and allelic effect was validated through analysis of additional segregating populations (including crosses with red-fleshed lines).Through deep sequencing of F3 bulks from the mapping population; we were recently able to narrow the QTL interval to less than 1Mbp with 50 annotated genes. This study is a step towards identification of a major gene involved in carotenoid biosynthesis and accumulation in watermelon. The co-dominant inheritance of β-carotene provides opportunities to develop, through marker-assisted breeding, β-carotene-enriched watermelon hybrids.

W273: Cucurbit Genomics

Cucurbit Genomics Database for Cucurbit Genomics, Genetics and Breeding

Zhangjun Fei, Boyce Thompson Institute, Cornell University, Ithaca, NY

The Cucurbitaceae family (cucurbit) includes several economically important crops, such as melon, cucumber, watermelon, pumpkin, squash, and gourd. During the past several years, genomic and genetic data have been rapidly accumulated for cucurbits. We have been developing the Cucurbit Genomics Database (http://cucurbitgenomics.org/) to store, mine, analyze, and disseminate the large-scale cucurbit genomic and genetic datasets in an efficient way and to provide a center portal for the cucurbit research and breeding community. The database currently contains all available genome and EST sequences, genetic maps, and transcriptome profiling for cucurbit species, as well as sequence annotations and biochemical pathways. A set of analysis and visualization tools and user-friendly query interfaces have been implemented in the database. Future development of the database will be discussed.

W274: CyVerse Education: Scaling Genomics and Data Science for the Biology Classroom

Next Generation RNA Sequencing: Facilitating Undergraduate Research

Carolina Sempertegui, Truman State University, Kirksville, MO

Professional development opportunities that allow faculty from primarily undergraduate institutions to collect data and learn about new technologies, combined with the availability of free analysis software, constitute excellent sources of research for students in these institutions. In addition, the use of model organisms that are easy and safe to manipulate for students, as well as simple experimental design, facilitate the learning process of students and encourages the development of ideas for research projects. After conducting ultra violet (UV) exposure experiments using *Saccharomyces cerevisiae* and collecting transcriptomic data, it was possible to observe that there was differential expression of genes involved in multiple cell functions. Students at Truman State University who have had the opportunity of observing these preliminary data took the initiative to construct research questions that were developed into research projects. These experiences have been of benefit to these students because they had the opportunity of learning first hand how to analyze these data and were able to acquire experience and training in the practice of the scientific method, laboratory techniques and analytical skills. These experiences have resulted in seminar presentations in local symposia and poster presentations in regional meetings.

W275: CyVerse Education: Scaling Genomics and Data Science for the Biology Classroom

Integrating Bioinformatics Tools in the High School Classroom

Victoria Hernandez, William Floyd High School, Mastic Beach, NY

As bioinformatics technology expands, computing skills needed to be college and career ready are becoming more rigorous. Bioinformatics tools developed by the Cold Spring Harbor Laboratory's DNA Learning Center (CSHL DNALC) through the NSF-funded CyVerse organization, including the genome analysis tool DNA Subway, are easily accessible for high school students to use in their research. These tools enhance student ability to go beyond the classroom and explore data extensively. As part of the DNA barcoding high school research programs run by the CSHL DNALC, the DNA Subway Blue Line was utilized by students in grades 9-12 for the purpose of analyzing DNA sequence data obtained from organisms collected throughout Long Island and New York City. Experienced students from the DNA barcoding programs were given the opportunity to develop microbiome analysis research projects. To conduct these projects, which included analysis of microbiomes from soil, aquatic plants, and vectors for disease, among other projects, students learned Python coding through the use of Jupyter Notebooks and performed QIIME analyses on their obtained microbial sequence data. The aims of the student microbiome research projects from William Floyd High School were as follows: a) Identify if embalming techniques impacted microbiomes in cemeteries from different time periods; b) Identify if a variation of pathogens present is evident within *Amblyomma cajennense*, a group of alpha-gal allergy associated ticks also known as or "Cayenne ticks"; and c) Observe if there is a relationship between microorganisms in sparsely populated *Zostera marina*

(eelgrass) beds, densely populated beds, and the microorganisms living within the surrounding sediments. Using QIIME analyses, the following conclusions were reached: a) Civil War cemetery soil exhibited high beta diversity in comparison to Revolutionary War and modern cemetery soils; b) collected *Amblyomma cajennense*, contained pathogens associated with Rocky Mountain Spotted Fever, Legionnaires, and Ehrlichiosis; c) a group of collected *Zostera marina* and associated sediment samples contained stramenopiles. *Labyrinthula zosterae*, is a stramenopile associated with an endophytic infection known as "eelgrass wasting disease." Students are currently investigating this population of eelgrass further to confirm if the disease is present. The results of this project are important because *Zostera marina* is a foundational marine species that increases habitat complexity and vastly impacts the biodiversity of coastal ecosystems and water quality. These eelgrass populations appear to be decreasing off of Long Island and identifying if a disease is present will aid in the replanting and restoration of this foundational species.

W276: CyVerse Education: Scaling Genomics and Data Science for the Biology Classroom
CURE-All: Large Scale Implementation of Authentic DNA Barcoding Research into a Freshman Biology Curriculum
Oliver Hyman, Elizabeth Doyle, Andrea Pesce and Ray A. Enke, James Madison University, Harrisonburg, VA
There is mounting evidence supporting the benefits that Undergraduate Research Experiences (UREs) can provide to STEM students resulting in a rising call to provide UREs to as many students as possible as early as possible in their college careers. Unfortunately, many traditional models of UREs lack the ability to scale to meet the needs of large course enrollments, inadequate student experience, limited instructor expertise, budgetary limitations, and departmental inertia. To help address these issues, we have designed a highly scalable and transferrable two semester Course-based Undergraduate Research Experience (CURE) based on DNA barcoding. Each year >500 freshman students at James Madison University use this DNA barcoding CURE to engage in authentic research and gain competencies, content knowledge, and skills from a variety of fields including ecology, molecular biology, and bioinformatics. This CURE was implemented as part of a complete course redesign to a large enrollment introductory course sequence in the biology department of a large, four-year, masters granting institution following the recommendations of Vision and Change. In this talk we outline the planning process, content, logistics, and implementation of this lab. To our knowledge this is one of the largest-enrollment CUREs currently being implemented in the United States, and serves as a model to demonstrate that all freshman STEM students, even in large enrollment courses, can participate in authentic research.

W277: CyVerse Education: Scaling Genomics and Data Science for the Biology Classroom
Building Capacity for Genomics Data Analysis in Resource-Limited Environments
Nicola Anthony, University of New Orleans, New Orleans, LA
DNA sequencing technologies have advanced in such a way that individual researchers and research teams can generate large genomic and transcriptomic datasets quickly and relatively inexpensively. This increased accessibility of genome-scale data has made significant contributions to many areas of plant and animal biology, including the evolutionary genomics of non-model organisms. However, the most common obstacle to performing these studies is the bioinformatics skills needed to make sense of these large datasets and diverse computational approaches needed to analyze the data. Many smaller research institutions, particularly in the countries where the actual data originates, may currently lack adequate local cyber-infrastructure and the necessary expertise to train the next generation of students in the computational biology skills needed to embrace genomics-era datasets. Here we describe lessons that we have learned in integrating ecological genomics into the undergraduate classroom as well as our experience leveraging existing training and cloud computing facilities through the CyVerse, Software/Data Carpentry, and VirtualBox platforms. Lastly, we also describe our collective experience of organizing bioinformatics workshops in central Africa and potential avenues to grow greater networking opportunities and genomics research capacity in sub-Saharan Africa.

W278: CyVerse - Software, Tools, and Services for Data-Driven Discovery
CyVerse: Looking Towards the Future
Parker Antin, University of Arizona, Tucson, AZ

W279: CyVerse - Software, Tools, and Services for Data-Driven Discovery
iVirus: Facilitating New Insights in Viral Ecology with Software and Community Data Sets Imbedded in a Cyberinfrastructure
Benjamin Bolduc[1], Ken Youens-Clark[2], Simon Roux[3], Bonnie L. Hurwitz[2] and Matthew B Sullivan[1], (1)The Ohio State University, Columbus, OH, (2)University of Arizona, Tucson, AZ, (3)Joint Genome Institute, Walnut Creek, CA
Microbes affect nutrient and energy transformations throughout the world's ecosystems, yet they do so under viral constraints. In complex communities, viral metagenome (virome) sequencing is transforming our ability to quantify viral diversity and impacts. Although some bottlenecks, for example, few reference genomes and nonquantitative viromics, have been overcome, the void of centralized data sets and specialized tools now prevents viromics from being broadly applied to answer fundamental ecological questions. Here we present iVirus, a community resource that leverages the CyVerse cyberinfrastructure to provide access to viromic tools and data sets. Through the CyVerse Discovery Environment, users can interrogate these data sets using existing analytical tools (software applications known as 'Apps') for assembly, open reading frame prediction and annotation, as well as several new Apps specifically developed for analyzing viromes. Because Apps are web based and powered by CyVerse supercomputing resources, they enable scalable analyses for a broad user base. This growing iVirus resource should help researchers utilize viromics as yet another tool to elucidate viral roles in nature.

W280: CyVerse - Software, Tools, and Services for Data-Driven Discovery
A High-Throughput Image Analysis Pipeline to Quantify Carrot Shoot and Root Morphology
Sarah D. Turner[1], Shelby Ellison[2], Douglas Senalik[3], Philipp W. Simon[4], Edgar Spalding[5] and Nathan Miller[5], (1)Division of Biological Sciences, University of Missouri, Columbia, MO, (2)USDA-ARS, Madison, WI, (3)USDA, ARS and University of

Wisconsin, Madison, Madison, WI, (4)USDA-Agricultural Research Service, Vegetable Crops Unit, University of Wisconsin-Madison, Madison, WI, (5)University of Wisconsin-Madison, Madison, WI
Carrot (*Daucus carota* L.) is a globally important root crop that is valued for its culinary and nutritional attributes. Selective breeding in carrot has focused primarily on characteristics of the storage root, such as size, length, shape, and uniformity, but we still lack detailed information about the genetic control of these traits and how they are influenced by shoot architecture (e.g. height, width, and biomass). To better characterize carrot morphology, we developed an automated, high-throughput image analysis pipeline that quantifies root and shoot architectural features. This pipeline is implemented on CyVerse as a computational workflow in which users can upload images of whole carrot plants, run the automated algorithm, and retrieve quantitative data for downstream analyses. We validated the workflow using 917 carrot images, which included individuals from a diallel mating design and an F_2 mapping population. Strong correlations were observed between manual and algorithm measurements, ranging from 0.77 for leaf number to to 0.93 for shoot biomass. Additionally, we captured components of shoot and root shape that cannot be measured manually. The availability of this pipeline on CyVerse establishes a resource for breeders and geneticists to efficiently and reliably catalog phenotypic diversity in large populations, thereby enabling novel genetic insights and expanding opportunities for future research.

W281: CyVerse - Software, Tools, and Services for Data-Driven Discovery
TIPS: A System for Automated Image-Based Phenotyping of Maize Tassels
Joseph Gage[1], Nathan Miller[1], Edgar Spalding[1], Shawn Kaeppler[2] and Natalia de Leon[1], (1)University of Wisconsin-Madison, Madison, WI, (2)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI
The maize male inflorescence (tassel) produces pollen necessary for reproduction and commercial grain production of maize. Mapping studies to understand the genetic basis of tassel size and morphology require numerous, precise phenotypes. Tassels are fragile and deform easily after removal from the plant, requiring timely measurement of any shape characteristics that cannot be retained during storage. Lack of any standardized tassel imaging software necessitated custom development of the Tassel Image-based Phenotyping System (TIPS) – a platform for imaging tassels in the field immediately following removal from the plant. TIPS leverages the CyVerse Discovery Environment and Data Store, as well as distributed computing through HTCondor and the Open Science Grid, to quantify morphological traits from profile images of freshly harvested tassels acquired with a standard digital camera. Traits measured by TIPS include those traditionally measured by hand (tassel weight, tassel length, spike length, and branch number), as well as additional tassel characteristics that cannot be quantified manually (curvature, compactness, fractal dimension, skeleton length, and perimeter). The computing infrastructures of CyVerse and Open Science Grid enable completely automated batch processing of tassel images, making TIPS a scalable tool for conducting accurate and efficient population-level studies of maize tassel morphology.

W282: CyVerse - Software, Tools, and Services for Data-Driven Discovery
Genetic Study of Maize Yield-Component Traits Measured by Automated Image Analysis
Celeste Marie Falcon, University of Wisconsin-Madison, Madison, WI
Studying yield component traits, like maize ear and kernel dimensions, allows plant breeders and geneticists to dissect the genetic and environmental factors that influence grain yield. Measuring yield components automatically instead of manually would enable more efficient phenotyping of larger populations in more environments. To this end, collaborators developed a high-throughput image analysis workflow that extracts trait measurements from thousands of digital images of maize ears, cobs, and kernels. This data storage challenge was met by the data management platform provided by CyVerse. Here, we present two studies that utilized this image analysis platform to extract eight phenotypic measurements (ear width and length, kernels per row, kernel weight, width, length, thickness, and area) for three ears and ~100 kernels per plot. In one study, 31 inbred lines were grown in 36 environments nation-wide to examine genotype-by-environment interactions (G × E) from several angles including the genotypes' stability/plasticity, environments' discriminability, and traits' sensitivity to G × E. The other study involved 837 inbreds grown in two environments in Wisconsin, as well as a subset of 500 inbreds grown in four environments in Minnesota and one in Iowa. In this study, along with exploring the G × E, we conducted genome-wide association analysis to identify genomic regions contributing to yield components. Improved understanding of the genetic and environmental factors influencing yield component traits, enhanced by this high-throughput image analysis method, will allow us to choose more effective breeding strategies and will contribute to marker-assisted breeding approaches to develop higher yielding cultivars.

W283: CyVerse - Software, Tools, and Services for Data-Driven Discovery
Machine Vision Phenotyping of Seedling Growth and Morphology
Cory D. Hirsch[1], Tara A. Enders[2], Nathan Miller[3], Elizabeth Sampson[1], Sara Tirado[2], Nathan M. Springer[2] and Edgar P. Spalding[3], (1)University of Minnesota, St. Paul, MN, (2)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (3)University of Wisconsin-Madison, Madison, WI
There is a need to integrate large-scale genomic information with phenotypic information to sustain and improve plant productivity and our understanding of plant biology. The efficiency, scale, and cost of obtaining genomic information has advanced greatly, while high-throughput phenotyping methods have remained labor intensive and often subjective and expensive. Towards alleviating these barriers, we have developed a user friendly and affordable platform to acquire highly standardized RGB images, while relying on minimal equipment and space in a laboratory setting. This system uses the storage, application deployment, and compute resources available through the infrastructure at CyVerse. These resources allow this platform accessibility to almost any researcher. Currently, we have developed algorithms to extract ten biologically relevant traits including plant height, width, stem diameter, and pixel area. The traits obtained with our image capturing technique and algorithm design correlate well with both traditional hand measurements and measurements using manual software. We have optimized the system for daily collection of images of multiple plants to easily look at plant development. This system is currently being used by multiple research groups across several institutions and has been applied to monitor developmental variation among different maize genotypes subjected to temperature stress and to measure heterosis for seedling growth using a panel of maize inbred and hybrid genotypes.

W284: CyVerse - Software, Tools, and Services for Data-Driven Discovery
Development of a Gene Family Toolkit for Exploring Diversity in New Sequence Data
Paul C. Bailey, Earlham Institute, Norwich, United Kingdom
As more genomes are sequenced it is becoming apparent that gene duplication and deletion are important drivers in evolution, with rapidly expanding gene families often a signature of their role in an organism's adaptation to the environment. To identify these events, we are building a gene family analysis toolkit which will be deployed on the Cyverse cloud infrastructure for use by the scientific community. Central to its purpose will be the ability to distinguish evolutionary relationships between genes within gene families for currently and newly sequenced species, both at the intra- and inter-species level.
As a pilot study, we are using a collection of landrace bread wheats (the Watkins collection) which have been sequenced by exome capture to explore the diversity of the large Nucleotide Binding Leucine Rich Repeat (NLR) family of plant intracellular immune receptors in the collection. Illumina read data from each wheat line have been assembled using various tools. The resulting contigs have been aligned to their corresponding subgroups within the NLR family. Looking at subgroups that are expanded relative to other monocot species, we can demonstrate that further novel gene duplication events have occurred in specific lines of the Watkins collection. The next step will be to understand whether specific genes in the family are under positive selection and therefore which genes have particular functional significance. The assembly and downstream procedures will be placed into a Docker container for use on Cyverse as a tool for exploring the diversity of any gene family in any species with sequence data.

W285: CyVerse - Software, Tools, and Services for Data-Driven Discovery
CyVerse UK: Widening the Scope to the UK and Beyond
Alice Minotto[1], Erik Van Den Bergh[2] and Robert P. Davey[1], (1)Earlham Institute, Norwich, United Kingdom, (2)EMBL-EBI, Hinxton, United Kingdom
CyVerse UK aims to provide geographically advantageous access to life science HPC infrastructure in the UK and Europe, keeping the same objectives as the parent CyVerse project. In addition to those tools already available within the CyVerse ecosystem, CyVerse UK now hosts 35 file-processing and bioinformatics applications, developed both on site at the Earlham Institute and elsewhere.
The EI execution system consists of an HTCondor pool. We are working on combining the US and UK pools through *flocking*, making it possible for users to run their jobs in the most favourable location. We already rely on Docker containers to dissociate applications and workflows from the underlying computing architecture.
CyVerse UK also provides data storage capability in the form of the CyVerse UK Data Store. We are working to federate the UK and US storage systems through iRODS to provide seamless and transparent data transfer and availability between geographically disparate CyVerse systems. We are also working to set up the CyVerse UK Data Commons, a public collection of community and curated datasets of interest to researchers.
CyVerse UK currently hosts other projects, such as COPO and SignaLink, and fosters collaborations throughout the research community. In the spirit of outreach we made a Galaxy instance available for researchers at ILRI-BecAHub in Kenya to submit jobs to the CyVerse UK pool.
We will also host the main infrastructure for the UK Wheat Information System node.

W286: CyVerse - Software, Tools, and Services for Data-Driven Discovery
Viewing Data Hosted at CyVerse on the UCSC Genome Browser
Brian Lee, UCSC, Santa Cruz, CA
CyVerse's support of byte-range requests makes it an optimal location for storing and viewing genomic data in genome browsers. Uploaded data can be viewed on the UCSC Genome Browser through the "Send to Genome Browser" link creation option. The UCSC Genome Browser can display a variety of formats from bigDataUrls in this way, including BAM, CRAM, and VCF formats as well as a collection of bigBed, bigBarChart, bigGenePred, bigPsl, bigChain, bigMaf, bigNarrowPeak, and bigWig track types. Assembly Track Hubs hosted at CyVerse, using the .2bit file format, allow one to further quickly view new genomes by building a few supporting files around the converted FASTA file.

W287: Data Resource Sustainability and Funding
Peanutbase and the Peanut Genome Project Consortium
Ethalinda Cannon, Iowa State University, Ames, IA
Over a period of several years in the early 2010s, a consortium of U.S. peanut growers, shellers, manufacturers, and researchers came to the consensus that there was a need to fund a vigorous basic research initiative in support of peanut genetics, breeding, and production. The centerpiece of this research effort would be to sequence the peanut genome. Many other resources would also be developed, including markers, genotyping platforms, and genotyped and phenotyped recombinant inbred lines. The sequencing effort was complex due to the repeat-rich, allotetraploid genome in cultivated peanut, so the consortium first sequenced two diploid relatives of peanut, before tackling the full tetraploid genome assembly. To house the data from this ambitious, multifaceted effort, an online database was established: PeanutBase.org.
Representatives from each of the funding groups have been involved in every step of the planning and scientific outcomes of this effort. This five year project has transformed peanut research and breeding, and the peanut community itself. As the fifth year has now ended, PeanutBase and the Peanut Genome Consortium is looking for ways to continue funding, including exploration of partnerships with NIFA, USDA-ARS, and NSF. Another avenue for database longevity is the use of well-supported, widely-used platforms and tools such as the GMOD suite, and allying with similar online database projects. To ensure long-term data availability, and interoperability with data for other legumes, PeanutBase is closely allied with the Legume Information System (LegumeInfo.org) and is part of the NSF-funded project, The Legume Federation.

W288: Data Resource Sustainability and Funding
Sustaining the CIPRES Science Gateway, a Public Resource for Creating Large Phylogenetic Trees.
Mark A. Miller, San Diego Supercomputer Center, La Jolla, CA

The CIPRES Science Gateway (CIPRES) is a highly accessed digital resource for phylogenetics research. CIPRES provides public access to community phylogenetics codes run on high-end HPC resources through browser and RESTful interfaces at no cost to the user. Over the past eight years, CIPRES has received more than 1 million job submissions from more than 25,000 unique users in 86 countries. The results of these analyses have appeared in more than 3,700 peer-reviewed publications. CIPRES operations are currently supported by grants from the National Science Foundation. This funding supports essential software development tasks including: routine adaptations to an ever-changing IT environment; updating existing codes and adding new ones; making innovations that improve the functionality of the underlying software; and adding features the resource to accommodate the evolving needs of phylogenetics researchers. While CIPRES operations are supported by research funding from the National Science Foundation at present, long-term operation of a research resource that supports 8,000 users per month is precarious at best, given the highly competitive and uncertain 3-5 year funding cycle of federal granting agencies. Moreover, CIPRES' position as critical infrastructure for biology does not necessarily map well to the mission of research funding agencies and foundations. This presentation will describe strategies used by the CIPRES team to minimize ongoing costs of operation, and articulate a strategy for funding CIPRES through a mixed funding model where CIPRES infrastructure is supported by a fee-for-service model and innovation is funded through competitive grant awards.

W289: Data Resource Sustainability and Funding
Ag Data Commons: Harnessing the Power of Digital Agriculture
Cynthia Parr, National Agricultural Library, US Department of Agriculture, BELTSVILLE, MD
Public access to data underlying research results is a growing expectation. In part, this is because federal funders are seeking accountability, but open science also enables re-use towards new discoveries and applications. Agriculture is poised for interdisciplinary and big data approaches to address critical societal needs. However, in most agricultural disciplines, public data are scattered, poorly documented; metadata and data are rarely standardized or interoperable. Ag Data Commons is an ecosystem of distributed platforms designed to support the FAIR data principles. In our pilot, initial users are USDA-funded researchers and partner repositories who are sharing data products. Many researchers are new to distributing data using data repositories. Some research groups need a single public point of access to their scattered data output. Some don't yet have a domain-specific repository and want a cost-effective place to make data files available. Growth will bring a broader audience of researchers, application developers, an citizens seeking to use the data. Our curators and planned tools make datasets more discoverable and increase quality, machine-readability, and understandability of both metadata and data. For now, Ag Data Commons focuses on data with few or no access restrictions, consistent with USDA's Public Access Implementation Plan. In-house funding enables us to move ahead on some of our vision, but not on intelligent tools or fully scalable storage. We plan a business model with a mix of funding sources including external deposit fees so that we can ensure enduring, free access to high quality federally-funded data.

W290: Data Resource Sustainability and Funding
Incentivizing Volunteer Curation through Micropublication
Paul W. Sternberg[1], Karen Yook[1], Daniela Raciti[1], Todd W Harris[2] and Tim Schedl[3], (1)California Institute of Technology, Pasadena, CA, (2)Ontario Institute for Cancer Research, Toronto, ON, Canada, (3)Washington University School of Medicine, St. Louis, MO
WormBase is the authoritative database of genomic and genetic data of the free living nematode, *Caenorhabditis elegans,* a major model organism for biomedical research. WormBase houses the entire genome for this nematode and related nematodes including many parasites and agricultural pests. The gene models of all of these species are continually updated with an emphasis on those species most intensively used. In addition, WormBase curators collect and annotate gene function data from the published primary literature. As a result, we provide user-friendly report pages for over 39 data classes, encompassing nearly 12 million unique pages. Curating and integrating biological and biomedical knowledge into computable publicly available resources is a key step to expedite new discoveries but is costly and time consuming. So-called community annotation in which researchers volunteer effort to associate metadata with data or integrate information has proven difficult to incentivize. We have established a new data capture and dissemination paradigm that automatically and simultaneously captures and ingests biomedical data into our repository and publishes them in an online, peer-reviewed, open access journal *'Micropublication: biology'*. This new platform introduces a curation paradigm shift, allowing authors to directly submit the output of their research into the database using pre-designed intelligent web forms. Simultaneously, the process will automatically generate a 'publication-like' PDF file that will be publishable and citable according to findable, accessible, interoperable and reproducible (FAIR) data principles. We call these single result experiments, streamlined with minimal narrative, "micropublications." Authors preserve provenance and establish credit for their research and the automated flow of data they submit will be made publicly accessible in WormBase, integrated with existing datasets that have been manually extracted from the literature for almost 2 decades. In addition, researchers will be able to share both positive and negative data with the scientific community, fulfilling funding agencies' requirements to share all data coming from publicly funded research for further re-use. While we, along with most information resources, have spent effort on eliciting curatorial action from our communities, the Micropublication:biology platform provides an added incentive of a bona-fide citation that will eventually be indexed and findable in major citation indexers such as PubMed. After establishing this data retrieval/publication pipeline with WormBase first, and the other Model Organism Database (MOD) members of the Allied Genome Resources (AGR): FlyBase, Mouse Genome Database (MGI), Rat Genome Database (RGD), Saccharomyces Genome Database (SGD), Zebrafish Model Organism Database (ZFIN), we will work to expand to non-member, but otherwise critical organismal databases, such as Xenbase (*Xenopus laevis* and *tropicalis* Database), DictyBase (*Dictyostelium discoideum* database), PomBase (*Schizosaccharomyces pombe* Database), among others. I will discuss our progress and potential broader use of this incentivization approach.

W291: Data Resource Sustainability and Funding
Presenter Unable to Attend: National Institute of Food and Agriculture
Parag Chitnis, National Institute of Food and Agriculture, Washington, DC

W292: Data Resource Sustainability and Funding
NSF's Division of Biological Infrastructure and the ABI Program Tracks
Jennifer Walsh Weller, National Science Foundation, Arlington, VA
The Advances in Biological Infrastructure program at the National Science Foundation is the only program whose purpose is to fund the development of infrastructure supporting computational needs for researchers in the biological sciences. The Division of Biological Infrastructure has funded development and deployment of a considerable and diverse collection of software tools and databases, computing hardware, and training to produce computationally adept research scientists. Because the mission of the NSF is to fund innovative research there is always a tension between funding the newest tools or keeping a well-recognized resource updated and accessible. Highly accessed resources tend to become like air – no one notices the dependency but if they disappear there is no obvious way to accomplish key tasks. However, it is not possible to sustain >10,000 biological databases and hundreds of thousands of software tools on current budgets, even if we gave up funding any new research. In order to select those resources to be maintained by funding from the ABI Sustainability track, we demand that an applicant demonstrate the relevance of the resource to current active research by scientists other than the applicants, and present a plan for how the resource will be sustained once the requested funding expires. It is already true that we ask reviewers to look at applications in innovation, development and sustaining tracks with respect to sharing responsibility for development and maintenance as broadly as possible, which implies using well tested standards for representation, standard I/O formats and methods, and container-base application deployment. ABI is tolerant of a wide range of data and software release models so long as the model is explicitly defined in the proposal and the choice of the method is clear and reasonable for the target users. ABI is also tolerant of mixed funding models so long as ownership of intellectual property and policies on data and code release are established at the time a proposal is submitted. ABI is interested in seeing proposals that try a range of approaches and carry out assessment and produce success/failure metrics that can be used to guide different research communities as to what works and what does not for their users. We are interested in hearing from the community what their routine needs are in all three areas (tools, hardware, people) both now and projected for the next 5 years.

W293: Degraded DNA and Paleogenomics
Domestication Bottlenecks: The Last Sacred Cow of Domestication?
Robin G. Allaby, University of Warwick, Warwick, United Kingdom
Domesticated crops show a reduced level of diversity which is commonly attributed to the 'domestication bottleneck', a drastic reduction in the population size associated with sub-sampling the wild progenitor species and the imposition of selection pressures associated with the domestication syndrome. A prediction of the domestication bottleneck is a sharp fall in genetic diversity early in the domestication process. Surprisingly, archaeological genomes of three major annual crops do not indicate that such a fall in diversity occurred. In light of this observation, we revisit the general assumption of the domestication bottleneck concept in our current understanding of the evolutionary process of domestication.

W294: Degraded DNA and Paleogenomics
Ancient DNA Meets Forensics
Samuel H. Vohr[1], Sidra Hussain[1], Joshua D. Kapp[1], Jelmer W. Eerkens[2] and **Richard E. Green**[3], (1)University of Califonia, Santa Cruz, Santa Cruz, CA, (2)University of Califonia, Davis, Davis, CA, (3)University of California Santa Cruz, Santa Cruz, CA
We have applied many of the advances in extraction and analysis of DNA from ancient remains to forensics samples. Forensic samples are often limited by the same factors that limit ancient DNA analyses: damaged and degraded DNA samples, limited material, and mixed samples. I will present two new algorithmic approaches suitable for both forensics and ancient DNA. The first is a method to determine the identity of a sample given extremely low amounts of input material. The second is a method for deconvolution of complex mixtures that is capable of accurate and complete determination of mtDNA haplotypes. Finally, I will show how these and similar approaches were used to solve a real life mystery of a historical burial in San Francisco in the 1870s.

W295: Degraded DNA and Paleogenomics
Caving for Ancient DNA – Looking for Human Impact on the Environment
Anna Linderholm, Texas A&M University, College Station, TX
The first successful use of sediment as a source of DNA was back in 2003, when researchers showed that it was possible to use DNA trapped in sediment to track changes in both taxonomic diversity and composition of Beringian vegetation and fauna. DNA has since then been extracted from ice, lake sediments, soils and coprolites. With the introduction of Next Generation Sequencing this subfield of ancient DNA research has gained a lot of traction. Different studies have found evidence of both extinct and current animals and traces of plants no longer found in the region. Complete reconstructions of long lost environments are now possible.
Hall's Cave is a unique paleo environment archeological site. It is an underground cave located in central Texas. It is a remarkable site due to its prime location on the central Edwards Plateau, the bedrock consists primarily of limestone hence DNA preservation is excellent. The site has been excavated extensively and the oldest layers has been proven to be at least 18,000 years old. Native Americans used the cave and surrounding area intermittently from the late Paleo Indian period to the Late Prehistoric period. This means that this site represents a unique place to investigate any human impact on the surrounding environment. Samples have been taken across the stratigraphy, thus representing both before and after human entered the cave and the surrounding environment. Initial results show a remarkable change in both flora and fauna at several distinct time periods.

W296: Degraded DNA and Paleogenomics
Picking the Bones out: Ancient Animal Genomics in the Fertile Crescent
Daniel G. Bradley, Trinity College Dublin, Dublin, Ireland
The origins of domestic ungulates have been explored for over two decades by extrapolating patterns from modern genomes. However, such patterns are an amalgram of overlaid processes and deciphering the past can only bereliably achieved by including reference to ancient genome

variation. With these data we can begin to answer questions about the nature of domestication such as: did it radiate from a central innovative core region or did separate Neolithic communities capture and tame wild progenitors across the Fertile Crescent? Also, can we deduce where African, European and Asian populations trace their ancestry back to at the beginnings of herding? Which other processes, such as wild introgression, may have played a role in establishing modern genomes?

W297: Degraded DNA and Paleogenomics

Genome-Based Sexing of Prehistoric Animals

Love Dalén, Swedish Museum of Natural History, Stockholm, Sweden

Ancient remains can provide a wide array of information about the ecology and behaviour of prehistoric animals. In recent years, biomolecular methods such as radiocarbon dating, palaeogenomics and stable isotope analysis have become increasingly popular to recover information from ancient remains. However, several of these applications would benefit from knowledge about the sex of each specimen that is studied, and recovering this information is not an easy task due to the often fragmentary nature of the fossil record. Shotgun-based DNA sequencing methods offer a solution to this issue, since low coverage genomic data mapped against sex-specific chromosomes can be used to accurately assign biological sex. The aim of this presentation is to present molecular sexing data from a large number of prehistoric samples. The results suggest that the available fossil record is not always a random representation of the past, and thus that care needs to be taken to avoid biases in population-level studies. However, such deviations from expected sex ratios might in turn provide clues about the behaviour and ecology of prehistoric populations.

W298: Degraded DNA and Paleogenomics

Ancient DNA of Grape Seeds Provides Insights into Viticulture and Cultivation Practices in Post-Roman France

Jazmin Madrigal, Natural History Museum of Denmark, København K, Denmark

The grapevine (*Vitis vinifera*) was one of the earliest fruit crops to have been domesticated and has since played an important role as both a food staple and the basis of wine production. Despite being clonally propagated, present-day grape cultivars hold morphological and genetic diversity comparable with that of annual crops such as maize, with thousands of different varieties having been described in both historic and contemporaneous records. Some of these varieties can be traced as far back as the Middle Ages, through morphological descriptions and historical records. However the genetic relationship between ancient and modern varieties remains unknown.

We explored the potential of ancient DNA in the identification of ancient grape cultivars by assembling a dataset of ancient samples from France: a region that is today home to one of the largest collections of grape cultivars. Specifically, we performed targeted-high-throughput sequencing of ten thousand recently described cultivar-diagnostic SNP sites, on 28 ancient and historic grape seeds from archaeological sites dated to the Roman and Medieval periods. We compared these ancient samples to a genotype panel containing present-day wild and domesticated varieties, and were able to assign them to a defined geographic cluster of domesticated grapes. Our analyses identified 16 different ancient cultivars and resolved their genetic relationships with modern grape varieties. Furthermore, we show genetic evidence of clonal propagation over at least one thousand years, which gave rise to the *Savagnin* cultivar as well as some of the most important cultivars found in the region today.

W299: Degraded DNA and Paleogenomics

Heirloom Dairy Cultures: The Prehistoric Origins and Modern Diversity of Eurasian Dairying

Christina Warinner[1], Shevan Wilkin[1], Richard Hagan[1], Björn Reichhardt[1], Ashley Scott[1], Cheryl Makarewicz[2], Soninkhishig Tsolmon[3] and Jessica Hendy[1], (1)Max Planck Institute for the Science of Human History, Jena, Germany, (2)Christian-Albrechts-Universität, Kiel, Germany, (3)Mongolian University of Science and Technology, Ulaanbaatar, Mongolia

Within and beyond the human body, human cultures enable microbial ecosystems to grow and thrive. In our cuisine, this process is evident in the creation of dairy products, which have enormous global complexity and diversity in taste, aroma and texture. Thousands of years ago, with the invention of products such as yoghurts and cheeses, people were domesticating and manipulating microbes before they even knew of their existence. However, modern dairy products in industrialized economies are highly regulated in order to maintain hygiene standards and reproducibility, leading to a striking reduction in the number of microbial strains involved in food fermentation. We remain unaware of the vast microbial diversity involved in ancient food preparation, and the impact this microbial diversity would have had on flavors and textures. Additionally, as a result of contemporary food globalization and industrialization, traditional methods of dairying and their unique microbial cultures (such as yoghurt strains passed down through generations) are being lost at an alarming pace. This study explores two non-European regions where dairying forms an integral part of cuisine but is highly distinctive: the Middle East and Mongolia. Here we present new paleoproteomic evidence of the antiquity of prehistoric dairy production in the Middle East and Mongolia, as well as our findings to date on the nutritional and microbial diversity of contemporary dairy products produced by small-scale nomadic pastoralists in these two regions.

W300: Degraded DNA and Paleogenomics

Tracking Six Millenia of Horse Selection, Admixture and Management with Complete Genome Time-Series

Ludovic Orlando[1,2] and The ERC PEGASUS consortium[2], (1)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, (2)Laboratoire d'Anthropologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse III Paul Sabatier, Toulouse, France

The domestication of the Horse and its impact on warfare, transportation and agriculture, have revolutionized human history. Even though most modern breeds have been engendered within the last couple of centuries, humans have managed horse livestock for over five millenia. Recent selective and management strategies have tremendously impacted the genetic structure of horse populations. As a result, modern patterns of genetic diversity can only partly help reconstruct the horse domestication process prior to the modern era. Recent research in our laboratory, carried out in the framework of the ERC PEGASUS programme, has endeavoured to sequence complete horse genomes from accross their whole temporal and geographical domestication range in order to identify how the many past human cultures progressively forged the horse genome by means of selection, drift and admixture. This work revealed two different dynamics at play within early and late domestication

stages, involving the selection for different functional pathways, different management strategies for the genetic resource available, including stallion diversity, and a recent increase in the genomic deleterious load. Our new genome dataset now allows us to document such changes at unprecedented scales and reveals unexpected features of the whole population dynamic underlying horse domestication.

W301: Developing and Executing Successful Broader Impact Programs for Current and Future Grants
Better Mentoring for Science and Life
Michelle A. Graham[1], Jamie A. O'Rourke[1] and Michael D Gonzales[2], (1)USDA-ARS-MWA-CICGRU, Ames, IA, (2)University of Georgia, Albuquerque, NM
The purpose of broader impacts is to promote scientific discovery while encouraging the education and training of students at all skill levels. Mentoring provides a safe learning environment for students to acquire new skills and knowledge, leading to professional development. For many scientists, mentoring is focused in and around the lab bench. However, for many students, personal circumstances, families and communities impact career decisions. This presentation will focus on improving mentoring by understanding the balance between science and every day life.

W302: Developing and Executing Successful Broader Impact Programs for Current and Future Grants
MutantMillets: A Platform for Gene Discovery in the Classroom
Thomas P. Brutnell, Enterprise Institute for Renewable Fuels Donald Danforth Plant Science Center, St. Louis, MO
The MutantMillets program (https://mutantmillets.org) was developed to introduce high school and undergraduate students to inquiry-based plant genetics. Setaria viridis (green bristlegrass) is the wild relative of foxtail millet, a crop grown for food security throughout Asia and Africa. In the MutantMillets program students screen chemically mutagenized populations and identify novel phenotypic variants segregating in families. These phenotypes are documented and often picked up for further characterization by scientists at the Danforth Center. In one example a novel phenotype identified by a student is now being fine mapped using bulk segregant analysis to reveal the underlying molecular lesion. A high school teaching training program has been integral to the success of the program which now has reached over 1000 students in the St. Louis region.

W303: Developing and Executing Successful Broader Impact Programs for Current and Future Grants
A Gaggle of Geese, a Murder of Crows, a Diversity of Impacts
Carolyn J. Lawrence-Dill, Iowa State University, Ames, IA
The NSF says, "While intellectual merit is about the potential to advance knowledge and encompasses the scientific research proposal, the **broader impacts** criterion deals with the potential to benefit society and contribute to the achievement of specific, desired societal outcomes." From outreach to American Indians to investigating (and changing) attitudes about gene editing, the potential to make broad and lasting impacts is there and the work can be both challenging and fun. This seminar will describe a number of projects that focus on how to ensure that plant biology research benefits students, other scientists, and society writ large.

W304: Developing and Executing Successful Broader Impact Programs for Current and Future Grants
NSF Funding Opportunities and Resources for Outreach and Training: Strategies for Success
Diane Jofuku Okamuro, National Science Foundation/BIO/Division of Integrative Organismal Systems, Alexandria, VA
The U.S. National Science Foundation (NSF) receives approximately 50,000 research proposals each year, of which approximately 12,000 are funded. The NSF employs the two National Science Board approved merit review criteria to determine which research has the greatest potential to promote the progress of science: Intellectual Merit and Broader Impacts. While most researchers know what is meant by intellectual merit, experience shows that many have a less than clear understanding of the meaning of broader impacts. NSF representative(s) will meet with workshop participants and present general information about current funding opportunities and resources available to facilitate and support outreach and training activities.

W305: Developing and Executing Successful Broader Impact Programs for Current and Future Grants
GETBIO-PGR: The Gateway for Education, Training, Broader Impacts and Outreach in Plant Genome Research
Carol Lushbough, University of South Dakota, Vermillion, SD
Increasingly, NSF investigators are encouraged to envision an integration of their own research with education, training, and outreach so that Broader Impacts are interwoven throughout. Broader impacts are of great importance but also pose many challenges to those seeking opportunities as well as for researchers planning and implementing programs such as: 1) finding and sharing resources; 2) locating and collating information that is diverse, widespread and presented in a variety of ways; 3) acquiring technical expertise and infrastructure for creating websites, videos, integrating applications such as social media, survey tools and analytic tools; 4) disseminating information and interacting with large, distributed groups; 5) meeting recruitment initiative goals to broaden participation and building collaborations in unfamiliar communities; 6) measuring success and promoting and highlighting successful projects; 7) developing creative Broader Impacts initiatives; and 8) locating requisite bioinformatics analytic tools to be applied specific research questions.
GETBIO-PGR, An integrated Gateway for Education, Training, Broader Impacts and Outreach platform has been created to provide an infrastructure for researchers, educators, students and the general public to access, create, share and exchange information about research project's broader impact activities. This Gateway will enable researchers to share the full range of their broader impact activities within a centralized, integrated infrastructure. GETBIO-PGR will elevate the visibility and importance of broader impact activities -- and their intertwined research activities -- making it possible to offer more rapid dissemination and more effective use of best practices for mentoring, workshop management, and data analysis.

W306: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture
Cross-Suppression of *AG* and *AG-like 11* Genes gives Sterility in Field Grown Poplar

Haiwei Lu[1], Amy Leigh Klocko[1], Amy Brunner[2], Anna Carlina Magnuson[1], Cathleen Ma[1] and **Steven H. Strauss**[3], (1)Oregon State University, Corvallis, OR, (2)Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, (3)Oregon State University, Oregon State University, Corvallis, OR 97331, USA, Corvallis, OR
Concerns over transgene dispersal have limited field studies and commercial use of genetically engineered trees. We seek to mitigate concerns by producing sexual sterility in poplar. Based on cDNA sequence of the *AGAMOUS 2* gene in *Populus trichocarpa*, we created two RNA interference (RNAi) constructs, PTG and its matrix-attachment-region flanked version MPG, and transformed *P. alba* genotype 6K10, an early flowering female clone. A total of 35 transformed events with four ramets per event and 24 wild-type control trees were planted as part of a larger field trial in 2011. Six out of 22 flowering PTG events and 11 out of 12 flowering MPG events showed a modified floral phenotype; their floral buds flushed early in the field and the capsules on each catkin often had “carpel-inside-carpel” phenotypes. A complete disruption of ovule and seed production was observed in a number of gene insertion events within both constructs. Quantitative RT-PCR (qRT-PCR) revealed suppression at two *AG* orthologs and two *AG-like 11* (*AGL11*) orthologs in all events with strongest suppression in sterile events. In all cases, trees appeared normal in their vegetative morphology and growth, and alterations in floral phenotypes were stable over multiple years. RNAi suppression of *AG*-like genes appears to be a safe and effective means of genetic containment in poplar.
We thank the USDA Biotechnology Risk Assessment Grants (no. 2010-33522-21736 and no. 2011-68005-30407) and the Tree Biosafety and Genomics Research Cooperative at OSU for support.

W307: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture
Enhanced CRISPR/Cas9-Mediated Precise Genome Editing by Improved Design and Delivery of gRNA, Cas9 Nuclease, and Donor DNA

Jason Potter, ThermoFisher Scientific, Carlsbad, CA

While CRISPR-based gene knock out in mammalian cells has proven to be very efficient, precise insertion of genetic elements via the cellular homology directed repair (HDR) pathway remains a rate-limiting step to seamless genome editing. Under the conditions described here, we achieved up to 56% targeted integration efficiency with up to a six-nucleotide insertion in HEK293 cells. In induced pluripotent stem cells (iPSCs), we achieved precise genome editing rates of up to 45% by co-delivering the Cas9 RNP and donor DNA. In addition, the use of a short double stranded DNA oligonucleotide with 3’ overhangs allowed integration of a longer FLAG epitope tag along with a restriction site at rates of up to 50%.We propose a model that favors the design of donor DNAs with the change as close to the cleavage site as possible. For small changes such as SNPs or short insertions, asymmetric single stranded donor molecules with 30 base homology arms 3’ to the insertion/repair cassette and greater than 40 bases of homology on the 5’ end seems to be favored. For larger insertions such as an epitope tag, a dsDNA donor with protruding 3’ homology arms of 30 bases is favored. In both cases, protecting the ends of the donor DNA with phosphorothioate modifications improves the editing efficiency.

W308: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture
On-Demand Trait Enablement in Crops Using Florian™ Switch Technology

Arianne Tremblay, AgBio Intrexon Corp, Davis, CA and Amanda Edwards, Rio Stamler, Trang Dang, Patrick Canlas, Jyoti Rout, Shiv Tiwari, Sekhar Boddupalli
Over the past 3 decades, agricultural biotechnology has revolutionized food and feed productivity with commercialization of crop protection and quality traits. Several of the crop protection traits are based on over-expression of transgenic pesticidal proteins. Rapid adaptation of these traits along with year-round crop production in several geographies has led to resistance development that is imminent across the globe. In addition, constitutive over production of the pesticidal proteins could negatively impact plant vigor and be yield limiting in some cases. In view of this, “on-demand” expression of transgenic traits, combined with the increasing adoption of precision agricultural tools could be particularly advantageous for future crop pipelines. In order to explore the potential of this concept, we have developed and tested Florian™ switch technology, a system which enables inducible control of gene expression using cost-effective and scalable chemistries for broad agricultural applications. Specific examples demonstrating flowering control, crop protection, and quality traits in two different plant systems will be presented. Potential commercial applications to further the Florian™ switch technology include: a) increasing biomass and feed quality in forage crops by prolonging the vegetative stage, b) enabling “on-demand” seed production with improved efficiency and effectiveness, c) synchronizing flowering in high value fruit and produce to aid in harvest timing (e.g. strawberries, pineapples), and d) bringing flexibility in crop protection against biotic and abiotic stresses based on disease pressures or climate conditions.

W309: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture
Precision Gene Editing in Agriculture

Steven L. Sanders, Christian Schöpke, Dave Songstad, Noel Sauer, Greg Gocal and Peter Beetham, Cibus US LLC, San Diego, CA
The field of precision gene editing has been enabled by a convergence of technological breakthroughs including the emergence of directed nucleases such as TALENs and CRISPRs. The potential benefits of precision gene editing are profound and widespread with applications across bacterial, fungi, mammalian and plant systems. This talk will focus on the many ways in which precision gene editing is revolutionizing agriculture by accelerating trait development and plant breeding. Cibus has developed an advanced non-transgenic plant breeding system called ***Rapid Trait Development System***™ (***RTDS***™) that precisely and predictably changes one or a few nucleotides within a plant gene to obtain a desired phenotype/trait. At the core of this technology is the Gene Repair OligoNucleotide (GRON). The GRON is a chemically synthesized oligonucleotide specifically designed to be used by the plants native DNA repair machinery as a template to precisely produce the targeted DNA change of interest within the host organism’s genome. The result is the development of a desired non-transgenic trait within a plant. These novel traits range from increased nutritional value, improved disease resistance and the ability to grow in challenging changing environments, all of which will help us to meet the future growing requirements for an adequate and sustainable global food supply.

W310: Development and Application of Genome Engineering and Transgenic Technology to the Agriculture
Transgenic Cavendish Bananas with Resistance to Fusarium Wilt Tropical Race 4
James Dale[1], Anthony James[2], Jean-Yves Paul[2], Benjamin Dugdale[2], Upendra Shekhawat[2], Peter Waterhouse[3] and Robert Harding[2], (1)Centre for Tropical Crops and Biocommodities, Brisbane, QUEENSLAND, Australia, (2)Queensland University of Technology, Brisbane, Australia, (3)Centre for Tropical Crops & Biocommodities, Queensland University of Technology, Brisbane, Australia
Fusarium wilt, caused by the soil borne fungus *Fusarium oxysporum* fsp *cubense* (Foc), is the most destructive disease of bananas. Foc race 1 infects a wide range of banana cultivars and was responsible for the destruction of the banana export industry based on the cultivar Gros Michel last century. This resulted in the adoption of the Foc race 1 resistant Cavendish as the now almost exclusive export banana. This one cultivar now represents more than 50% of the world's banana production. However, over the past 30 years, the impact of Foc tropical race 4 (TR4) has become increasingly devastating. TR4 infects and kills Cavendish. Very importantly, TR4 is moving and is now widespread in south east Asia, south Asia and present in the middle east and Africa. It is likely to continue to move. Resistance is clearly the most appropriate control strategy. Conventional breeding and genetic modification are valid approaches to generate resistant cultivars. We transformed Cavendish with two different transgenes, Ced9, a nematode derived anti-apoptosis gene, and RGA2, a NB-LRR resistance like gene from a TR4 resistant wild diploid banana, *Musa acuminata* ssp *malaccensis*. We took a small number of transgenic lines through to the field in northern Australia. After three years in the field, one line each of RGA2 and Ced9 lines had no disease in contrast to the more than 67% infection levels in controls. The level of resistance in the RGA2 lines was strongly correlated with the level of expression of the transgene.

W311: DivSeek – Addressing the challenges and opportunities for information and data sharing associated with plant germplasm
Seeds of Discovery (SeeD)
Sarah Hearne, CIMMYT International Maize and Wheat Improvement Center, Texcoco, Mexico
The SeeD initiative aims to explore and leverage high value novel diversity from germplasm bank collections for maize and wheat breeding application. SeeD has generated vast quantities of genotypic and phenotypic data from extensive evaluation of germplasm bank accessions. Since inception the effective collection and management of this primary data and the derivation of added value through interconnection and visualization have been initiative objectives. Strong emphasis has been placed on facilitating equitable access to knowledge and creating a framework where data remains in the public domain. We present some of the experiences of SeeD in the collection, curation, analysis and dissemination of germplasm bank collection derived data.

W312: DivSeek – Addressing the challenges and opportunities for information and data sharing associated with plant germplasm
The Germinate 3 Platform and Crop Diversity Informatics Tools at the James Hutton Institute
David Marshall[1,2], Paul Shaw[1], Iain Milne[1], Gordon Stephen[1] and Sebastian Raubach[1], (1)The James Hutton Institute, Dundee, United Kingdom, (2)SRUC, Dundee, United Kingdom
The development of technologies to rapidly characterise both *in situ* and *ex situ* genetic resources and breeding population of crop plants and their wild relatives provides an opportunity to improve the efficiency with which germplasm can be readily utilised in varietal development. However this in turn brings challenges in storing and integrating the resulting data in ways in which it can be readily accessed by plant geneticists and breeders and integrated with the necessary analytical software tools. At the James Hutton Institute in Scotland we have been working both breeders and germplasm curators to develop a set of software tools which can play a major role in several aspects of the analysis and utilisation pipeline. The major elements in out work revolve around the Germinate 3 database for storage and access to genotype and trait data and the Flapjack, CurlyWhirly and Helium software tools for visualisation and manipulation of analysed genotype, trait and pedigree information. We will demonstrate applications of these tools in both breeding and genetic resources context and indicate further development of our work through collaborations in the development of analysis pipelines with the Biometris and VSNi and more extensive integration with other tools and resources through the Plant Breeding API. Our software will run on a wide variety of platforms and is accessible through the following URL (https://ics.hutton.ac.uk/software/)

W313: DivSeek – Addressing the challenges and opportunities for information and data sharing associated with plant germplasm
The Digital Cassava Genebank
Sarah C. Dyer[1], Bruno A. Santos[1], Mohamed Abdelhalim[1], Pradeep Ruperao[1], David Marshall[2] and Peter Wenzl[3], (1)NIAB, Cambridge, United Kingdom, (2)The James Hutton Institute, Dundee, United Kingdom, (3)CIAT, Cali, Colombia
CIAT's Genetic Resources Program houses >6,000 accessions of cassava (*Manihot esculenta*) and its wild relatives. Relatively little is known about the accessions within the genebank, aside from their collection information, with the majority of accessions coming from Colombia and Brazil. We have genotyped >4,000 accessions from the collection using DArT-Seq, generating >30,000 high confidence allele calls mapped to the reference genome. Using this genotyping data we have also selected 25 genetically diverse individuals to be whole genome sequenced to allow us to explore their genomic diversity more fully.
During this project we have identified a number of potential sources of error which may occur with large-scale genotyping projects for germplasm collections. We will present an update on the project and outline these potential issues and measures which we recommend to reduce and check for such errors when undertaking similar projects.

W314: DivSeek – Addressing the challenges and opportunities for information and data sharing associated with plant germplasm
Divseek Canada Initiative

Richard M Bruskiewich[1,2], Kirstin Bett[3], Helen M. Booker[3], Sylvie Cloutier[4], Frank M. You[5], S. Evan Staton[6], Andrew Warfield[7], Emily Warfield[8] and Loren H. Rieseberg[6], (1)Crop Informatics / STAR Informatics / Delphinai Corporation, Sooke, BC, Canada, (2)Department of Botany/University of British Columbia, Vancouver, BC, Canada, (3)University of Saskatchewan, Saskatoon, SK, Canada, (4)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (5)Agriculture and Agri-Food Canada, Morden, MB, Canada, (6)University of British Columbia, Vancouver, BC, Canada, (7)Department of Computing Science/University of British Columbia, Vancouver, BC, Canada, (8)Peter A. Allard School of Law/University of British Columbia, Vancouver, BC, Canada

The challenges to global food security are well documented. Canada must dramatically expand agricultural production to meet the challenges by developing high yielding, climate-adapted and "planet friendly" varieties. Canada is also a signatory of the Convention for Biological Diversity and the International Treaty for Plant Genetic Resources in Food and Agriculture. To become fully compliant with the Treaty and other international agreements, we must develop effective mechanisms for sharing plant germplasm, as well as genotypic, phenotypic, and genomic information.

DivSeek (http://www.divseek.org) offers a potential pathway forward on both fronts by accelerating plant breeding though development of "*a unified, coordinated and cohesive information management platform to provide easy access to genotypic and phenotypic data associated with genebank germplasm*."

In the spirit of DivSeek, the goal of our pilot DivSeek Canada initiative is to develop an online crop informatics resource, hosted on ComputeCanada high performance infrastructure, to integrate plant genetic resource germplasm and genotypic and phenotypic metadata for three Canadian crop communities (flax, lentils and sunflower) and to deploy mapping, breeding, and visualization tools online to use this data, as a blueprint for the integration of genomics data from other crops in the future. We suspect that the utility of such an informatics resource will be greatest in medium to small crop communities that have not previously had the financial resources or bioinformatics skill set to fully exploit the enormous quantity of genomic information rapidly becoming available for essentially all crops and their wild relatives.

W315: DivSeek – Addressing the challenges and opportunities for information and data sharing associated with plant germplasm

Digital Object Identifiers for Plant Genetic Resources

Marco Marsella, Food and Agriculture Organization of the UN, Rome, Italy and **Ruaraidh S Hamilton**, International Rice Research Institute, Metro Manila, Philippines

Progress in genomics depends on sharing genetic materials and associated information between collaborating laboratories. This in turn requires accurate, reliable identification of the material and associated information. Effective standards are already in place and widely adopted to identify information in digital publications and datasets through Digital Object Identifiers (DOIs). However, standards for identifying material lag far behind. Even advanced laboratories find it challenging to track genetic samples across collaborators; and once data are published, the link between the data and the material it describes is usually quickly lost.

A solution to these problems is now available through the Global Information System (http://www.fao.org/plant-treaty/areas-of-work/global-information-system/en/), in the form of DOIs for genetic materials (https://ssl.fao.org/glis). Functioning since October 2017, these digital identifiers are used to identify samples of physical material uniquely, globally, permanently. Over 500,000 DOIs have already been registered. Everyone who holds genetic materials is encouraged to register DOIs for the material they hold, free of charge; the only obligation is to make sure that each DOI is used only for the material originally registered.

Two metadata fields provide much of the functionality. One points to URLs of data or publications, providing a permanent association between the material and associated information. A second, points to the DOI(s) of the progenitor from which the material was derived, enabling the documentation of relationships between genetic material across different laboratories.

W316: Domestication Genomics

Detecting Polygenic Adaptation in Maize

Emily B. Josephs, University of California, Davis, Davis, CA

Characterizing the genetic basis of adaptation is not only a longstanding goal of evolutionary biology, but is also an important component of understanding adaptation in domesticated plants. Many of the traits that are likely to contribute to adaptation in domesticated plants are quantitative and adaptation in these traits likely often occurs through subtle shifts in allele frequencies at many loci. However, we lack methods for detecting polygenic adaptation in domesticated plant species. In this talk I describe strategies for leveraging genotypic data at loci identified using genome-wide association studies (GWAS) to detect the coordinated shifts in allele frequency that we expect to occur during polygenic adaptation. I will discuss false positives that occur when population structure is shared between the GWAS panel and the genotypes investigated for selection and describe solutions to this problem. I will then present results using new methods about the specific traits that are locally adapted in maize inbred lines and landraces. Ultimately, the methods presented here will be applicable to a broad range of domesticated and wild plant species.

W317: Domestication Genomics

Deleterious Variants and the Genetic Cost of Domestication

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

The "cost of domestication" hypothesis posits that domestication could result in an increase in the number, frequency, and/or proportion of deleterious genetic variants that are fixed or segregating in the genomes of domesticated species. This cost may limit the efficacy of selection and thus reduce genetic gain in breeding programs. Understanding when and how deleterious mutations accumulate can also provide insight into fundamental questions about the interplay of demography and selection. Putatively deleterious variants can be readily identified as polymorphisms at phylogenetically-conserved nucleotide sites. Phenotype-changing variants are especially likely to annotate as a deleterious, suggesting both that variants important to domestication and improvement may be more likely to be identified as deleterious and that targeted identification and elimination of segregating deleterious variants is a potential novel approach to breeding.

W318: Domestication Genomics

Comparative Characterization of the Domestication Process in Two Independently Domesticated Pumpkin Species

Heather Rose Kates, University of Florida, Gainesville, FL

Population genomics approach to explore the genetic structure and domestication bottlenecks in two pumpkin species, buttercup squash (*Cucurbita maxima*) and cushaw (*C. argyrosperma*). Here we compare how different wild origins and breeding practices produce current crop diversity of these two independently domesticated pumpkin crops that share a domestication syndrome, but have differing domestication histories and breeding practices. We performed targeted genomic sequencing of 48 unrelated accessions for each species including wild, landrace, and improved lines and identify over 15,000 SNPs for each species. Population analysis of allelic diversity of SNP data shows a single domestication event in each species and suggests specific geographic regions where wild relatives are most similar to the respective domesticate. Results of these analyses also reveal that the locally important Mexican domesticate *C. argyrosperma* experienced a domestication bottleneck consistent with our expectations of the genetic consequences of domestication. In contrast, the phenotypically diverse and economically important *C. maxima* ssp. *maxima* does not exhibit a reduction in genetic diversity relative to its wild ancestor. These results improve our knowledge about the domestication of two *Cucurbita* species and add to our understanding of how the reduction of genetic diversity during the processes of domestication and trait improvement impacts the breeding potential and utility of current crops.

W319: Domestication Genomics

Phylogeny of the Genus *Gallus* and Contribution of the Different Species to Chicken Domestication

Frédéric Hospital[1], Marie Suez[2], Mahendra Mariadassou[2], Pierre Nicolas[2], Sambandam Sathyakumar[3], Alain Vignal[4], Xavier Rognon[5], Agathe Vieaud[1], Tatiana Zerjal[1] and Michèle Tixier-Boichard[1], (1)GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France, (2)MAIAGE, INRA, 78350, Jouy-en-Josas, France, (3)Wildlife Institute of India, Dehradun, India, (4)GenPhySe, INRA, ENVT, ENSAT, 31326, Castanet Tolosan, France, (5)GABI, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

The red junglefowl (*Gallus gallus*) is considered as the main ancestor of the domestic chicken, but the genus *Gallus* includes three other species : *Gallus sonneratii* (grey junglefowl), *Gallus varius* (green junglefowl) and *Gallus lafayetii* (Sri Lanka junglefowl). Deep genome sequencing (25-30X) has been undertaken for these species in order to analyze with a great accuracy the genetic diversity of the genus *Gallus* and to better understand the genetic make-up of domestic chickens. Wild individuals were sampled in their habitat (Thailand, India) or in zoological parks and a subset was chosen for sequencing on the basis of 57K genotypes. Domestic chickens were sampled according to their geographic origin (Europe, Africa, Asia, South America). In total 25 wild individuals and 18 domestic chickens, 1 per breed, mostly females, were sequenced in paired-end.

Phylogenetic trees were obtained with three types of genomic data: autosomal chromosomes, W chromosome, mitochondrial DNA. The comparison revealed an event of maternal introgression of *G. gallus* into *G. sonneratii* in some zoological parks: 9.6% of autosomes of *G. sonneratii* exhibited regions, of variable sizes, related to *G. gallus* genome. Introgressed regions were mapped and removed from the subsequent analysis of the demographic history of the genus *Gallus,* which was performed with PSMC. A hidden Markov chain model was implemented on the corrected data set of wild *Gallus* to analyze contributions to the genome of domestic chickens. *G. sonneratii* was the only species to exhibit a low but significant contribution to domestication (up to 1% of the genome).

W320: Domestication Genomics

Gene Flow between Rye and its Wild Relatives

Mona Schreiber, IPK Gatersleben, Stadt Seeland, Germany, Andreas Boerner, IPK Gatersleben, Germany and Martin Mascher, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany

Rye (*Secale cereale* L.) is a cereal grass that is an important food crop in Central and Eastern Europe. In contrast to its close relatives wheat and barley, it was not a founder crop of Neolithic agriculture, but is considered a secondary domesticate that may have become a crop plant only after a transitory phase as a weed. As a minor crop of only local importance, genomic resources in rye are underdeveloped, and few population genetic studies using genome-wide markers have been published to date. We collected genotyping-by-sequencing data for 603 individuals from 101 genebank accessions of domesticated rye and its wild progenitor *S. cereale* subsp. *vavilovii* and related species in the genus *Secale*. Variant detection in the context of a recently published draft sequence assembly of cultivated rye yielded 55,744 single-nucleotide polymorphisms with present genotype calls in 90 % of samples. Analysis of population structure recapitulated the taxonomy of the genus *Secale*. We found only weak genetic differentiation between wild and domesticated rye with likely gene flow between the two groups. Moreover, incomplete lineage sorting was frequent between *Secale* species either because of on-going gene flow or recent speciation. Our study highlights the necessity of gauging the representativeness of *ex situ* germplasm collections for domestication studies and motivates a more in-depth analysis of the interplay between sequence divergence and reproductive isolation in the genus *Secale*.

W321: Domestication Genomics

Cross-Species Hybridization and the Domestication of Crops

Michael D. Purugganan, New York University, New York, NY

Domestication is a unique co-evolutionary process, and genomic data has provided key insights into the evolutionary dynamics surrounding the origin and dispersal of domesticated crop species. An emerging theme in recent studies has been the importance of interspecies hybridization in the origin and spread of domesticated taxa. Here we discuss key examples in which genomic data has shown the extent to which hybridization and introgression underlies the evolution of key crop species.

W322: Duckweed Research and Applications

The *Wolffia australiana* Genome

Todd Michael[1], Sarah Gilbert[2], Philomena Chu[2], Doug Bryant[3], Florian Jupe[4], Joseph R. Ecker[4], Todd C. Mockler[3] and Eric Lam[2], (1)J. Craig Venter Institute, La Jolla, CA, (2)Department of Plant Biology, Rutgers, the State University of New Jersey, New Brunswick, NJ, (3)Donald Danforth Plant Science Center, Saint Louis, MO, (4)Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA

Wolffia is the smallest and most derived of the Duckweed genera, and it has genomes that span an order of magnitude from *Wolffia australiana* at 375 megabases (Mb) to *Wolffia arrhiza* at 1,881Mb (Wang et al., 2011). We sequenced two *Wolffia australiana* accessions (7733, 8730) leveraging two types of long read and physical mapping technologies: Pacific Bioscience single molecule real-time sequencing (SMRT) and BioNano Genomics optical mapping. Using Kmer genome size estimation both genomes are around 380 Mb with 26% (~100 Mb) of high copy number repeats. However, long terminal repeat (LTR) retrotransposon made up 45% (~175 Mb) of the genome, suggesting that many of the LTRs are lower copy number repeats. We resolved large contigs with ribosomal DNA (rDNA; 18S, 5.8S, 26S, 5S) and identified two high copy number tandem repeats (126 bp and 167 bp) with higher order repeat (HOR) structure consistent with centromeres. Like *Spirodela polyrhiza* that likely only has ~19,000 protein coding genes, *Wolffia australiana* also has a reduced gene repertoire. However, unlike Spirodela that has the lowest known global DNA methylation level (9%), *Wolffia australiana* has the highest global DNA methylation level of any plant tested to date at 75%. Comparative analysis with Spirodela, Lemna, and other monocot genomes revealed unique features of this fast growing aquatic plant. The Wolffia genome will not only augment information for the Duckweed research community, but it will also serve as an important plant model for growth, genome organization, evolution and synthetic biology.

W323: Duckweed Research and Applications
The Lemna Genome
Evan Ernst, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

W324: Duckweed Research and Applications
Cytogenetics of Duckweed
Phuong Hoang, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben,, Germany
Neotenous aquatic duckweeds belong to monocot order Alismatales and comprise 37 species of 5 genera: Spirodela, Landoltia, Lemna, Wolffiella and Wolffia. The genome size and chromosome number of duckweed species vary from 158 Mbp *(Spirodela polyrhiza)* to 1881 Mbp *(Wolffia arrhiza),* and from 2n = 20 to 126, respectively. We focused on Spirodela - the ancestral genus: (1) to establish a reference genome map for *S. polyrhiza* and (2) to study chromosome rearrangements between the only two species *S. polyrhiza* and *S. intermedia,* both with similar genome size (~158 Mbp).
We applied comparative FISH with BACs and pooled BAC probes on seven *S. polyrhiza* clones to address the reason of discrepancies between previous maps of *S. polyrhiza*. Our result revealed no chromosome rearrangements between the seven studied clones and integrated Oxford Nanopore sequencing data to establish an updated reference genome map for *S. polyrhiza*.
The same approach was used to investigate the chromosome homeology and karyotype evolution between *S. polyrhiza* (n=20) and *S. intermedia* (n=18). Two scenarios of karyotype evolution are supposed, considering the ancestral karyotype was similar either to that of *S. polyrhiza* or that *S. intermedia*.
Key words: *S. polyrhiza, S. intermedia,* cytogenetic map, chromosome rearrangements, karyotype evolution.

W325: Duckweed Research and Applications
Epigenomics in Duckweed
Alex Harkess, Donald Danforth Plant Science Center, St. Louis, MO

W326: Duckweed Research and Applications
Sequencing of *Spirodela polyrhiza* miRNAs and their Targets in Diverse Conditions
Paul Fourounjian[1], Jie Tang[2], M. Bahattin Tanyolac[3], Feng Yaping[4], Atul Kakrana[5], Brian Gelfand[4], Min Tu[4], Chris Wakim[1], Dibyendu Kumar[4], Jiong Ma[6], Blake Meyers[5] and Joachim Messing[1], (1)Waksman Institute of Microbiology Rutgers University, Piscataway, NJ, (2)School of Pharmacy and Bioengineering; Chengdu University, Chengdu, China, (3)Ege University, Izmir, Turkey, (4)Genomics Core Facility, Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, (5)Donald Danforth Plant Science Center, St. Louis, MO, (6)School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China
The Lemnaceae family, colloquially called duckweeds, have seen a recent surge in research interest. They have long been a convenient model plant due to their simplified morphology and neotenous lifestyle, and are now a promising crop species that can affordably clean nutrient rich wastewater while providing a biomass that can be used as animal feed or biofuel. *Spirodela polyrhiza* the most basal family member with the smallest genome has two fully assembled genome sequences, a transcriptomic study, and a sRNA-seq experiment establishing it as the reference genome of the family. To further characterize post-transcriptional regulation we analyzed sRNA-seq in 8 different conditions: control, abscisic acid, cold, copper, heat, kinetin, nitrate, and sucrose exposure, and identified 138 mature miRNA sequences and 148 predicted targets. Running degradome sequencing of the same 8 conditions verified 87 miRNAs, 29 of which were novel, and 170 targets based on precisely cleaved target mRNAs. This study helps illustrate the value of degradome sequencing to create a high confidence miRNA-target catalogue and measure miRNA activity.

W327: Duckweed Research and Applications
Establishing a Genomics-Based Platform as a Model System to Functionally Characterize Plant Microbiome

Eric Lam[1], Kenneth Acosta[1], Sarah Gilbert[1], Weijuan Huang[1], Todd P. Michael[2] and Sarah Lebeis[3], (1)Department of Plant Biology, Rutgers, the State University of New Jersey, New Brunswick, NJ, (2)J. Craig Venter Institute, La Jolla, CA, (3)University of Tennessee, Knoxville, TN
The aquatic habitat and simple morphology of duckweeds presents an excellent model system to study the complex interactions between microbial communities and their plant hosts. We have recently isolated over one hundred strains of bacteria associated with various species of duckweed and to date have produced complete genome sequences for 33 with another 17 draft genomes to be completed by the end of 2017. In addition, validated and comprehensive reference genome sequence and other genomic tools have recently been published or are coming on line for several species of duckweeds. Leveraging these resources, we are systematically comparing the ability of these bacteria strains to colonize and promote growth in gnotobiotic duckweed strains as well as *Arabidopsis thaliana* plants in order to identify highly conserved and species-specific interaction pathways between microbes and plants. As a complementary approach, using a bioinformatic pipeline and analytical chemistry techniques, we are mining our sequenced duckweed associated bacterial genomes to reveal the assortment of potential bacterial biosynthetic gene clusters. This will help identify candidate signal molecules that can be produced by these bacteria to mediate bacteria-plant interactions as well as with other microbes in the plant microbiomes. Our work will contribute to a functional understanding of the plant microbiome's impact on plant growth and provide new knowledge on the underlying mechanisms at the community level.

W328: Ecological Genomics
Adaptive Altitudinal Variation in Genome Size in Maize and Teosinte
Jeffrey Ross-Ibarra, University of California, Davis, CA

W329: Ecological Genomics
Gene Expression Plasticity and the Persistence of Plant Species across Environmental Variation
Hannah Marx, Katrina Dlugosch and Michael S. Barker, University of Arizona, Tucson, AZ

W330: Ecological Genomics
Inferring Constraint from Comparative Genomic Studies of Local Adaptation
Sam Yeaman, University of Calgary, Calgary, AB, Canada
TBA

W331: Ecological Genomics
Harnessing the Power of RADseq for Ecological Genomics
Kimberly Andrews, University of Idaho; University of Washington, Moscow, ID, Jeffrey Good, University of Montana, Missoula, MT, Michael Miller, Department of Animal Science, University of California, Davis, Davis, CA, Gordon Luikart, University of Montana, Polson, MT and Paul Hohenlohe, University of Idaho, Moscow, ID
The development of restriction site-associated DNA sequencing (RADseq) has been deemed among the most important scientific breakthroughs in the past decade. RADseq has fueled conservation genomic studies of hundreds of non-model organisms by harnessing the massive throughput of next-generation sequencing. This technique allows both discovery and genotyping of hundreds to thousands of polymorphic markers in coding and non-coding regions using a time- and cost-efficient approach, and requires no prior genomic information. These advantages have led to an explosion of studies of non-model organisms investigating population structure and delineating conservation units, and estimating gene flow and dispersal, demographic history, genomic diversity, inbreeding, relatedness, and phylogenetics. RADseq also enables investigation into questions that were virtually intractable for non-model species prior to the advent of next-generation sequencing technologies, including questions regarding the genomic basis of fitness, adaptation, and phenotypic variation. However, as with any new method, researchers should exercise caution when using RADseq data. Several sources of bias and error exist for RADseq data, and researchers should consider these carefully when designing and implementing a RADseq experiment. In addition, numerous variant RADseq methods have emerged in recent years (e.g. GBS, ddRAD, 2b-RAD, ezRAD), and it can be difficult for researchers to choose an approach. We provide an overview of the types of research questions RADseq can be used to answer, highlighting examples from the literature for ecological applications with non-model organisms. We provide a brief overview of the technical differences among the RADseq methods, and discuss the types of error and bias to which RADseq is susceptible. We also discuss recent controversial criticisms of RADseq, and key considerations for choosing among the growing number of methods when designing a study.

W332: Ecological Genomics
Testing Drivers and Constraints to the Diversification in Neotropical Palms (Geonoma) Using NGS Data
Margot Paris[1], Oriane Loiseau[2], Jonathan Rolland[2], Ingrid Olivares[3], Marylaure de la Harpe[4], Anna Weigand[3], Maria Jose Sanin Perez[5], Fabian Gregorio Mejia Franco[5], Jaqueline Hess[4], Michael Kessler[3], Nicolas Salamin[2] and Christian Lexer[4], (1)University of Fribourg, Fribourg, Switzerland, (2)University of Lausanne, Lausanne, Switzerland, (3)University of Zurich, Zürich, Switzerland, (4)University of Vienna, Vienna, Austria, (5)Facultad de Ciencias y Biotecnología, Medellín, Colombia
Understanding why the Neotropics host the greatest plant diversity on Earth is a question that has puzzled evolutionary biologists for a long time and is still a vibrant field of research. The genus Geonoma is one of the most diverse palm genera in the Neotropics with 68 species distributed from sea level to >3'000m in Central and South American rain forests. Species delimitation has proved challenging because of high intraspecific morphological variation and 20% of Geonoma species are in fact considered as species complexes. For these reasons, the genus Geonoma offers a powerful model to test the relative roles of geography, dispersal limitations, ecological and intrinsic factors in driving divergence of palm populations and species.
We used a target sequencing approach to sequence thousands of single copy nuclear DNA regions for more than 800 individuals including 95% of the described species in the genus Geonoma, and population samples of two widespread species complexes. Using analyses covering both

micro- and macro-evolutionary time scales, we investigated the biogeographical history as well as the drivers of diversification of *Geonoma* by exploring several intrinsic and extrinsic factors that could have played a role in the evolution of the group. Our results contribute to broadening the general knowledge about the emergence of the current outstanding neotropical plant diversity.

W333: Ecological Genomics
Pan-Genome of Silver Birch (*Betula pendula*)
Sitaram Rajaraman, University of Helsinki, Helsinki, Finland
Silver birch (*Betula pendula*) is an important pioneer tree species that grows in boreal forests across Europe and Asia. We recently assembled the reference genome and carried out a population genomics study on the species. Selective sweeps were found acting mostly on the genomic regions originating from whole genome duplications, whereas most recent tandemly duplicated genome regions were depleted of sweeps. Here we study the standing pool of genomic structural variation among silver birches by estimating the pan-genome of silver birch from individuals sampled from 12 different locations across Eurasia. The individual-specific reads were assembled into scaffolds using SPAdes assembler. After removing organellar and bacterial contaminants the de novo assembly added 6 Mbp of contigs per individual, containing approximately 2000 gene fragments predicted by AUGUSTUS. Tandemly duplicated genes were enriched among the individual-specific genomes, corroborating recent results that copy-number variation forms the largest pool of standing variation among individuals. Gene ontology categories related to environmental responses were found to be enriched among the dispensable part of the genome.

W334: Ecological Genomics
Lightning Talk: Demogenetic Models: A Simulation Approach to Explore Future Trends in Population Genetic Patterns
Stephen Rice, San Diego State University, San Diego, CA

W335: Engineering NUE
Manipulation of Amino Acid Transport Processes to Improve Plant Nitrogen Use Efficiency
Mechthild Tegeder, Washington State University, Pullman, WA

W336: Engineering NUE
Below-Ground Approaches to Improve Nitrogen Use Efficiency in Oilseed Rape
Julien Louvieaux, CARAH/HEPH-Condorcet and Université libre de Bruxelles, Ath, Belgium and Christian Hermans, Université Libre de Bruxelles, Lab of Plant Physiology and Molecular Genetics, Brussels, Belgium
In agriculture, synthetic nitrogen (N) fertilizers come with high costs and give rise to groundwater and air pollution. Thus, improving the nitrogen use efficiency (NUE) is essential to ensure the economic and environmental sustainability of crop production. While research has been mainly conducted on N assimilation and remobilization processes in aerial, breeders often consider the root organ as a black box. Yet, optimizing the degree of root branching for exploring a large soil volume may contribute to higher N uptake. Oilseed rape has a poor NUE with a low ratio of seeds produced per N unit applied, around half that for cereals. Our aim is to understand the elaboration of root traits in oilseed rape and how it impacts on nitrate uptake. We will provide an example of high throughput hydroponic screen of root morphology with a genetic diversity panel grown at divergent N supplies. Key observations are the considerable degree of variability in the root morphological traits among genotypes and the absence of tradeoff between profuse root branching and shoot biomass production. Furthermore, root morphological traits observed at a young developmental stage in laboratory conditions, positively correlate with seed N and protein concentrations measured in the field.
Mineral nitrogen (N) is the quantitatively most important nutrient in cropping systems. However, a considerable N fraction is lost through runoffs with detrimental consequences for the environment and human health. One way to reduce N fertilizer input is to breed for crops with better Nitrogen Use Efficiency (NUE). Increasing the plant N uptake by optimizing the degree of root branching for exploring a larger soil volume in search of the mobile nitrate resource may contribute to that purpose. Rapeseed (*Brassica napus* L.) is a major oil crop showing a poor NUE, which makes its production highly dependent on N fertilization. Our aim is to understand the genetic control of root system architecture and how it is impacted by N nutrition. We are currently exploiting the genetic diversity of root morphology in a panel of >400 double haploid accessions. Seedlings grew for seven days on vertical germination paper imbibed with a nutrient solution containing low or moderate nitrate concentrations. Seedlings cultivated with low nitrate supply developed more numerous and longer lateral roots than those grown with high supply. There was a wide diversity within the accessions for the root morphological traits. We are identifying genomic regions associated those traits by performing Genome Wide Association Studies. That information is complemented with root transcriptome sequencing data which will allow detecting gene copies differentially expressed between accessions and between N nutrition conditions. A measurable outcome in the mid-term will be to provide genetic markers for selecting new crop genotypes for smart farming.

W337: Engineering NUE
Genetic Gains of Nitrogen Use Efficiency - Empirical Evidence and Future Directions in Breeding Research
Andreas Stahl, Paul Vollrath, Mara Pfeifer, Benjamin Wittkop and Rod Snowdon, Justus Liebig University, Giessen, Germany
Nitrogen (N) is the most important plant nutrient fertilized to agricultural crops in order to achieve high yield and product quality. Over many decades increased N input was the key driver of enhanced crop yields. However, on the flip side of the coin, these inputs were associated with massive damage to ecosystems by greenhouse gas emissions and pollution of waterways. Unused N also represents an economic loss for farmers, meaning that increased N use efficiency is of high relevance both from an ecological and an agro-economic point-of-view.
In two year, multi-location field trials, we investigated elite oilseed rape varieties released to the European market during the last two decades for nitrogen uptake und utilisation efficiency traits under contrasting N fertilization levels. Furthermore, a large-container phenotyping system, comprising 120 wheelie-bins with 90 cm soil depth, was used to assess genetic differences in root morphology associated with N inputs.
The results provide strong empirical evidence for a genetic-driven improvement of seed yield and seed quality traits that has indirectly led to considerable increases in N use efficiency. In summary, analysis of more than 40 macro-physiological traits give strong hints of traits

associated to genetic improvement. For example, hybrid varieties show a higher average number of siliques than open-pollinated (OP) varieties, whereas modern varieties (hybrids and OP) outperform older varieties in the number of seeds per silique. In addition, we found interesting effects of unintended co-selection. The presentation will discuss how overcoming negative trait associations might allow further gains in N use efficiency.

W338: Engineering NUE

From Arabidopsis to Crops: The *Qqs* Orphan Gene Modulates Carbon and Nitrogen Allocation across Species

Ling Li, Mississippi State University, Starkville, MS

The *Arabidopsis thaliana QQS* orphan gene modulates carbon and nitrogen allocation to protein and starch[1]. Ectopic expression of *QQS* increases protein content[2] in leaf and seed in soybean Williams82[3], in soybeans with different high-/low-protein levels, and in rice and corn[4]. The QQS protein binds to a transcriptional regulator in Arabidopsis and its soybean, rice and corn homologs: Nuclear Factor Y subunit C4 (NF-YC4). Overexpression of NF-YC4 mimics *QQS*-overexpression phenotype, increasing protein content and decreasing carbohydrate[4]. Little is known about the functional significance of the species-specific orphan genes[1, 4, 5]. *QQS* transcript levels are altered in plants under stresses and in mutants of genes involved in all sorts of stress responses, indicating that *QQS* may integrate primary metabolism and environmental perturbations[6]. Our research reveals the core of a previously undefined network in which QQS participates[4], and opens a new non-transgenic strategy to create high-protein crops[7]. Deficiency in dietary protein is globally one of the most severe health problems; the ability to optimize protein productivity of plant-based foods has far-ranging impact on world health and sustainability. Our research presents QQS as a model plant orphan gene regulating plant metabolism and adaptation to environment, and illustrates an example of basic research in Arabidopsis applied in agriculture.

References

[1]Li et al. PJ (2009).
[2]Li et al. U.S. Patent 9157091 (2012).
[3]Li et al. PBJ (2015).
[4]Li et al. PNAS (2015).
[5]Jones et al. Frontiers Plant Science (2016).
[6]Arendsee et al. TIPS (2014).
[7]Li et al. U.S. Patent Office 62/244,131 (2015).

W339: Engineering NUE

Screening and Validation of Candidate Genes for Nitrogen Use Efficiency in Rice

Raghuram Nandula, School of Biotechnology, GGS Indraprastha University, New Delhi, India

W340: EPIC: the Plant Epigenome Project

Using Sequence Capture Bisulfite Sequencing to Uncover Variability in the Maize Methylome

Peter Crisp[1], Zhaoxue Han[2], Scott Stelpflug[3], Shawn Kaeppler[4], Qing Li[5] and Nathan M. Springer[1], (1)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (2)Northwest A&F University, Shaanxi, China, (3)Monsanto Company, Huxley, IA, (4)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI, (5)National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Hubei, China

W341: EPIC: the Plant Epigenome Project

GC-Rich Coding Sequences Reduce Transposon-like, Small RNA-Mediated Transgene Silencing in *Arabidopsis* and Maize

Lyudmila Sidorenko[1], Tzuu-fen Lee[2], Aaron Woosley[1], William Moskal[1], Scott Bevan[1], P. Ann Owens Merlo[1], Terence Walsh[1], Xiujuan Wang[1], Staci Weaver[1], Todd Glancy[1], PoHao Wang[1], Xiaozeng Yang[1], Shreedharan Sriram[1] and Blake Meyers[2], (1)DOW AgroSciences LLC, Indianapolis, IN, (2)Donald Danforth Plant Science Center, St. Louis, MO

The molecular basis of transgene susceptibility to silencing is poorly characterized in plants, thus we evaluated several transgene design parameters as means to reduce heritable transgene silencing. Analyses of *Arabidopsis* plants with transgenes encoding a microalgal polyunsaturated fatty acid (PUFA) synthase revealed that small RNA (sRNA)-mediated silencing, combined with the use of repetitive regulatory elements, led to aggressive transposon-like silencing of canola-biased PUFA synthase transgenes. Diversifying regulatory sequences and using native microalgal coding sequences (CDSs) with higher GC content improved transgene expression and resulted in a remarkable trans-generational stability associated with reduced accumulation of sRNAs and DNA methylation. To assess the translatability of our observations to other transgenes and plant species, we investigated the influence of CDS composition on the expression of transgenes encoding crystal proteins from *Bacillus thuringiensis* in maize. Our results from maize supported the observations from *Arabidopsis* and demonstrated a significant role of GC content in increasing transgene expression and protein accumulation. Therefore, our observations are applicable in a wide range of plant species and provide simple and cost-effective steps for improving transgene expression, even without attenuating endogenous RNA silencing mechanisms.

W342: EPIC: the Plant Epigenome Project

Chromosome Dosage and Segregation Mediated Small RNA

Rob Martienssen, HHMI-GBMF Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

W343: EPIC: the Plant Epigenome Project

Defining the Basis for Selective, Multi-Megabase NOR Silencing (Nucleolar Dominance) *via* Chromosome Engineering

Craig Pikaard, Indiana University/HHMI, Bloomington, IN

In eukaryotic genomes, hundreds of ribosomal RNA (rRNA) genes are clustered at multi-megabase loci known as nucleolus organizer regions (NORs). Despite being nearly identical in sequence, we have shown that rRNA genes are differentially expressed during development, with some rRNA gene subtypes expressed constitutively and others being inactivated via repressive chromatin modifications. This differential expression of rRNA genes is the molecular basis for the classic epigenetic phenomenon known as nucleolar dominance. A major unanswered question is: how can rRNA genes of essentially the same sequence be differentiated from one another such that some are activated and others are silenced?
In the *Arabidopsis thaliana* ecotype (strain) Columbia (Col-0), we recently defined more than a dozen rRNA gene subtypes based on SNPs or short indels. Half of the subtypes are selectively silenced during development via repressive chromatin modifications. By exploiting natural variation in the rRNA gene subtypes of different ecotypes, we used a genetic mapping approach that revealed that silenced rRNA gene subtypes map to the NOR on chromosome 2, *NOR2,* whereas active rRNA gene subtypes map to *NOR4*, on chromosome 4. We also identified a mutant in which millions of basepairs of *NOR4* were replaced by the corresponding sequences of *NOR2*. In this mutant line, the *NOR2*-derived rRNA genes translocated to the position of *NOR4* are no longer silenced. Collectively, these results indicate that selective rRNA gene silencing is dependent on chromosomal context, not sequence variation within individual rRNA genes.
Using a combination of chromosome engineering strategies, including induced chromosome deletions, segmental inversions, and reciprocal chromosome translocations that move the ~4 Mb NORs to other chromosomes, we are zeroing in on the cis-acting chromosome intervals necessary for *NOR2-NOR4* communication and selective *NOR* silencing.

W344: Equine 1
Epigenomic Diversity in the Mammalian Brain
Chongyuan Luo, Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA
The epigenome is an ensemble of chemical modifications of DNA and chromatin that modulates the activities of the genome, which plays instrumental roles in gene regulations in healthy and disease tissues. We have developed a sequencing based method to profile DNA methylation (mC) at single-base resolution across the whole mammalian genome. Using this approach, we found high levels of non-CG methylation (mCH) at locations throughout the genome in human and mouse brains. Mammalian brains contain numerous types of neurons that can be distinguished by their morphological, physiological and functional properties. Using an epigenomics dataset produced from nuclei of specific neuronal types purified by INTACT approach, we identified abundant epigenomic and gene-expression differences across three excitatory and inhibitory neuron types in adult mouse. Extending cell type specific mC analysis to all brain cell types requires unbiased single cell mC profiling. We developed a new method for high-throughput single nucleus methylcytosine sequencing (snmC-seq). Using >6,000 single cell methylomes, we identified 16 mouse and 21 human neuron types in the frontal cortex. The results suggest expanded neuronal diversity in the human cortex, which is consistent with the finding of a human-specific inhibitory neuron subpopulation. Our epigenomic analysis allowed the prediction of approximately ~500,000 cell type specific regulatory elements for mouse or human neuron types. Currently we are generating single cell methylomes for cell type classification of the whole mouse brain. Our single nucleus methylome approach can be applied to all human tissues for producing epigenomic profiles to inform the human cell atlas.

W345: Equine 1
EquCab3: Assembly Complete
Edward S. Rice[1], Michael S. DePriest[2], Ludovic Orlando[3], Ernest Bailey[4] and Theodore S. Kalbfleisch[2], (1)Department of Biomolecular Engineering, UC Santa Cruz, Santa Cruz, CA, (2)University of Louisville, Louisville, KY, (3)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, (4)University of Kentucky, Lexington, KY
The current reference assembly for the domestic horse *Equus caballus* was published in 2009. This assembly used Sanger (first-generation) sequencing along with bacterial artificial chromosomes (BACs) to produce an assembly with 112kb contig N50 and 46Mb scaffold N50. Since 2009, there have been enormous advances in sequencing technology, bringing the cost and time required for sequencing down by more than a factor of 10,000. We used the existing Sanger sequence data along with several of these new technologies to assemble a new reference genome. Illumina HiSeq short reads increased average assembly read depth six-fold, resulting in fewer mis-calls. We used CHiCago and Hi-C long-insert libraries to improve scaffold assembly, nearly doubling the scaffold N50 and increasing the amount of sequence assigned to chromosomes. Gap-filling with PacBio long reads greatly increased contig sizes. We used a 10x Chromium library to identify and phase variants. The final assembly has 4,493kb contig N50, 85Mb scaffold N50, and 70Mb more sequence assigned to chromosomes. It will be publicly available by the end of 2017. This talk will discuss the various approaches and software pipelines we used to leverage the strengths of our different data types to create the best possible final assembly, including those that did not work, with the hope that this information will be useful to those producing their own reference genomes using these new technologies.

W346: Equine 1
Update on the Equine FAANG Initiative: How the Community Is using the Biobank
Erin N. Burns[1], Michael J. Mienaltowski[1], Theodore S. Kalbfleisch[2], Jessica L. Petersen[3], Rebecca R. Bellone[4] and Carrie J. Finno[1], (1)University of California, Davis, Davis, CA, (2)University of Louisville, Louisville, KY, (3)University of Nebraska, Lincoln, NE, (4)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA
The Functional Annotation of the Animal Genome (FAANG) project aims to identify functional regulatory elements in animal genomes in both sexes and across multiple stages of development. For the horse, a biobank of tissues and cells was created, which includes 80 tissues and three cell types from two adult Thoroughbred mares. Last year, we reported on the creation of this biobank to be used by the equine community in the functional annotation of the genome in accordance with FAANG Consortium guidelines. Since the Plant & Animal Genome conference in January of 2017, the FAANG biobank was used in the generation of 7 different datasets. In collaboration with the equine genetics community, reduced representation bisulfite sequencing, microbiome sequencing, and chromatin immunoprecipitation sequencing have been conducted. In addition, to date, RNA has been isolated from 34 tissues of both biobank mares using Trizol® chloroform phase separation and clean up via

Zymo Research columns. Extracted RNA had an average RIN score of 8.2 prior to stranded library preparation. mRNA and smRNA were sequenced on the Illumina HiSeq 4000. Data are publicly available and linked to the phenotype information within the FAANG database. Analysis of the RNA-seq data is ongoing to determine specific signatures of each tissue type.

W347: Equine 1

Strategies for Finding Previously Unannotated Protein-Coding Genes in an Updated Mammalian Reference Genome, Featuring EquCab3

Michael S. DePriest[1], James N. MacLeod[2] and Theodore S. Kalbfleisch[1], (1)University of Louisville, Louisville, KY, (2)Department of Veterinary Science, University of Kentucky, Lexington, KY

As sequencing technology continues to progress, reference genomes assembled using legacy technology are constantly re-evaluated using new types of sequence data. New sequencing methods allow researchers to fill gaps and fix errors, resulting in better, more complete reference genomes. In the case of EquCab3, the forthcoming new reference genome for the domestic horse, *Equus caballus,* the filled gaps contain many genes of immediate interest to the equine research community. The presence of protein-coding genes in newly characterized regions present in EquCab3 has been demonstrated previously using a BLAST-based approach. However, due to similarities among protein sequences in the database, the resolution of this analysis was limited and the results contained many false positives. In the current study, we evaluated different strategies for finding genes in a mammalian genome assembly using EquCab3 as the primary example. We investigated the effect of different experimental designs, including changing score thresholds, masking the genome prior to searching, and searching transcripts rather than the genome sequence. We used RNA-seq data to validate our gene predictions and determine gene structure. Finally, we compared our results from the horse genome with other recently updated mammalian reference genomes.

W348: Equine 1

De novo Assembly of the Equine MHC Region using Linked-Read Genome Sequencing

Donald Miller[1], Peter Schweitzer[1], Theodore S. Kalbfleisch[2], James N. MacLeod[3], Ernest Bailey[4] and Douglas Antczak[1], (1)Cornell University, Ithaca, NY, (2)University of Louisville, Louisville, KY, (3)Department of Veterinary Science, University of Kentucky, Lexington, KY, (4)University of Kentucky, Lexington, KY

The Major Histocompatibility Complex (MHC) has been associated with more disease conditions than any other region of the equine genome, including susceptibility to equine sarcoid tumors, uveitis, insect bite hypersensitivity, and abortion. However, sequencing and assembly of the MHC presents a tremendous challenge in genomics because of the region's high degree of polymorphism, gene duplication, and structural variation between MHC haplotypes. Because of these features, it is not possible to assemble conventional short read (e.g. Illumina) sequencing data into reliable MHC region haplotypes. Here we produced a *de novo* assembly of the equine MHC using the 10x Genomics™ Chromium™ linked read gel-bead system. For input DNA we used Twilight, the DNA donor of the NCBI horse genome sequence, and compared our results to the EquCab2 assembly. We obtained 120 GB of sequence (~50x coverage) that was assembled *de novo* at the Cornell Biotechnology Resource Center using the 10x Genomics Supernova™ assembler program. This produced six long contigs in the MHC region with very few gaps, allowing us to correctly order the class I and class II genes on the ELA-A3 haplotype and to obtain high fidelity full length genomic sequences for all of those genes. The 10X Genomics™ platform produced data that is virtually identical to that found in the current NCBI EquCab2 equine genome assembly of the MHC region. These technologies should be useful for assembly of other equine MHC haplotypes, and for investigating the mechanisms of MHC disease associations in the horse.

W349: Equine 1

Determining Copy Numbers of Key Fertility Genes in the Equine Y Chromosome

Caitlin Castaneda[1], Andrew Hillhouse[2], Jan Janecka[3], Rytis Juras[1] 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, (3)Duquesne University, Pittsburgh, USA, Pittsburgh, PA

The mammalian Y chromosome houses essential male fertility genes. Previous mammalian studies provide evidence that copy number variations (CNVs) in these genes affect male reproductive phenotypes. Here, we initiate CNV analysis of multicopy genes in the horse male specific Y (eMSY) using digital droplet-PCR. Our goal is to determine a normal range of CNVs for all multicopy genes (total 15, of which 8 are novel and horse specific) in a cohort of normal male horses of diverse breeds. The data will form a baseline for the discovery of variants associated with reduced stallion fertility. Initial studies involved 64 normal males of seven breeds, and three infertile stallions. The breeds studied were Thoroughbred (including the Thoroughbred stallion *Bravo* - the DNA donor of the eMSY reference sequence), American Quarter Horse, Arabian, Suffolk, Caspian Pony, Mongolian Native, Icelandic, Shetland Pony, Mongolian and Lipizzaner. TaqMan assays were designed for five multicopy genes: *TSPY*, *RBMY*, *UBA1Y*, and *HSFY*, and for the *SRY* gene. Using autosomal myostatin (*MSTN*) as the single copy reference gene, multiple ddPCR iterations provided consistent CNs amongst all seven breeds and confirmed that *SRY* is a single copy gene. None of the studied genes showed different CNs in infertile stallions. However, for all genes we observed less copies by ddPCR compared to the Y reference assembly. Future studies include analysis of the remaining multicopy genes, including equine specific transcripts such at *ETSTY2*. Gaining information on eMSY CNVs would enable improving the assembly of eMSY multicopy regions, and can potentially lead to the discovery of CNVs affecting stallion fertility.

W350: Equine 1

Annotation of a Structural Polymorphism in the LATH Gene Region of the Equine and Related Species

Katelyn M. Palermo, Kelsey Yates, Laura Patterson Rosa and Samantha A. Brooks, University of Florida, Gainesville, FL

Cooling mechanisms are vital to thermoregulation and survival of all mammals. Equids and higher primate species have uniquely evolved to produce sweat as a primary means of cooling. Unlike humans, equine sweat is high in protein, specifically Latherin, a surfactant protein which increases pelt wetting and aids in evaporation. Encoded by the gene *LATH*, the locus is a paralog to human Bactericidal Permeability-Increasing Protein Family A Member 4 (*BPIFA4*) and representative of the rapidly evolving PLUNC (palate, lung and nasal epithelium clone)

protein family. A previous study observed a polymorphic copy number variant (CNV) encompassing the *LATH* gene region of the domesticated horse and varying in copy number among horses of diverse breeds. In the horse, this particular structural polymorphism is under positive selection, possibly as a result of adaptive evolution for survival, or due to human-selection throughout domestication. Here we will report the results of quantifying exact copy numbers of five BPI genes in multiple equid species and across the Perissodactyl family which includes the rhinoceros and tapir. Our use of the QX200™ Digital Droplet™ qPCR and Illumina's Eco Real-Time PCR has allowed us to generate Cq numbers that illustrate a large difference between gene expression of *LATH* in horses (25.73446529) versus that of rhinos (30.15830968). Future work will include testing such individuals using the Qualitative Intradermal Terbutaline Sweat Test (QITSTs) and quantification of the amount of *LATH* protein produced in sweat samples by western blot.

W351: Equine 1

Transitioning from Illumina to Affymetrix: Platform Concordance and Lessons Learned

Heather M. Holl and Samantha A. Brooks, University of Florida, Gainesville, FL

Reliable, high quality genotypes are vital for genomic studies. In the horse, two Illumina BeadChips, the Equine SNP50 and SNP70 arrays, have been successfully used in a variety of studies involving domestication, population genetics, genetic selection, and trait mapping. More recently, an Affymetrix 670k Axiom array was developed in order to increase marker coverage for genomic studies. However, concordance rates between legacy and whole genome sequencing genotypes have not yet been evaluated. We obtained Axiom 670k genotype data from 767 horses for use in multiple projects. DNA was extracted from multiple sample types, and arrays were genotyped by two providers in four batches. Initial quality control analyses indicated problematic batch effects across service labs, thus we evaluated more stringent filtering criteria. Genotypes from alternate platforms were available for a subset of horses: seventeen horses with Axiom 670k and Illumina SNP50, six horses with Axiom 670k and Illumina SNP70, and eight horses with Axiom 670k and Illumina HiSeq. Overall genotype concordance rates were determined separately for each Affymetrix defined SNP cluster classification. Additionally, we called variants from Illumina HiSeq lanes from twenty-one horses using the GATK HaplotypeCaller pipeline and identified variants incorrectly annotated as bi-allelic SNPs in the Axiom 670k reference files. We will present recommendations for genotyping and quality control for Affymetrix based projects.

This study was made possible in part by NPRP Grant 6-1303-4-023 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

W352: Equine 1

Selective Sweep Mapping using a Unique Nordic Horse Model Revealed EDN3 as a Candidate Gene for Harness Racing Performance

Kim J. Fegraeus[1], Brandon D. Velie[1], Jeanette Axelsson[1], Rachel Ang[2], Natasha A Hamilton[2], Leif Andersson[3], Jennifer R.S. Meadows[3] and **Gabriella Lindgren**[1], (1)Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, (2)School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia, (3)Dept of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

The unique origin of the Swedish-Norwegian Coldblooded trotter makes the breed ideal for studying the genetics of racing performance. By comparing the genomes of Coldblooded trotters, Standardbreds and North-Swedish draught horses for a large number of single nucleotide polymorphisms (SNPs), the aim of the study was to identify genetic regions that may be under selection for racing performance. A fixation index (Fst) analysis was performed and sliding window Delta Fst values were calculated across the breeds using data generated from the equine SNP50K array (Coldblooded trotters, n=11; North-Swedish draught horses, n=19; Standardbreds, n=12). The average Delta Fst was calculated for windows of five SNPs and the five top windows, where the Fst between Coldblooded trotters and Standardbreds was low and the Fst between Coldblooded trotters and North-Swedish draught horses was high, were chosen for further investigation. Associations between the top SNP from each region and harness racing performance was analyzed in 400 raced Coldblooded trotters. One SNP (g.22:45748491C>T) showed significant associations with racing performance, with the CC genotype appearing to be negatively associated with the majority of performance traits tested. In addition, four consecutive SNPs spanning 1.5 kilobases (kb) all showed significant associations with harness racing performance. Further, the SNP identified was genotyped in 1,634 horses of 14 different breeds. The frequency of the TT genotype was high in breeds typically used for racing and show jumping while the frequency of the CC genotype was high in most pony breeds and draught horses. The closest gene in the top region identified was the *Endothelin3* gene (*EDN3*).

W353: Equine 1

Utilization of the Foal Fecal Microbiota to Understand Gut Flora Transitions: From Birth to Weaning

Michael J. Mienaltowski[1], Ubaldo De La Torre[2], John D. Henderson[2], Kathleen Furtado[2], Madeleine Leahy[2] and Elizabeth A. Maga[2], (1)University of California, Davis, Davis, CA, (2)University of California Davis, Davis, CA

Few studies have been conducted to understand the gut microbiota of foals. A healthy gastrointestinal (GI) tract with a properly established microbiota is necessary for a foal to develop into a healthy weanling. We hypothesized that the establishment of the gut flora in foals is directly correlated to diet and environmental exposures, and could be assessed from analyses of the fecal microbiota during GI transition. Fecal samples from 42 sets of foals and mares were collected at multiple time points ranging from birth to weaning. Bacterial DNA was isolated and the V4 domain of bacterial 16S rRNA genes was amplified and then applied to next-generation sequencing to characterize the fecal microbiota in each foal. Microbial taxonomic assignment and relative abundance determination were performed using QIIME (Quantitative Insights Into Microbial Ecology). Specific comparisons in microbial populations were made using LefSe (LDA Effect Size). STAMP (STatistical Analysis of Metagenomic Profiles) was used to characterize functional roles of microbial populations in host biology. We found that bacterial population compositions followed a pattern throughout the early life of the foal in an age-dependent manner. Moreover, we were also able to recognize differences in microbial populations amongst diarrheic foals. Future efforts will better discern the effects of lesser abundant bacterial populations that may be just as essential to GI health, as well as characterize microbial populations at additional time points to further investigate the fecal microbiota as the foal transitions to weaning.

W354: Equine 2

Prevalence of the Mutation Conferring Susceptibility to Immune-Mediated Myositis in Seven Performance Subgroups of American Quarter Horses

Giuliana M. Gianino[1], Stephanie J. Valberg[2], Sudeep Perumbakkam[2], Marisa L. Henry[2], Keri Gardner[2], M. Cecilia Penedo[3] and Carrie J. Finno[1], (1)University of California - Davis, Davis, CA, (2)Michigan State University, East Lansing, MI, (3)Veterinary Genetics Laboratory, University of California, Davis, Davis, CA

Equine immune-mediated myositis (IMM) is a disease in the American Quarter Horse (QH) breed characterized by recurrent, rapid-onset muscle atrophy and lymphocytic infiltration of myofibers, often following an episode of respiratory infection, vaccination, or rhabdomyolysis. Recently, a genome-wide association study and whole genome sequencing identified a functional variant which was significantly associated with the IMM phenotype. The purpose of this study was to estimate the frequency of this variant suspected to confer susceptibility to IMM in seven elite performance subgroups of QHs. Selection of elite QHs for this study was based on records obtained of the most competition points (halter discipline) or money earned (all other disciplines) in the last three years by individual horses registered with the American Quarter Horse Association (AQHA). With permission from the AQHA, top-performing horses from the barrel racing (42), cutting (43), halter (50), racing (36), reining (35), Western pleasure (45), and working cow (36) disciplines were genotyped for the IMM variant using pyrosequencing and a basic variant frequency was calculated in each subgroup. Of the total 287 elite performance horses genotyped, 91% were wildtype, 9% were heterozygotes, and 1 individual was homozygous for the novel variant. Frequency of the IMM variant was highest in reining (0.114), working cow (0.097), and halter (0.080) horses, followed by cutting (0.047) and Western pleasure (0.011) horses. The IMM variant was not observed in horses from the barrel racing or racing disciplines in this survey.

W355: Equine 2

Additional Evidence Supports *DDB2 T338M* As the Genetic Risk Factor for Ocular Squamous Cell Carcinoma in Horses

Moriel Hana Singer-Berk[1], Mary Lassaline[2], Kelly Knickelbein[3], Nikhil Joshi[4], Monica T. Britton[4], Matthew Settles[4], Zachary T. Lounsberry[1], Christopher Reilly[5], Ellison Bentley[6], Alison Clode[7], Cord Drögemüller[8], Lucia Unger[9], Vincent Gerber[9], Tammy Miller-Michau[10] and Rebecca R. Bellone[1], (1)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (2)UC Davis School of Veterinary Medicine, Davis, CA, (3)Veterinary Medical Teaching Hospital, University of California-Davis, Davis, CA, (4)UC Davis Genome Center, Davis, CA, (5)Insight Veterinary Specialty Pathology, Austin, TX, (6)University of Wisconsin-Madison, School of Veterinary Medicine, Madison, WI, (7)Port City Veterinary Referral Hospital, Portsmouth, NH, (8)Institute for Genetics University of Bern, Bern, Switzerland, (9)Swiss Institute of Equine Medicine, University of Bern and Agroscope, Bern, Switzerland, (10)Bluepearl Veterinary Partners, Tampa, FL

Squamous cell carcinoma (SCC) is the most common periocular cancer, and the second most common tumor in horses. A missense mutation in *damage specific DNA binding protein 2 (DDB2, p.Thr338Met),* within an associated 483 kb haplotype on ECA12, was identified as a recessive genetic risk factor for limbal SCC in Haflinger horses. To determine if this variant also contributes to risk for SCC in other ocular and urogenital locations and in other breeds commonly diagnosed with SCC, it was genotyped in Haflingers (N=110), Belgians (N=36), Appaloosas (N=42), and Arabs (N=25). These horses were diagnosed with SCC or were classified as unaffected based on clinical examination. A significant association was detected in Haflingers and Belgians when considering all ocular locations ($P= 2.80 \times 10^{-16}$ and $P= 1.98 \times 10^{-5}$ respectively). This association was not perfectly concordant with phenotype, as 24% of the Haflingers and Belgians affected with ocular SCC were not explained by homozygosity for this mutation. Therefore, high throughput sequencing data from six Haflingers (four cases and two controls) were analyzed to identify additional variants for investigation. Sixty-seven polymorphisms from the previously associated locus on ECA12 were genotyped in 103 Haflingers. Analysis revealed that no other variant from this locus explained the genetic risk better than *DDB2 p.Thr338Met* ($P = 7.83 \times 10^{-15}$). These data provide further support that the *DDB2* variant is a contributing risk factor for ocular SCC in Haflingers and Belgians and can be utilized as a diagnostic tool to inform clinical and breeding decisions.

W356: Equine 2

RNA Sequencing Reveals HCN4 As a New Mediator of Airway Hyper-Responsiveness in a Spontaneous, Naturally Occurring Equine Asthma Model

Courtney Hunter[1], Lais Costa[2], Jim Cooley[3], Lauren Bright[4], Phillipe Chouvarine[5], Andrew Claude[6], Alison Eddy[7], Bindu Nanduri[7], Jacquelyn Bowser[8], Wellesley Dittmar[7], Alison Linch[7], Fiona McCarthy[9], Shane Burgess[9] and Cyprianna Swiderski[10], (1)Mississippi State College of Veterinary Medicine, Starkville, MS, (2)Mississippi State College of Veterinary Medicine, Mississippi State, MS, (3)Mississippi State University, Missisippi State, MS, (4)Rutgers University, New Brunswick, NJ, (5)Hannover Medical School, Hannover, Germany, (6)Michigan State University, East Lansing, MI, (7)Mississippi State University, Mississippi State, MS, (8)Johnson and Wales, Providence, RI, (9)University of Arizona, Tucson, AZ, (10)Department of Clinical Sciences, Mississippi State, MS

Equine Pasture Asthma (EPA), an asthma-like disease affecting horses housed on pasture in hot, humid climates, demonstrates key diagnostic features of severe, adult asthma. These include reversible airway obstruction, neutrophilic airway inflammation, and airway hyper-responsiveness (AHR) of a magnitude that is diagnostic of severe asthma (≤1mg/ml methacholine). Decreasing AHR decreases asthma severity, making it a primary goal of asthma therapy. To address this goal, we employed RNA sequencing of lung tissue from 2 horses with EPA and 2 controls to identify novel signaling pathways that contribute to AHR in EPA horses. By manually filtering 1376 differentially expressed genes (DEGs) that were a) conserved in diseased horses during clinical exacerbation, b) not differentially expressed by season in controls, c) had raw read counts approaching 0 during disease remission, and d) have known roles in autonomic signaling and muscle physiology, we identified hyperpolarization activated cyclic nucleotide gated potassium channel 4 (HCN4) as an overexpressed target for potentiating AHR in EPA horses. Histochemical staining of lung samples from a separate EPA and control cohort confirmed that increased HCN4 localizes to airway smooth muscle (ASM) in lung samples collected during seasonal pasture asthma exacerbations. Antagonist responses in isolated equine bronchi confirm that HCN4 signaling contributes to the contractile responses of ASM in horses. Coupled with the

physiologic role of HCN4 in raising resting muscle membrane potentials, our findings support that HCN4-mediated current has a role in AHR that characterizes EPA horses.

W357: Equine 2

Proteome and Transcriptome Profiling of Equine Myofibrillar Myopathy Identifies Diminished Peroxiredoxin 6 and Enhanced Cysteine Metabolic Pathways

Sudeep Perumbakkam[1], Stephanie J. Valberg[1], Erica C. McKenzie[2] and Carrie Finno[3], (1)Michigan State University, East Lansing, MI, (2)Oregon State University, Corvallis, OR, (3)University of California Davis, Davis, CA

Equine myofibrillar myopathy (MFM) causes exertional muscle pain and is characterized by myofibrillar disarray and ectopic protein aggregates of unknown origin. To investigate the pathophysiology of MFM, we compared the skeletal muscle proteome and 3 h post-exercise transcriptome of gluteal muscle in MFM and control Arabian horses using iTRAQ and RNA-sequencing analyses. Differential expression (DE) was evaluated using edgeR and pathway analysis using Cytoscape and Cluego. Proteome analysis revealed significantly lower antioxidant peroxiredoxin 6 content (PRDX6, ↓4.14 $\log_2$ fold change [FC]), sarcomere protein tropomyosin (TPM2, ↓3.24X) and higher fatty acid transport enzyme carnitine palmitoyl transferase (CPT1B, ↑3.49X) in MFM vs. control muscle at rest. Three hours after exercise, 191 genes were DE in MFM vs. control muscle with a remarkably focused > 1.5 $\log_2$FC in genes involved in sulfur compound/ cysteine metabolism such as cystathionine-beta-synthase [CBS, ↑4.51] and a cysteine and neutral amino acid membrane transporter [SLC7A10, ↑1.79]. In MFM vs. control at rest, 284 genes were DE with > 1.5 $\log_2$ FC in pathways for structure morphogenesis, fiber organization, tissue development and cell differentiation including> 2 $\log_2$ FC in alpha actin-cardiac [↑ ACTC1], cytoskeletal desmoplakin [↑ DSP], basement membrane usherin [↓ USH2A] and delta like non-canonical Notch ligand 1, [↓ DLK1]. In conclusion, myofibrillar disarray and protein aggregation in MFM horses was embodied by DE expression in pathways of structure/fiber organization and tissue regeneration. Reduced antioxidant capacity as a potential etiology for MFM was supported by diminished cysteine rich antioxidant peroxiredoxin 6 with compensatory increased cysteine synthesis following exercise.

W358: Equine 2

Elucidating the Etiology of Atypical Equine Thrombasthenia in Thoroughbreds

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

Bleeding disorders are frequently seen in Thoroughbred racehorses and can negatively impact their performance. Atypical Equine Thrombasthenia (AET) was elucidated as a frequent cause of bleeding, affecting approximately one in every 150 Thoroughbreds. AET can be diagnosed using flow cytometry to determine the extent that platelets bind fibrinogen. Affected platelets bind 35% or less fibrinogen as compared to healthy platelets. Pedigrees of affected horses indicate that the disease is heritable, though the underlying disease etiology remains unknown. AET platelets have been thoroughly phenotyped and the thrombin signaling pathway identified as the likely site of a causative genetic variation. Six AET affected and twelve unaffected Thoroughbreds underwent whole genome sequencing (50x coverage) and segregating variants were identified using both candidate gene and whole genome approaches. Due to the aberrant signaling in the thrombin pathway, fourteen genes that code for the primary proteins within the pathway were investigated using the candidate gene approach. There was no variation in the candidate genes significantly associated with AET. However, the whole genome approach identified a region of interest with segregating variants on chromosome 24. By identifying the genetic mechanism of AET, breeders can select against the disease to improve Thoroughbred health and performance.

W359: Equine 2

Genome Wide Association Study Identifies Locus for the Mushroom Coat Color Dilution in Shetland Ponies

Jocelyn L. Tanaka, University of California - Davis, Davis, CA, Maura Mack, Colorado State University Veterinary School, Fort Collins, CO, Tosso Leeb, Institute of Genetics, University of Bern, Bern, Switzerland, 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

Coat color is a trait of economic importance to horse breeders. Mushroom is a unique phenotype in the Shetland pony that is characterized by the dilution of the chestnut coat color to a light sepia tone, while leaving bay and black base coat colors unaffected. The molecular mechanism for this trait is unknown. Previous pedigree analysis suggested a simple recessive mode of inheritance. To identify a candidate locus for further investigation a genome wide association study (GWAS) utilizing the Affymetrix 670K array (MNEc670k) was performed with DNA isolated from 12 mushroom and 12 chestnut horses. To correct for population substructure, (l=1.11), a single locus mixed linear model analysis (EMMAX) approach was utilized. This approach identified a single region on ECA7 that reached genome wide significance ($P_{corrected}=7.64X10^{-5}$). This region contains a 328 Kb haplotype that was perfectly concordant with the mushroom phenotype. Replication testing of four SNPs from this haplotype using 42 additional horses confirmed this association ($P_{combined}=4.51X10^{-16}$). The haplotype spans eight genes, one of which is a putative functional candidate. Further research is needed to explore this candidate's role in phaeomelanogenesis and mushroom coat dilution.

W360: Evolution of Genome Size

Accessing Plant Genome Evolution using LTR Retrotransposons

Shujun Ou, Michigan State University, East Lansing, MI

W361: Evolution of Genome Size

Flightless Birds have Larger Genomes Despite Lower Transposable Element Accumulation

Aurelie Kapusta, Department of Human Genetics, University of Utah, Salt Lake City, UT

W362: Evolution of Genome Size

Differential Drought Response and Transcriptome Size Plasticity between Diploid and Autotetraploid *Tolmiea*

Clayton J. Visger[1], Pamela S. Soltis[2] and Douglas E. Soltis[2], (1)California State University, Sacramento, Sacramento, CA, (2)University of Florida, Gainesville, FL

We investigated the transcriptomic basis of the ecophysiolgical divergence in drought response between diploid and polyploid Tolmiea. Using polyethylene glycol treated hydroponic cultures, we subjected T. diplomenziesii and T. menziesii to negative osmotic potential, inducing extreme drought stress. We then compared gene expression over time in response to our treatment and determined the gene functions most likely to contribute to the physiological differences between T. diplomenziesii and T. menziesii. Using recently developed methods, we accounted for variation in transcriptome size and cell size/density, enabling our comparisons of gene expression to take place in the context of change per cell, per biomass, and per transcriptome.

We found that in response to drought, tetraploid *Tolmiea* exhibits an extreme degree of transcriptome size plasticity, both between individuals and within individual drought responses. Additionally, we found that between the diploid and tetraploid, 9.2% of all loci investigated were differentially responsive to drought. Based on the integration of a functional enrichment analysis and prior physiological investigations, our results suggest that the tetraploids may reduce their photosynthetic machinery in response to drought.

W363: Evolution of Genome Size

Unusual Biogeographic Disjunction in the Cotton Tribe Yields Insight into Genome Downsizing

Justin Conover, Iowa State University, Ames, IA

The cotton tribe (Gossypieae) has experienced multiple trans-oceanic dispersals, generating an aggregate geographic range that encompasses much of the tropics and subtropics worldwide. Two genera in the Gossypieae, *Kokia* and *Gossypioides*, exhibit a remarkable geographic disjunction, being restricted to the Hawaiian Islands and Madagascar/East Africa, respectively. Here, *de novo* genome sequences of representative species within these genera are used to investigate questions underlying the genome downsizing that characterizes *Kokia* and *Gossypioides* relative to other genera in the tribe. We report that while the transposable element fraction of each genome is relatively static, each genus has experienced a loss of approximately 30% of the gene content as compared to their sister group (*Gossypium*), and a history of genome-wide accumulation of deletions. In both genera, there is a genome-wide bias toward deletions over insertions, and the number of gene losses exceeds the number of gene gains by about two- to four-fold. The genomic analyses presented here elucidate genomic consequences of the demographic and biogeographic history of these closest relatives of *Gossypium*, and enhance their value as phylogenetic outgroups.

W364: Evolution of Genome Size

The Evolutionary Maintenance of Unreduced Gametes and their Role in Whole Genome Duplication

Julia M. Kreiner[1], Paul Kron[2] and Brian C. Husband[2], (1)University of Toronto, Toronto, ON, Canada, (2)University of Guelph, Guelph, ON, Canada

Fertilization involving unreduced (2*n*) gametes is considered the dominant mechanism of polyploid formation in angiosperms; however, our knowledge of the prevalence of and evolutionary mechanisms maintaining 2*n* gametes in natural populations is limited. The product of irregularities in meiosis, we expect 2*n* gametes to be rare and deleterious in most natural populations, contrary to their wide taxonomic distribution and the prevalence of polyploidy. We used flow cytometry across 60 populations and 24 species of the Brassicaceae to examine how variation in the frequency of unreduced gamete varies along with life history traits such as reproductive mode in a phylogenetic context. Most individuals produced < 2% 2*n* male gametes with only rare individuals with elevated production. Variation in 2*n* gamete production was significant among species and related to reproductive system; asexual species produced significantly more 2*n* gametes sexual species. These results are consistent with 2*n* gametes being deleterious but maintained when opportunities for selection are limited. In addition, we reviewed and synthesized contemporary evidence relating to factors that should influence the rate of 2*n* gamete production and polyploid formation, and outline outstanding gaps in our knowledge for understanding the origins of polyploidy and its prevalence.

W365: Exploring Phytobiomes

Challenges and Opportunities in Characterizing Microbiomes

Mihai Pop, University of Maryland, College Park, MD

The complex microbial communities associated with virtually all living organisms are increasingly easy to explore through modern genomic technologies. Initial studies of these communities have highlighted their important contributions to the health of plants, animals, and humans, and have opened up tantalizing prospects for the use of microbiome science in agriculture and medicine. Despite initial successes, numerous challenges remain that limit our current knowledge and the potential for impact of this emerging field. These challenges span the entire analytical pipeline, from the technologies and devices used to measure the microbiota, to the analytical software used to process the raw data, and to the statistical approaches used to derive new scientific insights. In this talk I will highlight some of the key challenges and outline new opportunities afforded by the new data being generated. I will primarily focus on computational problems arising from the sheer size of the data sets generated in metagenomic studies, and describe recent results from my lab in the reconstruction of genome sequences from complex samples.

W366: Exploring Phytobiomes

The Poplar Tree Microbiome: Implications of the Ecosystem within

Sharon L. Doty[1], Andrew W. Sher[2], Pierre M. Joubert[1], Matthew Joseph[1], Andrea Firrincieli[3], Dehong Hu[4], Adam Deutschbauer[5], Zareen Khan[1] and Galya Orr[4], (1)University of Washington, Seattle, WA, (2)University of California, San Diego, La Jolla, CA, (3)University of Tuscia, Viterbo, Italy, (4)Pacific NW National Lab, Richland, WA, (5)Lawrence Berkeley Natl Lab, Berkeley, CA

With the relatively long life cycles of plants, symbiosis with microorganisms may allow plants to rapidly overcome environmental challenges. Endophytes are bacteria and fungi that live in intimate association within plants. The plant microbiota provide numerous benefits to the host plant including N-fixation, phytohormone production, reduced stress responses, antimicrobial production, tolerance to heat, salt, and drought, and pollutant degradation. Although some plant species are leguminous or actinorhizal, associating with rhizobia or Frankia, respectively, in root nodules, many pioneer plant species are non-nodulating and yet thrive in low-nutrient settings. For these plants, N-fixing (diazotrophic) endophytes and other closely associated microorganisms may be the source of this essential macronutrient. We study the diazotrophic endophytes of poplar (*Populus*) and willow (*Salix*), pioneer plant species able to colonize the rocky substrates deposited following riparian flooding. The $^{15}N_2$ incorporation assay was used to directly demonstrate N-fixation in cuttings of wild poplar. The microbiota varies considerably within the plants in number and in species with the intriguing possibility that only specific communities are active. Putative diazotrophic microorganisms were isolated from wild poplar plants, characterized, and sequenced. The presence of nitrogenase (*nif*) genes, in vitro assays for N-fixation activity, and quantitative *nifH* fli-FISH supported the hypothesis that the endophyte strains are diazotrophic. Random barcoded TnSeq is underway to elucidate the endophyte genes required for plant colonization and N-fixation. A consortium of the strains was added to hybrid poplar, increasing growth and N-fixation under greenhouse conditions. Not only do the microbes improve growth of this important bioenergy plant species, they also increase growth, health, and yields of an exceptionally broad range of plant species, including rice, tomato, pepper, strawberries, ryegrasses, and Douglas-fir. Inoculation of plants with endophytes improved water use efficiency and drought tolerance of the host plant. With the increased stress of climate change, the implications of plant-microbe symbioses for agriculture, forestry, and bioenergy production are profound.

W367: Exploring Phytobiomes

Plant-Microbe Communication in the Shrub Willow Rhizosphere: Microbiome Structure, Function and Crop Yield

Wanyan Wang[1], John E. Carlson[2], Eric S. Fabio[3] and Lawrence B. Smart[3], (1)Pennsylvania State University, University park, PA, (2)Dept. of Ecosystem Science & Management, Pennsylvania State University, University Park, PA, (3)Cornell University, Geneva, NY

Plant roots are colonized by tens of thousands of microorganisms and these microbes and their metabolites have profound effects on host physiology and development. However, the factors which determine the rhizosphere microbial profile are ambiguously understood. This study compared the effects of geography and host genotype on rhizopheric microbial communities of shrub willow grown in plantations at various sites in Northeast US. In total, 130 rhizoshpere soil samples representing three geographic locations (Rock Springs in Pennsylvania, Fredonia in New York and Mylan Park in West Virginia) and twelve willow genotypes were analyzed via metagenomics technology, spanning a 3-year period from planting time to harvest time. Geography and host genotype were both found to have significant effects on the structure and function of willow rhizospheric microbiomes. The effect of geography predominated. The bacterial genera *Mycobacterium, Methylobacterium, Fankia, Rhodopseudomonas and Nitrobacter*, which include well-studied plant growth-promoting microorganisms, showed significant correlations with willow biomass yields in the states of NY and PA, indicating that they may have vital roles in willow growth and yield.

W368: Exploring Phytobiomes

A Directed Selection Approach to Engineering Beneficial Plant Microbiomes

Susan Turner, BioConsortia, Inc., Davis, CA

The prospect of improving agricultural productivity and sustainability through the use of microbes has driven exponential growth in the number of research initiatives focused on the phytobiome. A challenging goal is the identification of commercially usable microbes that perform effectively and consistently. The need to deliver these microbial solutions in formats that align with agronomic practices, such as seed treatments, is a further challenge that often prevents the translation of lab-based discovery into commercial products. BioConsortia has developed a directed selection strategy that addresses these issues by building selection for these factors into early phases of the discovery process.

This process, termed Advanced Microbial Selection (AMS), involves iterative rounds of plant-based assays in which the microbiome of superior plants is harvested and used as the inoculum in subsequent rounds of screening. Through this process, the microbiome is evolved to include effective plant-colonizers that confer beneficial properties to the plant. By subjecting the plants to environmental stressors, and manipulating agricultural inputs, the process can be tailored to co-select for a wide range of beneficial microbes. Selection can also be structured to build resilience to downstream processes such as seed treatment formulations. Coupling this process with microbiome analyses provides a window into microbiome dynamics of both beneficial and inferior plants. This information can then be used to inform isolation strategies and reconstruction of beneficial consortia for commercial applications. Aside from the commercial benefits, the process also offers a tractable experimental system for addressing key questions about the factors that drive microbiome evolution.

W369: Exploring Phytobiomes

Phytobiomes Flagship Initiative: The Discovery of Universal Host Mechanisms for Plant-Microbiome Association.

Piet Jones, Ben Garcia, Jessy Labbe, Gerald Tuskan, Timothy J. Tschaplinski and **Daniel Jacobson**, Oak Ridge National Laboratory, Oak Ridge, TN

We have used a genome wide association study (GWAS) with the detected members of the poplar phytobiome (including bacteria, archae, fungi, viruses, nematodes, insects and environmental interactions) used as phenotypes in order to identify plant genes that are important for the association with each of the taxa. Thus the use of phytobiome constituents as GWAS phenotypes is helping to elucidate the host mechanisms responsible for host-phytobiome interactions (and the regulation thereof) and provides indications of possible interactions among different members of the phytobiome itself. We have also scaled-up a kmer-based taxonomic identification method in order to use full kmer profiles of all of the publically available genomes in order to do taxa identification in metagenomics and metatranscriptomics datasets. This method is running on the supercomputers at ORNL and has already been used to screen metatranscriptomes and megenomes from various plant hosts. We believe that many of the genes responsible for host-phytobiome association have orthologs across the plant kingdom. Host and phytobiome association may be tailored by variance in the gene sequences present in different host species. However, many of the fundamental mechanisms

are likely to be conserved across species. We believe that our work can provide key mechanisms that can be validated and used in a variety of crop species. We already have momentum for such a project with agreements for collaboration with groups in the US and Europe who are working on *Brachypodium, Arabidopsis, Medicago,* potato and radishes and the *Gramineae* family of grass-based plant systems.

W370: Farm Animal Genome Editing
Editing: What Is Possible
Karl J. Clark, Mayo Clinic, Rochester, MN

W371: Farm Animal Genome Editing
Current Opportunities for Accelerated Livestock Breeding using CRISPR-Cas9
Vrushali Patil[1], Simon Lillico[1], Chris Proudfoot[1] and John Hickey[2], (1)The Roslin Institute, Midlothian, United Kingdom, (2)The Roslin Institute, Edinburgh, United Kingdom
Genetic engineering has seen unparalleled success in the last decade with the advent of CRISPR-Cas9 technology. Due to the demands of food supply around the world, accelerated improvements in agricultural productivity are needed if the mounting human population has to be sustained on this planet. Desirable traits are often complex and regulated by multiple genes hence it becomes time consuming to unravel how a particular gene influences a trait and how we can exploit it to produce the ideal phenotype. Precise gene editing – enabling indel production, sequence deletion and SNP introgression – offers the opportunity to induce targeted alterations in a genome sequence which can be then reliably quantified, thus adding greatly to our knowledge of gene function and its associated trait. This technology has been successfully applied to provide insights into reproductive efficiency, neurological stability and even resistance to deadly pathogens. The simplicity of CRISPR-Cas9 has also allowed us to introduce desirable mutations directly into the target species, saving considerable time and effort. We present our current projects aiming to enhance livestock breeding and consequently, our society.

W372: Farm Animal Genome Editing
Editing Enables Synthetic Speciation
Maciej B. Maselko[1], Siba R. Das[1], Stephen C. Heinsch[1], Jeremy M. Chacon[2], Aidan J. Peterson[3], James E. Parker[4], Michael B. O'Connor[3], Przemyslaw G. Bajer[5], William R. Harcombe[2] and Michael J. Smanski[1], (1)Department of Biochemistry, Molecular Biology, and Biophysics and BioTechnology Institute, University of Minnesota, St. Paul, MN, (2)Department of Ecology, Evolution, and Behavior and BioTechnology Institute, University of Minnesota, St. Paul, MN, (3)Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, (4)Computer Science and Engineering, University of Minnesota, Minneapolis, MN, (5)Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, MN
We present data that demonstrate the feasibility of Synthetic Incompatibility (SI); a method of engineering species-like barriers in sexually reproducing organisms. SI utilizes genome editing to introduce a silent mutation in a conserved region of a promoter followed by the expression of programmable transcriptional activator targeted to the non-edited sequence. Hybridization between an edited organism expressing the programmable transcriptional activator and an organism with a non-edited promoter results in lethal overexpression caused by transcriptional activation of the non-edited promoter. Animal applications of SI include genetic biocontrol of pest species, replacing disease vector populations with engineered non-vector organisms, preventing gene flow between genetically engineered fish or livestock and their non modified counterparts as well as the genetic biocontainment of experimental systems such as gene-drives. Agricultural applications include preventing transgene flow from engineered crops which may enable the use of herbicide resistant traits in plants with closely related weeds and large scale production of crops engineered to make high-value compounds. Results from proof of principle experiments in *Saccharomyces cerevisiae* and *Drosophila melanogaster* will be presented.

W373: Farm Animal Genome Editing
Evaluation of the Offspring of Genome Edited and Control Bulls
Amy Young[1], Tad S. Sonstegard[2] and **Alison Van Eenennaam**[1], (1)University of California, Davis, Davis, CA, (2)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN
Genome editing using transcription activator-like effector nucleases (TALENs) was used to introgress the dominant P_C *POLLED* allele into the genome of bovine embryo fibroblasts which were cloned by somatic cell nuclear transfer to produce two bull calves in 2015. Semen from one of these bulls was used to inseminate Horned Hereford cows, resulting in the birth of 6 calves (5 male, 1 female) in September, 2017. At the same time, three control calves sired by a Horned Hereford bull were born to Horned Hereford dams. All calves were deemed healthy upon veterinary examination, no horn buds were apparent by palpation for the calves from the gene edited bull but were present in the control calves. Complete blood counts and routine hematology parameters were within normal limits for all calves. DNA was extracted and used to verify parentage and confirm that the calves from the gene edited bull were all heterozygous for the polled allele by PCR. Quantitative PCR assays were designed for P_C, horned and a Y chromosome marker to assess fetal microchimerism, the presence of fetal DNA in maternal circulation, which has been identified in some mammalian species during and after pregnancy. DNA samples from the Hereford dams were assayed at six time points prior to and following delivery. No evidence of fetal microchimerism, specifically the presence of the polled allele or Y chromosome in the dams' blood, was observed. DNA samples from all animals have been whole genome sequenced and will be analyzed to evaluate genomic alterations.

W374: Farm Animal Genome Editing
Editing Disease Resistance in Livestock
A. Mark Cigan, Genus, DeForest, WI

Improvements in livestock genetics through classical methods of breeding and selection to produce high quality meat and increased milk production is a slow and resource intensive process. Animal agriculture stands to gain a huge leap forward with the adoption of new technologies like gene editing to improve disease resistance which will not only improve the health and well-being of the animal but reduce the need for antibiotics. Importantly, rather than backcrossing, gene editing enables the testing and rapid introduction of resistance alleles into the most genetically advanced, commercial lines of pigs and cattle. Gene editing projects at Genus will be presented which aim to improve animal health by delivering disease resistant livestock to farmers while maintaining food quality and enhancing sustainability.

W375: Farm Animal Genome Editing
Future Potential of Genome Edited Farm Animals
Tad S. Sonstegard, Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN

W377: Farmed Insects to Feed Future Populations
Generating Novel Phenotypes in the Non-Model Insect, *Gryllodes sigillatus*, using EMS Mutagenesis
Adam Session, Tiny Farms, San Leandro, CA

W378: Farmed Insects to Feed Future Populations
Engineering Insects for Food, Feed, Pharma and other Valuable Applications
Aaron T. Dossey, All Things Bugs LLC, Midwest City, OK

W379: Farmed Insects to Feed Future Populations
Genetic Engineering of Edible Insects for the Greater Good
Marce Lorenzen, North Carolina State University, Raleigh, NC

W380: Farmed Insects to Feed Future Populations
Repetitive Sequences and the Challenges of the Mealworm Genome Assembly
Brenda Oppert, USDA ARS Center for Grain & Animal Health Research, Manhattan, KS

W381: Feline & Canine Workshop
Phylogenomics of the Cat Family
Gang Li[1], Henrique Figueiró[2], Eduardo Eizirik[2] and **William J. Murphy**[1], (1)Texas A&M University, College Station, TX, (2)Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil
Attempts to resolve the evolutionary history of the cat family have encountered manifold challenges from multiple sources of phylogenetic discordance, including incomplete lineage sorting and cytonuclear discordance induced by rapid lineage diversification and complex patterns of hybridization. Here we used recombination-aware phylogenomics to test the validity of the total evidence approach by analyzing whole genome sequences from 28 species representing all eight extant lineages of the cat family. When divergence times were considered within and across felid clades, we found that the majority of the phylogenetic signal within the autosomes (>90% of the genome) is often not reflective of the inferred history of speciation events, due to pervasive episodes of ancient admixture within and between multiple lineages. Phylogenetic signal on the X chromosome was strongly partitioned by local recombination rate, which under different scenarios retained or obscured the species tree. Our results revise a number of accepted concepts of felid evolutionary history, and illustrate that the total evidence approach to phylogenetic inference may be highly misleading in lineages with complex histories, without considering the temporal and recombinational partitioning within genomes.

W382: Feline & Canine Workshop
Canine Athletes: Whole Genome Sequence Reveals Selection for Muscle and Cardiovascular Genes in Sport Hunting Dogs
Jaemin Kim, NIH-National Human Genome Research Institute, Bethesda, MD

W383: Feline & Canine Workshop
Genomic Evidence for the Chinese Mountain Cat (*Felis bieti*) as a Conspecific of the Wildcat (*F. silvestris*) with Contemporary Introgression to Local Domestic Cats
He Yu, Peking University, Beijing, China
Of the 37 species of Felidae, the Chinese mountain cat (*Felis bieti*) is among the most enigmatic and only found on the Qinghai-Tibet Plateau, China. Its evolutionary ancestry and taxonomic status remain controversial, with debates centered on whether it is an independent species or conspecific with the wildcat (*F. silvestris*) and whether it may have contributed to the origin and domestication of cats (*F. catus*) in Asia. Here, we conducted multi-locus sequencing in Chinese mountain cats (N = 26) collected range wide and its close relatives, the Asiatic wildcat (*F. s. ornata*, N = 1) and domestic cat (N = 238) to elucidate the genetic ancestry and evolutionary dynamics of this lineage. We further conducted whole-genome resequencing in 51 of the cats and proposed a formal reclassification of the Chinese mountain cat as a subspecies of the wildcat (*F. s. bieti*). A complex hybridization scenario among wildcat lineages was also revealed, including an ancient introgression event from *F. s. ornata* into *F. s. bieti* and contemporary gene flow between *F. s. bieti* and domestic cats across but not beyond the range of *F. s. bieti*. No evidence was found that *F. s. bieti* contributed to the origin and domestication of cats in East Asia, confirming that domestic cats worldwide share a single Near Eastern origin from the African wildcat (*F. s. lybica*). Our study describes the evolutionary process and demographics of the wildcat during the Pleistocene, sheds lights on the origin of domestic cats in East Asia and provides new evidence for Felidae taxonomic reclassification.

W384: Feline & Canine Workshop
Pet Dogs, Citizen Science, and the Genomics of Behavior
Elinor K. Karlsson, Broad Institute & U Mass Med School, Cambridge, MA

W385: Feline & Canine Workshop
Progress on the Reassembly of the Domestic Dog Genome
Vidhya Jagannathan, Institute of Genetics, University of Bern, Bern, Switzerland

W386: Feline & Canine Workshop
Variant Discovery in Preparation for Feline Precision Medicine
Wesley Warren, McDonnell Genome Institute at Washington University, St. Louis, MO

W387: Feline & Canine Workshop
Population Genomics of Wild Tigers: using Genomic Data to Inform Conservation
Ellie Armstrong, Department of Biology, Stanford University, Stanford, CA

W388: Flax Genomics
Polymorphisms and Induced Mutations in the Flax (*Linum usitatissimum*) Genome
Michael K. Deyholos, University of British Columbia, Kelowna, BC, Canada; Univ British Columbia Okanagan, KELOWNA, BC, Canada

W389: Flax Genomics
First Insights into the Genetic Factors Underlying Post-Flowering Drought Tolerance in Flax
Braulio J. Soto-Cerda, Centro de Genómica Nutricional Agroacuicola, Temuco, Chile
Drought stress acts simultaneously on many traits and developmental stages which ultimately affects seed yield of crops worldwide. In particular, drought stress at the reproductive stage (flowering and seed development) can result in the most significant reductions in crop production.In this study, yield and yield-related drought tolerance indices were assessed on 120 flaxseed accessions under mild drought and irrigated conditions at flowering time. A set of ~700,000 single nucleotide polymorphisms (SNPs) was screened for marker-trait associations using general models. In average, drought stress reduced seed yield by 21% and start of flowering was brought forward by 5 days. Based on the stress tolerance index (STI), three genomic regions on linkage groups 2, 7 and 9 were identified which explained 68.2% (R^2) of the phenotypic variation. Ten candidate genes were located nearby peak SNPs (~50 kb either side) where a clathrin heavy chain 1 (CHC1), a mitogen-activated protein kinase (MAPK), a peptide chain release factor and a CCCH-type zinc finger protein genes, which have previously been involved in drought response in other crops were the most promising candidates for further studies. This information provides important genetic insights into the natural variation of flaxseed drought tolerance. The identified SNPs or candidate genes could serve as direct targets for both genetic engineering and selection for flaxseed trait improvement.
This research was supported by Fondo Nacional para el Desarrollo Científico y Tecnológico (FONDECYT) Chile, project N° 1161133.

W390: Flax Genomics
Multiple Approaches Lead to an Improved Understanding of Lignification in Flax
Simon W Hawkins, University of Lille 1, Villeneuve d'Ascq, France
Lignin is a major factor impacting flax fibre quality and a better understanding of its biosynthesis is therefore important. Analysis of the sequenced flax genome shows that lignin genes belong to multigene families and targeted engineering and/or selection programmes aimed at improving fibre quality therefore require the identification of the *bona fide* lignin genes. We have used a combination of expression analyses, proteomics and *in situ* hybridisation to identify family gene members most likely to be involved in lignin biosynthesis. The recent development of a bioorthogonal (click chemistry) imaging approach is also leading to increased understanding of lignification in flax.

W391: Flax Genomics
Oligonucleotide-Mediated Genome Editing to Develop Traits in Flax
Javier Narvaez-Vasquez, Cibus US LLC, San Diego, CA and Yohannes A. Mihiret, Noel J. Sauer, Jerry Mozoruk, Melody J. Woodward, Lindsay Shaw, Kara Sarver, Andrew Walker, Zhixia Niu, Tracey A. Lincoln, Steven L. Sanders, Keith A. Walker, Peter R. Beetham, Christian R. Schöpke, and Greg F.W. Gocal
Breeding is limited by the available variation that can be leveraged from within a crop and from its wild relatives. The *Rapid Trait Development System* (***RTDS***™) is a system for accelerating breeding that enables diversity to be increased within a crop to develop precise, non-transgenic traits to benefit consumers, processors and farmers. ***RTDS*** can employ a combination of Gene Repair OligoNucleobases (GRONs) and double strand breakers to make defined spelling changes in genomic DNA in single cells and with advanced cell culture, regenerate those cells into whole plants. We applied our ***RTDS*** technology to develop an herbicide tolerance trait in *Linum usitatissimum* (flax) by precisely editing *5'-enolpyruvylshikimate-3-phosphate synthase* (*EPSPS*). *EPSPS* edits occurred at sufficient frequency that we could regenerate whole plants from protoplasts without employing selection. These plants were subsequently determined to be tolerant to the herbicide glyphosate *in vitro* and in greenhouse spray tests. Progeny (C1 – conversion generation 1) of these plants showed the expected Mendelian segregation of *EPSPS* edits. Our findings show the enormous potential of using a genome editing platform for precise, reliable trait development in flax and in other plants.

W392: Flax Genomics

Translating Genomic Advances in Flax/Linseed for Fibre Development

Prasanta Dash, ICAR-National Research Centre on Plant Biotechnology, New Deli, India

Genomic research in flax over last five years has transformed flax/linseed (*Linum usitatissimum*), the genomic poor "Orphan crop" to a burgeoning model crop. Comprehensive genetic, genomic and biochemical characterization along with discovery of biotic and abiotic trait governing genes have opened up research vistas for comprehensive flax development. The momentum started with decoding of flax structural genome in 2012 and flax community researchers tapped the genome sequence information for translational research. Incorporating drought tolerance along with enhanced fibre development has been the cardinal objective of our research. In this regard, we have delineated involvement of phospholipase in drought tolerance in flax and aquaporins in fibre development. Deploying genetic, genomic, biochemical and molecular approaches we have cloned and characterized the flax aquaporins genes (*TIP* and *NIP*) in fibre development and phospholipase (*phlA*, *phlB*) genes in drought tolerance in flax. We envisage, deployment of these monogenic dominant genes will generate a cultivar endowed with drought tolerant and better fibre quality in flax. The Development of transgenic flax is in progress in our laboratory.

W393: Flax Genomics

The Flax Genome – a Revolution in Evolution through Natural Genome Editing

Christopher Cullis, Case Western Reserve University, Cleveland, OH

The flax genome can be rapidly modified within a single generation in response to the growth environment. The sites of variation have been characterized in 5 lines, the progenitor line (Pl), 3 derived lines (genotrophs) and a tissue cultured line, by whole genome sequencing. The variations do not appear to be due to either random mutations or the movement of transposable elements, but rather that the genome appears to be able to be edited resulting in two well-defined, different sequences at many loci. These edited regions can extend over many kilobases. Such large-scale, reproducible variation is unlikely to occur through multiple independent events. Therefore an editing mechanism by which long tracts of the genome can be replaced with an alternative structure is being proposed. This editing mechanism is activated somatically, in the apical meristem, by the sub-optimal growth of an individual, with a set of genomic sites available for modification that have the potential to generate phenotypic diversity without lethality. These sites can be independently modified until a 'new' genome, that functions more efficiently in a meristematic cell the selecting environment, is generated. At that stage the editing mechanism is silenced and the 'new' genome replaces the previous version. If this occurs prior to flowering, the edited genome is transmitted to the progeny. Both the initial and the 'new' genomes have been observed in flax varieties and the wild progenitor of cultivated flax.

W394: Flax Genomics

Fun with Flax (*Linum usitatissimum* L.): Dirigent (*DIR*) Protein Gene Identification, *DIR* differential Regulation and Evolution, and *DIR*-Derived Oligomeric Lignan Imaging *in situ*

Norman Lewis[1], Cyrielle Corbin[2,3], Samantha Drouet[2,3], Lucija Markulin[2,3], Daniel Auguin[2,3], Éric Lainé[2,3], Laurence B. Davin[1], John R. Cort[4], Kye-Won Kim[1], Doralyn S. Dalisay[1], Syed G. A. Moinuddin[1], Oliver Rübel[5], Benjamin P. Bowen[5] and Christophe Hano[2,3], (1)Washington State University, Pullman, WA, (2)Université d'Orléans, Chartres, France, (3)COSM'ACTIFS, Chartres, France, (4)Pacific Northwest National Laboratory, Richland, WA, (5)Lawrence Berkeley National Laboratory, Berkeley, CA

Flax is a rich natural grain source of health-protecting 8-8′-linked secoisolariciresinol diglucoside (SDG) dirigent protein (DIR)-derived lignan oligomers. A genome-wide characterization of the predicted flax 44-membered *DIR* multigene family was performed. All predicted flax *DIR* sequences, including the promoters, were analyzed together with their public gene expression datasets. Expression patterns of selected *DIR* genes were examined using both qPCR and clustering analysis of *DIR* gene expression. Phylogeny and gene expression analysis segregated flax *DIR* genes into six distinct clusters with new cluster-specific motifs identified.

Our analyses further implicated roles for specific DIRs in formation of (–)-pinoresinol in seed coats, as well as of DIR-engendered formation of (+)-pinoresinol in vegetative organs and/or in specific responses to stress transducers. Pinoresinol is metabolized in the seed coats to produce SDG-derived lignan oligomers, whereas in vegetative tissues cytotoxic lignans, such as (–)-yatein, are produced. Metabolite imaging of developing seeds, using MALDI MS/MS and Ion Mobility analyses gave new temporal and spatial insights into SDG-derived oligomeric lignan formation.

W395: Forage, Feedstocks & Turf

Shoot Architecture and Biomass Yield in Switchgrass and Alfalfa

Zeng-Yu Wang and Jiqing Gou, Noble Research Institute, Ardmore, OK

Axillary bud development determines plant architecture and biomass yield. While performing micropropagation in the dedicated bioenergy crop switchgrass (*Panicum virgatum*), we identified genotypes that failed to form aerial buds. Overexpression of *miR156* in one of such genotypes induced aerial bud formation and increased the number of basal buds. Microarray and other expression analyses indicated that one of miR156 targets, SQUAMOSA PROMOTER BINDING PROTEIN LIKE4 (SPL4), directly regulated axillary bud formation. Downregulation of *PvSPL4* in the same genotype promoted aerial bud formation and increased basal buds, while overexpression of *PvSPL4* suppressed tillering and branching. We show that unlike all previously reported genes associated with basal bud formation, the *miR156-SPL4* module predominantly controls aerial bud formation and partially regulates basal bud development. Genetic manipulation of *PvSPL4* led to altered plant architecture with increased branching, enhanced regrowth after cutting and improved biomass yield in switchgrass. In a separate study, we identified three mutants with enhanced branching in the model legume *Medicago truncatula by screening a Tnt1* retrotransposon–tagged mutant population. Molecular analyses revealed that the mutations were caused by *Tnt1* insertions in the *MtSPL8* gene, an ortholog of *PvSPL4*. The *M. truncatula mtspl8* mutants had increased biomass yield, while overexpression of *MtSPL8* in *M. truncatula* suppressed branching and reduced biomass yield. Based on the *MtSPL8* sequence, an alfalfa *MsSPL8* was cloned and transgenic alfalfa plants were produced. The *MsSPL8* downregulated alfalfa plants showed large increases in biomass production in the first and second harvests compared to the control.

W396: Forage, Feedstocks & Turf

Development of a Commercial F1 Hybrid Ryegrass Breeding Scheme That Incorporates Advanced Phenomics and Genomics

Luke Pembleton[1], Hiroshi Shinozuka[1], Yvonne Ogaji[1], Carly Elliot[2], Preeti Thakur[1], Maiko Shinozuka[1], Michelle Drayton[1], Junping Wang[2], Hans D. Daetwyler[3,4], Pieter Badenhorst[2], German C. Spangenberg[3,4], John Forster[1,5] and **Noel O.I. Cogan**[4,5], (1)Agriculture Victoria Research, Bundoora, Australia, (2)Agriculture Victoria Research, Hamilton, Australia, (3)School of Applied Systems Biology, La Trobe University, Bundoora, Australia, (4)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Australia, (5)La Trobe University, School of Applied Systems Biology, Bundoora, Australia

Conventional ryegrass breeding has been limited in its ability to deliver genetic gain, partly due to the inability to effectively exploit and deliver heterosis through the formation of F_1 hybrids. Controlled crossing of geographically distant ryegrass lines demonstrated heterosis can result in c.25% increases in dry matter production. To date there have been no effective and commercially suitable methods of obtaining high proportions of F_1 hybrid seed. Advances in the understanding of the genes regulating the self-incompatibility loci in grasses (*S* and *Z*) have enabled the development of initial diagnostic markers and a methodology for the efficient generation of high proportions of F_1 hybrid ryegrass seed on a commercial scale. Initial parental populations have been developed and experimental F_1 varieties are being evaluated, that have already shown superior performance compared to reference commercial cultivars. Through the application of plant phenomics detailed trials are being undertaken to evaluate a large range of F_1 hybrid test crosses along with the parents and current market leading varieties, all of which have been comprehensively genotyped. To complement these approaches genomic selection is being developed to deliver gains directly into the parental pool development methodology, to deliver maximal advances in productivity. This technology has the ability to revolutionize ryegrass breeding, introducing a new paradigm of breeding methodologies with rapid genetic gains, supporting increased agricultural production.

W397: Forage, Feedstocks & Turf

Development of a High-Energy Red Clover

Bruno Studer, ETH Zurich, Zurich, Switzerland

Red clover (*Trifolium pratense* L.) is one of the most important forage legumes in grassland agroecosystems worldwide. Although red clover has a high biomass potential and is particularly valued for its high protein content, red clover herbage lacks the high-energy carbohydrates required to meet the productivity potential of modern livestock breeds. Like most plants, red clover accumulates diurnal starch in its leaves during the day as a temporary carbon store of photosynthesis, but harvesting this starch is challenging. To develop a harvestable high-starch agronomic trait in red clover, a reverse genetic pipeline based on Targeting Induced Local Lesions in Genomes (TILLING) was establish to identify beneficial alleles, which naturally accumulate starch. Implementation of a high starch trait is envisioned to maximize the protein and energy content of forage crops in order to deliver a higher proportion of the feed intake from environmentally sustainable and locally produced roughage.

W398: Forage, Feedstocks & Turf

Optimising Water Stress Tolerance in Energy Grasses

Iain Donnison, Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth University, United Kingdom, Paul R. H. Robson, Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, United Kingdom and Marta Malinowska, Aarhus University/ Department of Molecular Biology and Genetics, Slagelse, Denmark

The fundamental challenge for energy crops is how do we produce sufficient feedstock without negatively impacting food production and whilst maximising environmental benefits. One means by which this may be achieved is by growing energy crops on marginal land, however this may mean that water availability is limited even though irradiation and temperature are favourable. This creates a challenge as yield is strongly linked to water availability. Candidate energy grasses that grow across a wide range of environments and therefore have the potential to include diversity for drought tolerance and water use include Miscanthus and reed canary grass. In separate studies, diverse populations of these grasses have been studied in a high throughput phenomics facility under drought treatments (both mild and severe) to generate high-quality time-course data for biomass accumulation and water use. Plants were characterised by the collection of daily visual spectrum images which were then calibrated to actual harvested biomass so that biomass accumulation could be predicted over the experiment. Image analyses were used to determine both growth and leaf senescence over time and with treatment. Microclimate/ geographical modelling using this data indicated that the origin of Miscanthus genotypes could be associated with drought tolerance and helped explain responses. Water use efficiency could also be correlated with summer rainfall of origin of genotype. Phenomics therefore has the potential to help select candidate genotypes for crop breeding for improvement in water stress tolerance and use efficiency.

W399: Forage, Feedstocks & Turf

Genome-Wide Association and Transcriptome Analyses of Heat Tolerance in Perennial Ryegrass

Gang Nie[1], Xinquan Zhang[1] and **Yiwei Jiang**[2], (1)Sichuan Agricultural University, Chengdu, China, (2)Department of Agronomy, Purdue University, West Lafayette, IN

Perennial ryegrass *(Lolium perenne* L.) is widely cultivated in the temperate regions. Susceptibility to high temperature in summer limits its broad use. The objective of this study was to identify genetic variants and candidate genes associated with heat tolerance of perennial ryegrass. Large variations in plant height, chlorophyll index, chlorophyll fluorescence (Fv/Fm), leaf fresh weight, leaf dry weight, and leaf water content (LWC) were found among 271 accessions under the control and heat stress conditions (35ºC/30ºC, day/night for 14 days). A total of 24,995 SNPs were obtained by genotyping-by-sequencing. Genome wide association study (GWAS) detected 11 signals associated with fresh and dry weight, LWC, and Fv/Fm under heat stress or relative values to the control. Transcriptome analyses identified 3783 and 1844 differentially expressed genes in the heat tolerant and sensitive accession, respectively. The integrated GWAS and transcriptome analyses identified several candidate genes associated with growth and physiological traits. The functions of candidate genes will be verified for further revealing heat tolerance mechanisms of perennial ryegrass.

W400: Forage, Feedstocks & Turf

Development of the Alfalfa Breeder's Toolbox: Integration of Genomic, Genetic and Germplasm Resources for Alfalfa Improvement

Maria J. Monteros[1], Chunlin He[1], Jaeyoung Choi[1], Patrick X. Zhao[1], Nadim Tayeh[1], Xinbin Dai[1], Andrew Farmer[2], Joann Mudge[2], Haibao Tang[3], Junil Chang[1], Nick Krom[1], Justin Vaughn[4], Perdeep K. Mehta[1], Christy Motes[1], Michael Trammell[1], Brian Motes[1], Shawn Sullivan[5], Ivan Liachko[5], E. Charles Brummer[6], Nevin Young[7], Christopher Town[8] and Michael K. Udvardi[1], (1)Noble Research Institute, Ardmore, OK, (2)National Center for Genome Resources (NCGR), Santa Fe, NM, (3)Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fuzhou, AZ, China, (4)USDA - ARS - GBRU, Athens, GA, (5)Phase Genomics, Seattle, WA, (6)University of California, Davis, Davis, CA, (7)Department of Plant Pathology, University of Minnesota, St. Paul, MN, (8)J. Craig Venter Institute, Rockville, MD

Genetic and genomic resources can be used to increase genetic gains in plant breeding programs. Alfalfa (*Medicago sativa* L.; 2n=4x=32) is a perennial forage legume with global agronomic importance. Strategies to accelerate alfalfa improvement include sequencing and assembling the alfalfa genome and integrating genomic sequences with other datasets in the Alfalfa Breeder's Toolbox (ABT; https://www.alfalfatoolbox.org). Two different alfalfa genotypes, one diploid (2x) and another tetraploid (4x), were used for genome sequencing. The 2x genome was generated using PacBio sequencing, while Illumina and PacBio sequencing were used for the 4x alfalfa genome. The sequence data these technologies generated resulted in a highly fragmented tetraploid genome assembly due to high levels of heterozygosity and the complexity associated with haplotype variation. Integrating a Hi-C based proximity guided assembly enabled the *de novo* assembly of chromosome-scale scaffolds corresponding to the eight alfalfa chromosomes. The 2x genome v1.0 was further analyzed with a gene annotation pipeline so that it could anchor additional datasets in the ABT. The ABT provides access to and visualization of (1) the 2x alfalfa genome sequence, (2) gene models, (3) gene expression profiles in response to abiotic stress conditions presented in the gene expression atlas, (4) molecular markers, (5) shifts in allele frequencies obtained through cycles of selection, and (6) phenotypic data from field-based germplasm evaluation trials. Specific functionalities of the ABT include test case scenarios to address practical plant breeding applications and these can be used for genomics-based breeding approaches to develop improved alfalfa cultivars.

W401: Forest Tree

Genomic Selection in Forest Trees: Lessons Learned and the Way Ahead

Dario Grattapaglia, Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brazil

We and others have shown that genomic predictions can match phenotypic prediction for growth and wood quality traits in a number of forest tree species and populations. However, a number of questions await more data. These include validation across multiple generations, the impact of model updating to counterbalance the decay of accuracy and cost/benefit financial analyses. The successful implementation of GS will vary on a case- by-case basis. Based on a review of all published studies to date a tentative roadmap is presented for tree breeders that are contemplating exploring this breeding method: (1) **Training population:** sample preferably ~2,000 individuals from existing progeny trials derived from the top parents, covering the relevant spectrum of genetic variation; it is highly desirable to clone the 2,000 individuals and obtain more accurate phenotypes if selection target are clones. (2) **Relatedness**: prospective selection candidates must be genetically related to the training population for optimal accuracy. (3) **Genotyping**: fixed-content SNP arrays are the gold standard for data quality, breeder friendliness and speed of data generation; array costs have plummeted and thus highly competitive with GbS; for effective population sizes (N_e= 20-80) densities of 10,000-30,000 SNPs will usually provide satisfactory predictions. (4) **GxE and age interactions**: train models on traits measured at the same age and environment as the ones where predictions on selection candidates are planned. Data from G*E or age-age correlation studies adequately inform what to expect from GS. (5) **Model retraining**: no experimental data are available yet but simulations suggest that models will have to be periodically retrained with phenotypes collected in breeding generations closer to the selection candidates, such that decline in accuracy due to decay of relatedness and most importantly, to the constantly changing environment, is mitigated. (6) **Data analysis**: as most traits in forest trees fit the infinitesimal model, use RR-BLUP as a starting point from which to explore additional alternative methods. (7) **Logistics**: sample collection and tracking systems, DNA extraction, genotyping and data analysis pipelines can be accessed through service providers. (8) **Cost-benefit analysis**: with GS costs and accuracies in hand, carry out net-present-value analyses benchmarking against conventional breeding before considering the implementation of GS.

W402: Forest Tree

Breeding for the Future: Responding to New Challenges through Shortened Breeding Cycles

Natalie Graham, Emily J. Telfer, Jaroslav Klapste, Mari Suontama, Nari Williams and Heidi Dungey, Scion, Rotorua, New Zealand

The ability to rapidly introduce new selection traits into several breeding programs in New Zealand will be critical to the establishment of forests that are resilient to climate change. Traditional breeding cycles often span multiple decades, with intensive progeny testing. However, some forestry species now benefit from the availability of SNP array genotyping resources. For example, in *Eucalyptus nitens,* now entering the fourth generation of breeding, we demonstrated the potential of genomics to increase genetic gains through more accurate estimates of breeding values, and the possibility of halving the breeding cycle from 14 to seven years. In *Pseudotsuga menziesii*, still in its first generation, we are employing genomics to assist with first generation selections from native provenance collections and establish populations adapted to local conditions. However, for *Pinus radiata*, now in its third generation breeding cycle, no genomics resources were available. Scion and the New Zealand Radiata Pine Breeding Company therefore developed a 49K probe exome capture genotyping panel. To date, we have screened over 4000 industry-relevant individuals using this panel, collecting data on ~1 million SNPs and generating genomic estimated breeding values for a range of existing selection traits. In addition, we are now exploring new traits such as drought tolerance and resistance to red needle cast, a relatively new foliar Phytophthora disease. We report on the progress and challenges within the breeding programs for each of these species, with some key learnings for the implementation of genomics.

W403: Forest Tree

Accessible Chromatin Mapping in *Eucalyptus grandis* Vascular Tissue

Steven G Hussey, Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Pretoria, South Africa

W404: Forest Tree

Expression Genome-Wide Association Study (eGWAS) to Uncover Regulators of the Transcriptome and Whole-Plant Phenotypes

Kelly Balmant[1], Jerald Noble[2], Fabio Marroni[3], Christopher Dervinis[1], Daniel Conde[1], Paolo Triozzi[4], Marcio F. R. Resende[1], Brad Barbazuk[1], Gustavo de los Campos[5], Ana Ines Vazquez[5] and Matias Kirst[1], (1)University of Florida, Gainesville, FL, (2)Univesity of Florida, Gainesville, FL, (3)IGA Istituto di Genomica Applicata, Udine, Italy, (4)Universidad Politecnica de Madrid, Madrid, Spain, (5)Michigan State University, East Lansing, MI

Gene expression regulation has been shown to be largely genetically controlled and heritable, as described previously in yeast, animals and several plant species. Multiple genetic and environmental factors are likely to contribute to the expression of a gene. Therefore, transcript abundance values can be treated as any other quantitative phenotypes and subjected to genetic mapping to identify polymorphisms that contribute to variation of expression level in a population. Many of these polymorphisms may also contribute to variation in whole-plant phenotypes by acting pleiotropically. Studies that mapped the genetic regulation of transcripts and plant phenotypes in tree species have used bi-parental, single-generation populations. However, the high linkage disequilibrium in these populations hampers the detection of causative variants because of the low resolution. Here we performed high-resolution mapping of variants that regulate gene expression in an unstructured association population of 192 unrelated individuals of *Populus deltoides* individuals. We identified differentially regulated genes measured in xylem, and their heritability were estimated. Furthermore, we detected and quantified the contribution of putative *cis* and *trans*-regulatory variants to transcript abundance. The data from the eGWAS analysis is now being integrated with association genetic analysis results from whole-plant traits (biomass productivity and chemical properties) to discover putative regulatory genes.

W405: Forest Tree

Pine SNP Chip Consortium: Progress on Pine SNP Discovery and Array Design in Loblolly Pine

Fikret Isik, North Carolina State University, Raleigh, NC, Juan José Acosta, NC State University, Raleigh, NC, Andrew J. Eckert, Department of Biology, Virginia Commonwealth University, Richmond, VA, Richard Sniezko, USFS Dorena Genetic Resource Center, Portland, OR and Jill Wegrzyn, University of Connecticut, Storrs, CT

Previously generated NGS data sets originating from the USDA funded PINEMAP project and exome capture data from the University of Florida were aligned to the loblolly pine genome assembly v2.01. The new version has been annotated and serves as the new target for variant detection and genotyping array design in this project. Following the selection of optimal SNP calls, the individual SNPs were evaluated in regards to their flanking regions. Selections were further processed using Illumina's ADT scoring software. This resolved 169K potential targets for array development. The new annotation was curated with SNPeff in order to identify synonymous/non-synonymous sites as well as identify SNPs upstream of high quality gene annotations. The new annotation released with v2.01 was functionally re-annotated in order to provide comprehensive information on the gene space. From the ADT score filtered SNPs, a total of 52K were within or directly upstream of high quality gene annotations. Patterns of LD were quantified. Initial analysis of LD decay show rapid to moderate rates of decay, with a much larger variance than reported previously. The final size and selections will be derived from the TreeGenes database which is storing the full profile of quality, coverage, population, and annotation information on each loci. The pine consortium is currently working on choosing the platform for pine SNP chip design.

Funding: USDA-NIFA Award Number 2016-67013-24469.

W406: Forest Tree

Insights into Tree Lifespan Evolution from the Oak Genome

Christophe Plomion[1], Jean Marc Aury[2], Joelle Amselem[3], Thibault Leroy[1], Florent Murat[4], Sébastien Duplessis[5], Sébastien Faye[2], Nicolas Francillonne[3], Karine Labadie[2], Grégoire le Provost[1], Isabelle Lesur[1], Francis Martin[5], Patrick Wincker[2], Hadi Quesneville[3], Antoine Kremer[1] and Jérôme Salse[4], (1)INRA, Université de Bordeaux, Cestas, France, (2)CEA - Genoscope, Evry, France, (3)URGI, INRA, Université Paris-Saclay, Versailles, France, (4)GDEC, INRA, Université Clermont II Blaise Pascal, Clermont-Ferrand, France, (5)IAM, INRA, Université de Lorraine, Champenoux, France

A long-standing question in plant evolution concerns the ability of long-lived sessile species to sustain continuous exposure to biotic and abiotic stresses. We addressed this question, by generating a high-quality chromosome-scale genome assembly for the pedunculate oak (*Quercus robur* L.), an emblem of longevity, and studying its evolutionary history. We detected expansion through tandem duplications and relaxed purifying selection for disease resistance (R) genes. A comparison of the genomes of tree and herbaceous species showed that such expansions were particularly pronounced in trees. We also provide evidence for the accumulation and transmission of somatic mutations, a mechanism that may contribute to higher mutation rates in trees. These genomic features are probable signatures of the evolution of a diversified arsenal of weapons against biotic attacks, underpinning tree resilience and longevity.

W407: Forest Tree

Transcriptomic Diversity in a Mapping Population of *Populus deltoides*

Jerald D. Noble, University of Florida, Gainesville, FL

Alternative splicing (AS) is a phenomenon enabling multiple pre-mRNA isoforms to be transcribed from a single gene. AS changes the nucleotide sequence of pre-mRNA thus expanding the diversity of the proteome without increasing the size of a eukaryotic organism's genome. Previous research has investigated the dynamics of AS events between plant tissues, between plants under various growing conditions, and

between plant species. However, research focused on AS and the resulting transcript isoforms within a population of a single species is scant. We have profiled the diversity of AS events within an association population of Populus deltoides, investigated the effect of AS events on the translated sequence of transcript isoforms, and identified their putative genetic regulators.
Contrary to previous findings in other plant species, alternative 3' splice events were the most abundant AS type in the population rather than intron retention. Approximately 3000 genes were found to have isoforms that resulted in loss of protein domains relative to the reference gene sequences, and approximately 2800 genes express a transcript isoform with a premature stop codon, as a result of AS. Genetic variation can affect AS events via polymorphisms occurring near splice junctions and by impacting proteins involved in the spliceosome machinery. We have identified splice QTLs for genes that undergo AS and are currently inferring their putative biological impact. These findings will elucidate the genetic regulation of AS events and their biological impacts.

W408: Forest Tree

Two Rounds of Ancient RNA-Mediated Gene Duplication Revealed by Whole Genome Mining of TE-Enriched Conifer Genomes

Jian-Feng Mao, Beijing Forestry University, Beijing, China

Gene duplication is an important mechanism for new gene origination and adaptation, in turn driving organismal evolution. RNA-mediated gene duplication or gene retroposition, the formation of gene retrocopies via the enzymatic machinery encoded by retrotransposable elements, was observed as key evolutionary process for developmental innovation and adaptive transition in primarily mammals, fruit flies and model plant. However, knowledge on expansion and evolutionary dynamics of RNA-mediated gene duplication is still limited, especially for the TE-enriched non-model genomes (like conifer genomes) for which retrotransposable elements are prevalent and TE removal is inefficient. By whole genome examining, we detected tens thousands of intact, pre-stop and frame-shit conveying retrocopies for each of the four released conifer (*Pinus taeda*, *Pinus lambertiana*, *Picea abies*, *Pseudotsuga menziesii*) genomes and Ginkgo genome. Two ancient rounds of retrocopy expansion were recovered by plotting Ks distribution, with the most ancient round associates with the origination of land plant, the other with separation between angiosperm and gymnosperm. Further survey in TE-enriched seed plant genomes like wheat, maize and cotton support all two expansion events revealed here, fern and moss genomes support the most ancient one, but no ancient expansion detected in ancestry lineage of land plant, like alga. Tissue specific expression of the retrocopies was also analyzed in conifer genomes. Our study provide valuable information on understanding the complexity of conifer genome and gene retroposition in plant.

W409: Forest Tree

TreeGenes: Turning over a New Leaf

Emily Grau[1], Sean Buehler[1], Nic Herndon[1], Taylor Falk[1], Peter Richter[1], Margaret Staton[2], Stephen P. Ficklin[3] and Jill L. Wegrzyn[1], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Tennessee, Knoxville, Knoxville, TN, (3)Washington State University, Pullman, WA

TreeGenes is a web-based information resource designed to serve the diverse needs of the forest tree genomics research community by uniting information resources with tools visualization and analysis tools. TreeGenes has recently undergone a complete redesign using Tripal, a tool to create and manage genomic database websites. An open source project, Tripal allows developers the flexibility to create their own tools (modules) as well as open communication among Tripal supported repositories.
TreeGenes hosts a range of modules that have expanded functionality to deliver genomic and phenotypic data on >1700 forest tree species. The new Galaxy module allows users to execute analytical workflows with next generation sequencing datasets on high performance computing resources with the click of a button. The Elasticsearch module allows flexible searching and data retrieval between sites such that TreeGenes can share data with Hardwood Genomics Project and the Genome Database for Roseaceae.
TreeGenes is also creating new modules, including: CartograTree which integrates envrionmental, phenotypic, and genotypic data for georeferenced trees. The module works with the Tripal Galaxy to facilitate association mapping and landscape genetics analysis. Source data for CartograTree is imported via the Tripal Plant PopGen Submit module, a pipeline for accepting direct submissions from researchers. This module collects relevant metadata while reducing the burden on the researcher for submission. The new TreeGenes Tripal DIAMOND module offers speed improvements over BLAST and allows sequence similarity searches across numerous genomes and transcriptomes. TreeGenes looks forward to providing the forest tree research community with expanded data and analytical toolsets.

W410: Forest Tree

Updates to the Hardwood Genomics Project, a Website and Database for Tree Genetic and Genomic Data

Ming Chen[1], Bradford Condon[1], Abdullah Almsaeed[1], Jiali Yu[1], Jill Wegrzyn[2], Dorrie Main[3], Stephen P. Ficklin[3] and Margaret Staton[1], (1)University of Tennessee, Knoxville, TN, (2)University of Connecticut, Storrs, CT, (3)Washington State University, Pullman, WA

A major increase in genomic resources for tree species has led to the need for better community access to these massive data sets. The Hardwood Genomics Database (HWG) provides web interfaces to genomic data from economically, ecologically and phylogenetically important angiosperm tree species. The HWG currently includes reference genomes from 4 species, reference transcriptomes from 18 species, and low coverage whole genome sequence data from 10 species. The HWG adds value to this data by providing relevant metadata, unique searching and visualization of the data [1], and opportunities for data analysis. Site data and metadata utilize standard ontologies, paving the way for automated cross-site communication. A BLAST tool [2] is available for searching and comparing sequences, and genomic scaffolds and annotations can be explored through JBrowse [3]. In a major recent addition, the HWG now incorporates gene expression data derived from RNASeq experiments, enabling users to build gene expression heat maps across different sequenced tissues using their own lists of genes. We now offer a bridge to the Galaxy [4] bioinformatics platform, allowing users to not only discover, visualize and download genomic information, but to utilize it directly in predefined analysis workflows. Site visitors will be able to select custom datasets or upload their own data and run common bioinformatics analyses on connected Galaxy instances and view analysis results directly on the HWG website. With the

ongoing addition of new datasets and tools to HWG, tree researchers now have more ways to interact with and utilize public genomic data sets to advance their own research.
1. Chen M, Henry N, Almsaeed A, Zhou X, Wegrzyn J, Ficklin S, Staton M: New extension software modules to enhance searching and display of transcriptome data in Tripal databases. Database 2017, 2017.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J. Mol. Biol. 1990, 215:403–410.
3. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH: JBrowse: a next-generation genome browser. Genome Res. 2009, 19:1630–1638.
4. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A: Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005, 15:1451–1455.

W411: Forest Tree
Sequencing Individual Loblolly Pine Chromosomes using Laser Capture Microdissection Microscopy
Nurul Islam-Faridi, Southern Institute of Forest Genetics, US Forest Service, College Station, TX, Claudio Casola, Texas A & M University, College Station, TX, Craig S. Echt, Southern Institute of Forest Genetics,, Saucier, MS, Wayne Raska, USDA Forest Service, Southern Research Station, College Sttaion, TX, M. a. Majid, Depart. of Ecosystem Science & Management, College Station, TX, Thomas Byram, Texas A&M Forest Service, College Station, TX, Konstantin Krutovsky, Georg-August University of Gottingen, Gottingen, Germany and C. Dana Nelson, Southern Institute of Forest Genetics, US Forest Service SRS, Lexington, KY; USDA Forest Servce, Saucier, MS
Proper sequence assembly of a large genome is a challenge especially for pines that contain considerable repetitive DNA (~80%). The loblolly pine (*Pinus taeda*) genome ($2n = 2x = 24$; 22 Gb/1C) has recently been sequenced, but the ~1.8 million scaffolds in the current assembly are mostly unordered and unmapped on their respective linkage groups (LGs). Sequencing individual chromosomes, obtained through Laser Capture Microdissection Microscopy (LCMM), promises to reduce the complexity of genome assembly and allow for the direct assignment of contigs and scaffolds to LGs. However, obtaining karyotype-quality mitotic spreads is notoriously difficult in plants compared to animals, where chromosome-based sequencing has been done using LCMM (e.g., in frog, *Xenopus tropicalis* and salamander, *Ambystoma mexicanum*); and the task is even more difficult when the chromosomes are large, such as for conifers. To test LCMM-based sequencing in pine, we first developed a protocol for preparing high-quality chromosome spreads on pen-membrane slides. The protocol was then used to micro-dissect loblolly pine (clone 20-1010) chromosome 12 (Ch12) which was facilitated by its special characteristic as being the smallest and the only sub-metacentric chromosome in pines, and thus easily identified under a 40X objective. The micro-dissected Ch12 DNA was then amplified using a whole genome amplification kit (WGA4, Sigma) and verified using LG-specific PCR markers. Forty μl of the verified Ch12 DNA was re-amplified, and 20 μg were generated for DNA sequencing using next generation sequencing platforms. We will report the progress of the project.

W412: Forest Tree
Purifying Selection Patterns in Genes of Two Distantly Related Conifers
Mengmeng Lu, University of Calgary, Calgary, AB, Canada

W413: Forest Tree
RNAi Suppression of Agamous-like Genes Causes Field Sterility in Populus
Haiwei Lu[1], Amy Leigh Klocko[1], Amy Brunner[2], Cathleen Ma[1], Anna Carlina Magnuson[1] and **Steven H. Strauss**[1], (1)Oregon State University, Corvallis, OR, (2)Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA
Concerns over transgene dispersal have limited field studies and commercial use of genetically engineered (GE) trees. We seek to mitigate concerns by producing sexual sterility in poplar. Based on gene sequences from *Populus trichocarpa*, we created two RNA interference (RNAi) constructs, PTG and its matrix-attachment-region flanked version MPG, to suppress expression of the two duplicate *AGAMOUS* (*AG*) orthologs in *P. alba* genotype 6K10, an early flowering female clone. A total of 35 transformed events with four ramets per event and 24 wild-type (WT) control trees were planted as part of a larger field trial in 2011. Six out of 22 flowering PTG events and 11 out of 12 flowering MPG events showed a modified floral phenotype; their floral buds flushed early in the field and the capsules on each catkin often had “carpel-inside-carpel” phenotypes as expected from impairment of AG activity. A complete disruption of ovule and seed production was observed in a number of gene insertion events within both constructs. We also discovered suppression at two AG-like genes (*AGL11*), in sterile events. In all cases, trees appeared normal in their vegetative morphology and growth, and alterations in floral phenotypes were stable over multiple years. RNAi suppression of *AG*-like genes appears to be a safe and effective means of genetic containment in poplar.
We thank the USDA Biotechnology Risk Assessment Grants (no. 2010-33522-21736 and no. 2011-68005-30407) and the Tree Biosafety and Genomics Research Cooperative at OSU for support.

W414: Forest Tree
Genomic Analysis of the Endodormancy-Ecodormancy Transition (EET) in Trees
Anna Conrad[1], Jiali Yu[2], Margaret Staton[2], Jean-Marc Audergon[3], Veronique Decroocq[4], Kevin Knagge[5], Huadong Chen[5], Tatyana Zhebentyayeva[6], Zongrang Liu[7], Christopher Dardick[8], C. Dana Nelson[9] and Albert G. Abbott[10], (1)The Ohio State University, Columbus, OH, (2)University of Tennessee, Knoxville, TN, (3)UR1052 GAFL, INRA, Montfavet, France, (4)UMR BFP1332 - INRA-Universite de Bordeaux, Villenave d'Ornon Cedex, France, (5)David H. Murdock Research Institute, Kannapolis, NC, (6)Clemson University Genomics & Computational Biology Laboratory, Clemson, SC, (7)USDA-ARS, AFRS, Kearneysville, WV, (8)USDA-ARS Appalachian Fruit Research Station, Kearneysville, WV, (9)Southern Institute of Forest

Genetics, US Forest Service SRS, Lexington, KY, (10)Forest Health Research and Education Center, University of Kentucky, Lexington, KY
Due to the broad climate adaptation of perennial trees, phenological traits (e.g. chilling requirement-CR, bloom date-BD) exhibit complex inheritance patterns. Conceptually, these are adaptive responses to abiotic stress. Since in fruiting trees production depends on traits like CR, varieties are selected to phenotypically/genotypically match particular geographic/temperature zones. These genotypes are ideal for study of the genes and gene networks that govern these climate-critical traits. Using genetic approaches, genome-wide association analyses, and functional and comparative genomics in fruit and forest trees, we are investigating the foundational networks of genetic activity that link winter cold stress response, control of the endodormancy-ecodormancy transition (EET), and seed stratification.
We will present our recent results employing specific tree genotypes that exhibit differences in CR and BD to: 1) characterize genotypic effects on the phenylpropanoid gene network transcriptome and metabolome during endodormancy and the EET, 2) examine the genotypic effects on the flux of specific phenylpropanoid intermediates and the timing of resumption of growth post-dormancy, and 3) characterize specific dormancy stage gene network activities, using RNAseq analyses on floral buds sampled through endodormancy, ecodormancy and bloom stages. Currently, we are developing vector/candidate gene constructs to characterize the phenotypes of expression perturbations for specific pathway genes potentially implicated as control points for the EET. Using these functional genomics/metabolomics results, we propose a model for the genetic control of EET that sets the stage for detailed model testing in additional tree species.

W415: Forest Tree
A Chilling-Responsive Demeter-like DNA Demethylase Mediates in Poplar Bud Break
Daniel Conde[1], Anne-Laure Le Gac[2], Mariano Perales[3], Christopher Dervinis[1], Matias Kirst[1], Stéphane Maury[2], Pablo Gónzalez-Melendi[3] and Isabel Allona[3], (1)University of Florida, Gainesville, FL, (2)University of Orleans, Orleans, France, (3)Universidad Politécnica de Madrid, Madrid, Spain
Coordination between growth and environment is essential for the survival of trees in temperate and boreal latitudes. In these regions, the growth rate of perennial plants is markedly regulated by the establishment of periods of dormancy, which consist in the repression of cell division in meristem-containing structures. For many trees, the perception of short days is sufficient to promote cessation of growth and apical bud formation. Restoration of meristem activity and bud break during the spring are trigged by favourable environmental conditions (long days and warm temperatures), once the chilling requirement has been fulfilled.
Active chromatin rearrangement has been proposed as a key step in growth-dormancy transition in trees. We recently uncovered the dynamic changes in genomic DNA methylation (gDNA) levels during dormancy-growth cycle in poplar apex. We showed that the reactivation of cell division in the apical shoot during spring is preceded by a progressive decline of gDNA methylation in the apex tissue. According to this observation, we found that the shoot growth reactivation is preceded by the chilling-dependent induction of poplar DEMETER-LIKE 10 (PtaDML10) DNA demethylase in the apex. Transgenic poplars with down-regulation of PtaDML8/10 showed delayed bud break. By genome wide transcriptome and methylome analysis we identified target genes of the active DNA demethylation trigged by DMLs genetically associated to bud break. These results indicate that a chilling- responsive DEMETER-like DNA demethylase controls the genetic shift from winter dormancy to a condition that promotes apical growth in poplar.

W416: Forest Tree
Characterization of the Molecular and Cellular Role of EVE1 in Plant Vascular Development
Johnathon Blahut, Kelly Balmant, Daniel Conde, Christopher Dervinis, David Oppenheimer and Matias Kirst, University of Florida, Gainesville, FL
A remaining hallmark of the colonization of terrestrial environments is the variation in complexity of the water-conducting tissues across the plant kingdom. The vascular cells of mosses and liverworts can be surprisingly complex with thickened cell walls and plasmodesma-derived perforations. These features mimic the water-conducting tissues observed among the angiosperm clade which has adopted vessels for improved conductive efficiency. In most cases, vessels show greater hydraulic conductivity owing to intervessel pits and perforation plates. To better understand this complexity, we mapped quantitative trait loci associated with vessel-related traits and discovered the *Enlarged Vessel Elements 1* (*EVE1*) gene in poplar. Stems of hybrid poplar trees overexpressing *EVE1* display higher hydraulic conductivity, vessel number per sapwood area and vessel diameter without a concomitant increase in cavitation. Further analysis of relative transcript abundance among unrelated individuals suggests a role of *EVE1* in determining vessel size. Recent protein electrophoresis studies have revealed that the *EVE1* locus produces a small molecular weight membrane protein in *Populus deltoides* stems. Immunolocalization performed with sapling stem sections suggest that the EVE1 protein localizes specifically to developing vessel elements but is largely absent of mature vessels. The discovery and characterization of EVE1 represents a starting point towards identifying other genes involved in plant vascular transport.

W417: Forest Tree
GCMS-Based Metabolomics of Populus deltoides Plants with Modified Gene Activity Prior to and within the Lignin Pathway Reveals Alterations in Carbon Flux to Secondary Metabolism and the Underlying Basis of Altered Biomass Recalcitrance
Timothy J. Tschaplinski[1], Nancy Engle[1], Lee Gunter[1], Robert Sykes[2], Sara S. Jawdy[1], Xiaohan Yang[1], Wellington Muchero[1], William H Rottmann[3], Maud Hinchee[4] and Jerry Tuskan[1], (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)National Renewable Energy Laboratory, Golden, CO, (3)Arborgen, Ridgeville, SC, (4)Agricen Sciences, Pilot Point, TX
The generation of transgenic plants with reduced lignin production and/or altered lignin composition has been a major focus of research to reduce the recalcitrance of cell walls of biomass crops for conversion to biofuels. We have targeted gene knockdowns both within the lignin biosynthetic pathway, as well as upregulation of genes remote to the lignin pathway. In this study, we specifically targeted the knockdown (KD) of cinnamyl alcohol dehydrogenase (CAD) in the lignin pathway and the overexpression (OE) of 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase of the shikimic acid pathway of *Populus deltoides* 'WV94', both of which resulted in the reduced recalcitrance of stemwood. Gas chromatography-mass spectrometry (GCMS) of trimethylsilyl derivatives of metabolites was conducted on aqueous ethanol (80%) extracts

of fast-frozen stem biomass that had been previously lyophilized and milled. Compared with empty vector control plants, down-regulated CAD stem bark peels demonstrated elevated hydroxycinnamates upstream of the lignin pathway, offset by declines in some higher-order salicylates and flavonoids. In contrast, plants with up-regulated EPSP had greatly increased flavonoids and other closely-related metabolites, but also had increases in hydroxycinnamates, including caffeic acid and its ester conjugates, but to a lesser extent. Therefore, this latter transgenic construct with up-regulated activity in the shikimic acid pathway increased carbon flux to the lignin pathway, but diverted that flux to flavonoid metabolism. The results of these analyses will be discussed in the context of reducing biomass recalcitrance and the potential use of these transgenic lines as potential biomass feedstocks.
Key words: metabolomics, mass spectrometry, lignin, hydroxycinnamates, flavonoids, *Populus*

W418: Forest Tree

Identification of *Populus* Small RNAs Responsive to Symbiosis with Mycorrhizal Fungi *Laccaria bicolor* and *Rhizophagus irregularis*

Ritesh Mewalal[1], Hengfu Yin[2], Rongbin Hu[1], Sara S. Jawdy[1], Patrice Vion[3], Jerry Tuskan[1], François Le Tacon[3], Jessy Labbe[1] and **Xiaohan Yang**[1], (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)Chinese Academy of Forestry, Fuyang, China, (3)INRA, UMR 1136 INRA-University of Lorraine, Nancy, France

Ecto- and endomycorrhizal colonization of *Populus* roots have a positive impact on the overall health and stress tolerance of this economically important crop. Establishment and maintenance of mutualistic symbiosis between plant and microbes are associated with extensive morphological changes orchestrated by the genetic reprogramming in both organisms. While previous transcriptome sequencing effort has revealed multiple small proteins associated with this intricate interaction, the small RNA (sRNA) landscape, including micro RNAs (miRNAs), in this interaction has been unexplored. In this study, we aimed to determine the impact of sRNAs in the roots of two poplar species (*i.e*. *Populus deltoides* and *P. trichocarpa*) which were colonized by an ectomycorrhizal fungus (EMF), *Laccaria bicolor* and an arbuscular mycorrhizal fungus (AMF), *Rhizophagus irregularis*. Through sequencing of sRNA-enriched libraries obtained from poplar-AMF or ECM symbioses, we identified differential expression of sRNAs between fungal treatment and control. We predicted 287 putative miRNAs in *P. deltoides* and 357 putative miRNAs in *P. trichocarpa* , among which 34 and 130 unique miRNAs were significantly differential expressed, respectively. Also, we predicted 61 gene targets for 41 unique miRNAs in *P. deltoides*, and 157 gene targets for 58 unique miRNAs in *P. trichocarpa* (p-value <0.05). Finally, since sRNAs may be a source of novel small open reading frames (sORFs), we examined the coding potential of the differentially expressed sRNAs and predicted sORFs from some of these sRNAs. The findings of the current study highlight the important role of sRNAs in *Populus*-microbe interactions.

W419: Forest Tree

Multiple Approaches to Dissect Fusiform Rust Resistance in *Pinus taeda* L

Daniel Ence, University of Florida, Gainesville, FL

Mapping the specific loci that regulate phenotypic traits in conifers is a major undertaking because of their very large genomes. However, the release of the annotated loblolly pine (*Pinus taeda* L.) genome may allow fine-mapping of Mendelian traits that are economically critical, such as for disease-resistance. Here we present the results of work mapping Fusiform rust resistance locus 1 (*Fr1*) in elite rust resistant loblolly pine trees. Fusiform rust is a disease incited by the fungus *Cronartium quercuum f.sp. fusiforme* (Cqf) on southern pines (where it causes galls on stems and branches) and on oaks (where it causes minimal leaf damage). Fusiform rust is a major disease threat to the timber industry in the US. Rust galls cause yield losses that exceed US$100M/year. During the genome annotation process, an expressed sequence tag (EST) was identified that contains a single nucleotide polymorphism (SNP) mapping to the locus (*Fr1*) that interacts with the fungal avirulence gene, *Avr1*. This EST aligns to a transcript from RNA-sequencing data and a TIR-NB-LRR protein, thus identifying it as a candidate *Fr1* gene.
In order to further characterize the *Fr1* locus, we assembled the transcriptomes of 92 elite rust-resistant loblolly pine genotypes from five pine-growing regions, identifying candidate resistance genes in the process. Next we aligned these transcripts to the loblolly pine genome and calculated population genetic parameters. These results enable analysis of the diversity and conservation of resistance genes that interact with Cqf and present a foundation for further characterization of *Fr1* and other resistance loci.

W420: Forest Tree

Host Responses to the Pitch Canker Pathogen in a Resistant and Susceptible Pine Species

Erik A. Visser[1], Jill L. Wegrzyn[2], Alexander A. Myburg[1] and Sanushka Naidoo[1], (1)Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Pretoria, South Africa, (2)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

The fungal pathogen *Fusarium circinatum* threatens pine species. Transcriptome assembly from RNA sequencing data provides a platform to investigate host responses to pests and pathogens in non-model species. We investigated transcriptomic responses at three and seven days post inoculation (dpi) in a resistant (*Pinus tecunumanii*) and susceptible (*P. patula*) pine species during *F. circinatum* challenge through transciptome assembly. Twelve *de novo* and eight genome guided Trinity assemblies were generated for each species from RNA-seq data and combined, resulting in 28,622 contigs for *P. tecunumanii* and 52,734 for *P. patula*. Read data was mapped to the assembled transcriptomes and differentially expressed genes (DEGs), inoculated relative to mock-inoculated, identified. A larger disparity in the amount of DEGs between timepoints was seen for *P. patula* (323 at 3 dpi; 7,453 at 7dpi; 155shared) than *P. tecunumanii* (735 at 3 dpi; 2,499 at 7dpi; 63 shared). At 3 dpi in *P. patula* a large number of expansins showed up-regulation, indicating cell wall expansion, while in *P. tecunumanii* there were enriched gene ontology (GO) terms associated with jasmonic acid (JA) signalling and induced systemic resistance (ISR), terms classically associated with defence against necrotrophs. Host defence and JA signalling terms only appear in *P. patula* at 7 dpi, suggesting a delayed response. At 7 dpi, ISR terms are absent in *P. tecunumanii*, instead there is enrichment for salicylic acid signalling and systemic acquired resistance. Along with the low number of shared DEGs between time points, this could suggest a switch in defence response.

W421: Forest Tree

Introducing the Redwood Genome Project
Alison Dawn Scott, Department of Plant Sciences, University of California, Davis, Davis, CA

W422: Fruit/Nuts
Genome-Wide Association Study of Partial Resistance to Bacterial Canker of Apricot
Mariem Omrani[1,2], Guillaume Roch[3], Alain Blanc[1], Cindy E. Morris[2] and Jean-Marc Audergon[1], (1)UR1052 GAFL, INRA, Montfavet, France, (2)UR407 Pathologie Végétale, INRA, Montfavet, France, (3)CEP Innovation, Lyon, France
Apricot, a highly valuable crop is threatened by the growing importance of bacterial canker caused by *Pseudomonas syringae*. Among the key factors able to control the disease, genetic improvement is a promising measure. The variability of susceptibility on branches and the characterization of genetic determinants through a genome wide association study were thus investigated on a core-collection.
73 accessions were annually inoculated in the orchard with an aggressive strain of the bacterium for 4 years. Phenotypic data about the length of both external canker (lgc) and superficial browning (bs) of tissues were collected. The analysis displayed a highly environmental-dependent genetic variation with broad-sense heritabilities of lgc and bs reaching respectively 59% and 78% for the most severe year. Considering the two variance-maximizing years, genetic (G) and genetic x year (GxY) BLUP were predicted for each variable using a linear mixed model. Association analysis were performed with a 63,236 SNP set through both a multi-variate (GEMMA) and a multi-locus genome-wide analysis. By exploiting the between-years (multi-locus model on G and GxY terms) and between-phenotypes (multivariate model on lgc-bs G terms) correlations, 11 significant associations have been detected. Among them, two SNP impacting both lgc and bs expressions over the two studied years and explaining 41% and 31% of the total phenotypic variance were identified on chromosomes 5 and 6. A long-range linkage disequilibrium had been noticed between these two markers suggesting a co-selection effect. The associated SNP reported from this work will open up new opportunities for a Marker-assisted selection strategy.

W423: Fruit/Nuts
Differential Transcriptomic Responses of a Tolerant and Susceptible Avocado Cultivar to Phytophthora Root-Rot
Alice Hayward, Stephen Fletcher, Madeleine Gleeson, Christopher O'Brien, William Nak and Neena Mitter, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia
Avocado (*Persea Americana* Mill.) is one of the fastest growing horticultural industries globally. However, despite a 30% increase in the last decade, production is unable to keep pace with the booming demand. One of the most serious threats to productivity is the ubiquitous soil borne pathogen *Phytophthora cinnamomi*, causal agent of Phytophthora Root Rot (PRR). An invasive Oomycete, Phytophthora exhibits a complex life strategy of hemibiotrophy, eventually destroying the feeder roots that are essential for plant water and nutrient acquisition. At present there is no cost-effective prevention strategy for PRR. Control relies on the selection of tolerant avocado rootstocks, integrated with optimised drainage, mulching, nutrition, and chemical treatment of trees.
A number of breeding programs globally are focused on developing PRR tolerant avocado rootstocks. These show reproducible tolerance under field conditions with improved outcomes on fruit yield and quality. However, the fundamental basis of this tolerance is only just beginning to be explored. In this study, the transcriptome and small RNA profiles of a tolerant avocado cultivar (T) to PRR was compared with that of a highly susceptible cultivar (S) at different time-points after infection. Both cultivars were colonised by the pathogen and exhibited activation of many cognate stress and pathogen response genes. However, the tolerant cultivar showed reduced pathogen titre and root damage, and activation of specific genes not shown in the susceptible cultivar. These genes included members of the MATE and ABC toxin transporter families as well as Cytochrome P450s, Lipoxygenases, Transferases, Receptor-Like Kinases and Transcription Factors. Candidate genes from these families are being investigated further for potential roles in PRR tolerance in avocado.

W424: Fruit/Nuts
Genomic Analysis for Molecular Breeding of Almond
Diane Mather, School of Agriculture, Food and Wine, Waite Research Institute, The University of Adelaide, Glen Osmond, Australia
The self-incompatibility (SI) system of almond presents constraints in almond production (requiring pollinator clones in orchards) and breeding (limiting the cross combinations that can be used). Using the self-fertile cultivar Lauranne as a parent, the Australian Almond Breeding program has developed self-fertile cultivars. These will be important parents in future crossing cycles. For efficient selection among juvenile progeny, it is important to have high-throughput molecular markers that can distinguish the S_f (self-fertility) allele of the stylar RNase (*S-RNase*) gene from SI alleles of that gene.
We subjected amplicons from 48 almond clones (representing 16 *S*-locus haplotypes) to short-read sequencing and assembled complete sequences for several *S*-locus haplotypes and 11 *S-RNase* alleles. Novel variation was detected in the *S*-locus structure and the *S-RNase* sequence. KASP™ assays were designed for a SNP that distinguishes S_f from other *S-RNase* alleles. To investigate how the S_fRNASE differs from SI RNASEs, we constructed three-dimensional models of predicted products.
To support the extension of molecular breeding to other traits in almond, we applied genotyping-by-sequencing to Nonpareil × Lauranne F_1 progeny. Linkage maps were constructed. SNPs were selected for KASP™ assay design. Progeny from several Nonpareil crosses were genotyped and linkage maps were improved. QTL were mapped for nut and kernel traits. Markers were identified that that could be used to select for or against the 'papershell' trait of Nonpareil. With additional phenotyping and genotyping, QTL will be mapped for additional traits.

W425: Fruit/Nuts
The Pomegranate Genome and the Genomics of Punicalagin Biosynthesis
Zhen Yue, BGI genomics, Shenzhen, China
Pomegranate (*Punica granatum L.*) is a perennial fruit crop grown since ancient times that has been planted worldwide and is known for its functional metabolites, particularly punicalagins. We have sequenced and assembled the pomegranate genome with 328 Mb anchored into nine pseudo-chromosomes and annotated 29 229 gene models. A Myrtales lineage-specific whole-genome duplication (WGD) event was detected

that occurred in the common ancestor before the divergence of pomegranate and Eucalyptus. Repetitive sequences accounted for 46.1% of the assembled genome. We found that the integument development gene INNER NO OUTER (INO) was under positive selection and potentially contributed to the development of the fleshy outer layer of the seed coat, an edible part of pomegranate fruit. The genes encoding the enzymes for synthesis and degradation of lignin, hemicelluloses, and cellulose were also differentially expressed between soft- and hard-seeded varieties, reflecting differences in their accumulation in cultivars differing in seed hardness. Candidate genes for punicalagin biosynthesis were identified and their expression patterns indicated that gallic acid synthesis in tissues could follow different biochemical pathways. The genome sequence of pomegranate provides a valuable resource for the dissection of many biological and biochemical traits and also provides important insights for the acceleration of breeding. Elucidation of the biochemical pathway(s) involved in punicalagin biosynthesis could assist breeding efforts to increase production of this bioactive compound.

W426: Fruit/Nuts
Citrus Genomes and Excavation of Apomixis Genes
Qiang Xu[1], Xia Wang[1], Yuantao Xu[1], Siqi Zhang[1], Li Cao[2] and Xiuxin Deng[1], (1)Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China, (2)Southwest University, Chonqing, China
Citrus is one of the most important fruit crops worldwide. The best known citrus species are the sweet orange, mandarin, pummelo, grapefruit and lemon. Change of reproduction mode from seed- to grafting propagation and emergence of apomixes is a notable feature of Citrus domestication. In the past five years, we de novo assembled genomes from five citrus species ranging from primitive, wild and modern citrus. High-quality genomes of sweet orange and pummelo were built up by single molecule sequencing data (PacBio long reads) in combination with the high-throughput short-read (Illumina) sequencing technology. The pummelo genome is of high quality, with the contig N50 and N90 are 2.2 Mb and 70 kb, respectively. The sequence contiguity represents at least 18-fold improvement of the published genomes as evaluated by contig N50. The genome sequence, annotation files and raw data were deposited at http://citrus.hzau.edu.cn/orange/. Toward understanding the asexual reproduction caused by apomixes, we narrowed the nucellar polyembryony locus to a region of 80kb and 11 candidate genes, and one of them, CitRWP, shows specific expression in ovule and asexual embryos. Interestingly, a MITE insertion has been observed at the promoter region of CitRWP, and this polymorphism is cosegregated with the polyembryony. Transcriptome analysis of four tissues and embryos from seven species revealed the close developmental relationship of sexuality and apomixes in citrus. The genomic analysis of citrus ranging from primitive to modern provides new insights into the apomixes and as an example, a better understanding of domestication characteristics of perennials.

W427: Fruit/Nuts
Chromosome-Scale Genomic Resources for High Resolution Genetic Analysis in Pistachio

William J. Palmer[1], Ewelina Jacygrad[2], John E. Preece[3], Deborah Golino[2], Salih Kafkas[4] and Richard Michelmore[1], (1)Genome Center, University of California, Davis, Davis, CA, (2)Foundation Plant Services, Davis, CA, (3)Nat'l Clonal Germplasm Repository, Davis, CA, (4)University of Cukurova, Adana, Turkey
Over 99% of the $1.6 billion American pistachio industry occurs in California's Central Valley. The majority of scions here are clones of *Pistacia vera* cultivars. *P. vera* scions are typically grafted on to UCB-1, an interspecific hybrid rootstock. UCB-1 rootstock is the offspring of two clonal *P. atlantica* and *P. intergerrima* parents, which are both outbreeding and heterozygous.
In orchards, there is considerable variation in tree size and yield that often necessitates costly removal and replanting of smaller individuals. In the absence of disease or environmental factors, it seems likely that genetic segregation is driving variation in rootstock and overall performance. Development of diagnostic markers for rootstock vigor would therefore be highly beneficial, allowing removal of seedlings with expected poor performance *prior to* planting. Genomic resources are currently limited for this valuable crop. We are therefore developing foundational resources to enable genetic analysis and development of molecular markers for commercial traits.
We have generated chromosome-scale draft genome assemblies for *P. vera (scion)* as well as *P. atlantica*, and *P. integerrima (UCB-1 parents)* together with linkage data generated from a mapping population of over one thousand UCB-1 seedlings in both experimental and commercial orchards. To generate these data, we leveraged short and long read technologies including PacBio, 10X, Dovetail Hi-C, GBS and Skim-Seq. In addition, the UCB-1 mapping population has been phenotyped for numerous traits over several years - enabling the identification of commercially relevant QTLs, as well as dissection of the interactions between scion, rootstock, and the environment.

W428: Fruit/Nuts
The Genetics of Ester and Phenylpropene Volatile Production in Ripe Apple Fruit
Ross G. Atkinson[1], David Chagné[2], Xiuyin Chen[1], Sumathi Tomes[1], Mindy Y. Wang[1], Martin B. Hunt[2], Ratnasiri Maddumage[1], Adam J. Matich[2], Daryl D. Rowan[2], Edwige J.F. Souleyre[1] and Yar-Khing Yauk[1], (1)The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, (2)The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand
Flavor/aroma production is a complex trait consisting of compounds produced by many biosynthetic pathways, each under different genetic, hormonal and environmental control. With recent advances in genome sequencing, transcriptomics and transgenesis the process of flavor gene identification in non-model crops has accelerated and a more gene centric approach can now be considered for breeding high-flavor fruit. Our research is focused on unravelling the genetics of ester, phenylpropene and terpene volatile production in apples. In 'Royal Gala' apples, the major 'fruity' esters are hexyl acetate, butyl acetate and 2-MBA. The production of these esters is controlled by a major QTL on LG2 that co-locates with the ester biosynthetic gene *alcohol acyl transferase1* (*MdAAT1*). A similar approach was used to identify genes controlling the production of estragole, a phenylpropene that contributes 'spicy' notes to some varieties. One QTL for estragole production co-locates with the *o-methyl transferase1* gene on LG1, whilst a second major QTL co-locates with *MdAAT1*. Biochemical analysis shows that *MdAAT1* is required for the production of volatile esters and for the production of *p*-hydroxycinnamyl acetates that are substrates for producing phenylpropenes such as estragole in ripe apple fruit. The importance of the *MdAAT1* gene in ester and phenylpropene production was validated

in ‘Royal Gala’ knockdown lines that produced significantly reduced 2-MBA and estragole levels in ripe fruit. Transient over-expression of AATs from strawberry and tomato showed these enzymes can also produce *p*-hydroxycinnamyl acetates, indicating that ripening-related AATs are likely to link volatile ester and phenylpropene production in many different fruit.

W429: Fruit/Nuts

Subgenome Dominance in Interspecific Hybrids and Allopolyploids

Patrick Edger, Michigan State University, EAST LANSING, MI

The importance and applications of polyploidy have long been recognized, from shaping the evolutionary success of flowering plants to improving agricultural productivity. Recent studies have shown that one of the parental subgenomes in ancient polyploids is generally more dominant - having both retained more genes and being more highly expressed - a phenomenon termed subgenome dominance. How quickly one subgenome dominates within a newly formed polyploid, if immediate or after millions of years, and the genomic features that determine which genome dominates remain poorly understood. To investigate the rate of subgenome dominance emergence, we examined gene expression, gene methylation, and transposable element (TE) methylation in natural allopolyploids, resynthesized allopolyploids, and interspecific hybrids. We show that subgenome expression dominance occurs instantly following the hybridization of two divergent genomes and that subgenome expression dominance significantly increases over generations. Our analyses reveal that the subgenome differences in levels of TE methylation mirror the increase in expression bias observed over the generations following the hybridization. These findings not only provide important insights into genomic and epigenomic shock that occurs following hybridization and polyploid events, but may also contribute to uncovering the mechanistic basis of subgenome dominance. The presentation will include results from analyses of *Brassica*, *Fragaria*, and *Mimulus*.

W430: Fruit/Nuts

Genotyping By Sequencing and QTL Analysis for Phenolic Compound Content in Japanese Plum (*Prunus salicina* Lindl.)

Igor A. Pacheco Cruz, INTA - Universidad de Chile, Santiago, Chile

Flavonoids are key elements for colour and functional compound content in plant-derived foods; among them, anthocyanins have a bioactivity potential and are key contributors for fruit color in Rosaceae. Discovery of genetic markers for assisted selection of new varieties with increased functional compound content, as well as identification of candidate genes controlling fruit secondary metabolite content, are central objectives in our research. In order to pursue these aims in a commercially important species for Chilean fruit production, we performed an initial genetic analysis for individual anthocyanin content in a Japanese plum (*Prunus salicina* L.) F1 progeny obtained from the cross between selection '98-99' and the cultivar 'Angeleno'. UHPLC-DAD-Orbitrap-MS analyses were employed for compound identification in selected samples. HPLC-DAD was employed for anthocyanin profiling in skin (SK) and flesh (FL) methanolic extracts from fruits harvested at commercial maturity in 90 F1 individuals, for which genotypic data for 4058 SNPs were available after Genotyping-By-Sequencing. In addition, spectrophotometric methods were employed for total phenols, total procyanidins, total flavonoids and antioxidant activity in the F1 family. Two HPLC signals (peaks), previously identified as cyanidin-3-glucoside (C3G) and cyanidin-3-rutinoside (C3R) were detected in concentrations hundreds to thousand times higher than the other detected peaks. C3G and C3R, as well as minor-content anthocyanins (here called "M" and "K"; identification in progress) showed significant genetic contributions to phenotypic variance ($p < 0{,}001$). Parametric and non-parametric QTL-analyses (Interval Mapping and Kruskal-Wallis, respectively) were performed using linkage maps for both parents, separately. Significant marker-trait associations ($p < 0{,}05$) were detected mainly for compounds present in skin samples. QTLs located in linkage-groups (LG) 1, 3, 4, 5 and 7 of *Prunus* genome explained between 38% and 45% of the total phenotypic variance. The results here presented constitute the first elements in a study for genetic architecture dissection of potentially bioactive secondary metabolite content in Japanese plum fruits. This work is being supported by grants FONDECYT inicio 11150662, PAI-CONICYT 79140020, FONDECYT postdoctorate 3160080 and FONDAP Center for Genomic Regulation.

W431: Fruit/Nuts

The Apple Genome and Methylome : New Gene Regulatory Roles for DNA Methylation ?

Nicolas Daccord[1], Jean-Marc Celton[1], Gareth Linsmith[2], Claude Becker[3], Nathalie Choisne[4], Elio Schijlen[5], Henri van de Geest[5], Luca Bianco[2], Diego Micheletti[2], Riccardo Velasco[2], Erica Di Pierro[6], Jérôme Gouzy[7], Jasper Rees[8], Philippe Guérif[1], Hélène Muranty[1], Charles-Eric Durel[1], Francois L. Laurens[1], Yves Lespinasse[1], Sylvain Gaillard[1], Sébastien Aubourg[1], Hadi Quesneville[4], Detlef Weigel[9], Eric van de Weg[10], Michela Troggio[2] and Etienne Bucher[1], (1)IRHS, INRA, Université d'Angers, AgroCampus Ouest, Beaucouzé, France, (2)Research and Innovation Centre, Edmund Mach Foundation, San Michele all'Adige, Italy, (3)Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna Biocenter, Vienna, Austria, (4)URGI, INRA, Université Paris-Saclay, Versailles, France, (5)Plant Research International, Wageningen University & Research, Wageningen, Netherlands, (6)University of Milan, Milan, Italy, (7)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (8)Agricultural Research Council of South Africa, Pretoria, South Africa, (9)Max Planck Institute for Developmental Biology, Tübingen, Germany, (10)Plant Breeding Wageningen University & Research centre, Wageningen, Netherlands

Accurate sequence information and genome assemblies are critical for studies on genomic and epigenomic variations. Although partial genome information was already available for apple (Velasco et al. 2010), the assembled sequence is fragmented and lacks repeated regions. First, using second- (Illumina) and third-generation sequencing and optical mapping technologies (PacBio and BioNano), we have generated a high quality genome assembly of a 'Golden Delicious' doubled haploid tree (Daccord, Celton et al. 2017). Our de novo assembly resulted in a genome of 649.7 Mb, with a N50 of 5.6 Mb. Seventeen pseudo-chromosomes were constructed and validated using a high density integrated genetic linkage map (Di Pierro, E. A. et al. 2016) and linkage disequilibrium analysis. Using similarities, transcript resources and the EuGene predictor/combiner (Foissac et al. 2008), 45,115 protein coding genes were predicted and tagged, when possible, by putative function.
In a second time, to understand the potential role of epigenetic marks on fruit development, we constructed genome-wide DNA methylation maps that compared different tissues and two isogenic apple lines that produce large or small fruits. We established general correlations

between methylation patterns in promoters and gene expression. Moreover, this led to the identification of differentially methylated regions that may be associated with genes involved in fruit development.

W432: Fruit/Nuts

The Sweet Cherry Genome for Genomics-Based Breeding

Kenta Shirasawa[1], Kanji Isuzugawa[2], Mitsunobu Ikenaga[3], Yutaro Saito[2], Toshiya Yamamoto[4], Hideki Hirakawa[1] and Sachiko Isobe[1], (1)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (2)Yamagata Integrated Agricultural Research Center, Sagae, Yamagata, Japan, (3)Hokkaido Research Organization, Naganuma, Hokkaido, Japan, (4)National Agriculture and Food Research Organization, Tsukuba, Ibaraki, Japan

Genome sequence information on sweet cherry (*Prunus avium*) had been lacked, even though those on many Rosaceae crops, apple, pear, peach, strawberry etc., have been released. This situation made it difficult to perform efficient breeding program in sweet cherry, an economic important fruit crop in Japan. To assist future breeding as well as genetic and genomic studies for sweet cherry, we therefore established genomic resources such as whole-genome sequence data, high-density genetic maps with single nucleotide polymorphisms based on restriction-site associated DNA sequencing technology, DNA markers based on whole-genome resequencing analysis of modern varieties. In addition, we developed a user-friendly genome database, DBcherry (http://cherry.kazusa.or.jp). The information obtained from this study was really helpful and useful to identify genetic loci for agronomically important traits in our breeding programs for sweet cherry.

W433: Fruit/Nuts

Genomics-Based Tools for the Walnut (*Juglans regia* L.) Breeding Program in California

Annarita Marrano[1], Pedro J Martínez-García[1], Luca Bianco[2], Gina M. Sideli[1], Charles A. Leslie[1], Kristian A. Stevens[3], Marc Crepeau[3], Michela Troggio[2], Charles H. Langley[3] and David B. Neale[4], (1)Department of Plant Science, University of California, Davis, Davis, CA, (2)Research and Innovation Centre, Edmund Mach Foundation, San Michele all'Adige, Italy, (3)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (4)Department of Plant Sciences, University of California, Davis, Davis, CA

Persian walnut (*J. regia* L.) is horticulturally the most developed and widely cultivated for nut production. The USA is one of the leading countries in the global production and export of walnut crop. California leads the nation accounting for 99% of US walnut production, and owns one of the most significant walnut breeding programs worldwide. The Walnut Improvement Program (WIP) of University of California, Davis works to address the needs of Californian walnut growers by emphasizing yield, harvest date, kernel color and in-shell traits. Thus far, the WIP has implemented controlled outcross pollinations and phenotypic selection over several generations. However, selection based only on phenotyping can be inefficient, notably for complex polygenic traits. Nowadays, genomics provides innovative tools to dissect the genetic architecture of complex traits and, thus, assist breeders in the selection of superior genotypes for future breeding cycles. The primary goal of our research is the integration of conventional breeding techniques with genomics tools in the WIP. We re-sequenced 27 founders of the WIP at high coverage (80X), identifying over 17.8M Single Nucleotide Polymorphisms (SNPs). We selected 609,749 high-quality SNPs to build a new Axiom® Walnut700K SNP array. We then applied this genomic tool for genotyping over 1200 accessions, including 1130 progeny of 48 WIP families and 24 founders. Thanks to this highly informative tool, we could re-construct the WIP pedigree, which is a pivotal step for carrying out gene mapping and genomic prediction studies towards the full application of molecular breeding for walnut in California.

W434: Fruit/Nuts

Towards Developing a Chromosome Scale Reference Genome Sequence of Blueberry

Massimo Iorizzo[1], Hamed Bostan[2], Rishi Aryal[3], Lisa J. Rowland[4], Juan Zalapa[5] and Hamid Ashrafi[3], (1)Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC, (2)Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, (3)North Carolina State University, Raleigh, NC, (4)USDA-ARS Genetic Improvement of Fruits and Vegetables Laboratory, Beltsville, MD, (5)University of Wisconsin-Madison, Madison, WI

Blueberry (*Vaccinium corymbosum*) is one of the few fruit crops native to North America. Driven by recognition of the health benefits associated with blueberry consumption, blueberry production and demand continue to expand globally. To sustain this growth, there is a growing need of new blueberry cultivars integrating traits that can provide benefits to both stakeholders and consumers. Use of molecular breeding techniques can significantly expedite the conventional breeding process to fulfill this need. However, limited genomic resources are available to implement modern breeding techniques for blueberry. We used a diploid, highly heterozygous accession (W85-23), to establish a chromosome scale assembly by integrating PacBio, Chicago and Hi-C sequences. The genome size is estimated 670 Mb. In total, 55 Gb of PacBio sequences, corresponding to 83 fold nt coverage, were used for assembly using FALCON assembler and FALCON_UNZIP. The first draft of the blueberry genome assembly covered 635 Mb in 2,939 primary contigs (>1kb), with an N50 length of 423 Kb and representing 95% of the estimated genome size. Over 417 Mb, accounting for 65% of the primary contigs lenght, was assembled into associated contigs (10,450), confirming the heterozygous nature of this genome. Systematic integration of Hi-Raise and Hi-C data allowed us to assemble 349 scaffolds increasing N50 of the assembly to 50 Mb with longest scaffolds representing the twelve blueberry chromosomes. Efforts to anchor the current genome to linkage maps, develop a chromosome scale phased genome and integrating transcriptome data from multiple sources will be discussed.

W435: Functional Annotations of Animal Genomes (FAANG)

Introduction of the Functional Annotation of Animal Genomes (FAANG) Project

Christopher K. Tuggle, Iowa State University, Ames, IA

This presentation will provide a short introduction to the Functional Annotation of Animal Genomes (FAANG) Consortium, including the history and goals of the FAANG Project.

W436: Functional Annotations of Animal Genomes (FAANG)
The Central Role of RNA in Understanding of the Central Dogma
Zhihua Jiang, Washington State University, Pullman, WA
The central dogma of biology popularly known as "DNA makes RNA makes protein" was proposed by Francis Crick sixty years ago. However, research has clearly shown that many biological systems are pretty complex. Due to use of alternative promoters, polyadenylation and splicing, for example, our genome expresses a large number of alternative transcripts, but uses a limited number of genes. Consequently, a gene is functionally more diverse as alternative transcripts act quantitatively, qualitatively and/or epigenetically. In order to thoroughly study RNA diversity and dynamics, we have successfully developed a whole transcriptome termini site sequencing (WTTS-seq) method to capture 3'-ends of transcripts. At the workshop, I will demonstrate how gene knockout, binge feeding, developmental transition and gender affected and/or altered use of alternative polyadenylation sites, which in turn revealed potential links from genome to phenome. In addition, we have also successfully developed a method called whole transcriptome start site sequencing (WTSS-seq), which is ready for the community to use to capture the 5'-ends of transcripts. Based on our current research, we propose that alternative transcripts are minimal functional units in genomes and the traditional central dogma concept should be now examined under a systems biology approach.

W437: Functional Annotations of Animal Genomes (FAANG)
Facilitating the Generation of a Bovine Methylation Array
Stephanie McKay, University of Vermont, Burlington, VT

W438: Functional Annotations of Animal Genomes (FAANG)
Genome-Wide Identification of Active Enhancers in Bovine Skeletal Muscle and Adipose Cells
Honglin Jiang, VirginiaTech, Blacksburg, VA

W439: Functional Annotations of Animal Genomes (FAANG)
Herpesvirus Reactivation as a Model to Study Spatiotemporal Gene Regulation
Yoshihiro Izumiya, University of California, Davis, Davis, CA

W440: Functional Annotations of Animal Genomes (FAANG)
Genome Wide Identification and Annotation of Functional Regulatory Regions in Livestock Species
Huaijun Zhou[1], Pablo J. Ross[1], Colin Kern[1], Perot Saelao[1], Ying Wang[1], Michelle M. Halstead[1], Kelly Chanthavixay[1], Ian Korf[2], Mary E. Delany[1], Hans H. Cheng[3], Juan F. Medrano[4], Alison Van Eenennaam[5], Catherine W. Ernst[6] and Christopher K. Tuggle[7], (1)Animal Science, University of California, Davis, CA, (2)The Genome Center, University of California, Davis, Davis, CA, (3)USDA, ARS, ADOL, East Lansing, MI, (4)University of California, Davis, CA, (5)University of California, Davis, Davis, CA, (6)Department of Animal Science, Michigan State University, East Lansing, MI, (7)Iowa State University, Ames, IA
The recent international FAANG (Functional Annotation of ANimal Genomes) initiative has stimulated efforts to functionally annotated important livestock species, which will ultimately be leveraged to improve production efficiency, animal welfare, and food safety. As one of the FAANG pilot projects coordinated by UC Davis, we present the current progress in generating and analyzing data from chicken, cattle, and pig. Samples were collected from adipose, cerebellum, cortex, hypothalamus, liver, lung, muscle, and spleen in two male biological replicates from each species, allowing the identification of both universal and tissue-specific functional elements. High depth of RNA-seq identified 9,393 long non-coding RNAs in chicken, 7,235 in cattle, and 14,428 in pig. From DNase-seq in chickens, 132,362 open chromatin regions were identified, many of which are tissue-specific. Genes present in these open chromatin regions show generally higher expression in our RNA-seq data. In chicken, we have identified a total of 31,174 H3K4me3 peaks, 79,144 H3K27me3 peaks, 34,091 H3K27ac peaks, 44,664 H3K4me1 peaks, and 21,710 CTCF peaks in liver, lung and spleen. For the same tissues in pig, 35,081 H3K4me3 peaks were identified, 104,640 H3K27me3 peaks, 133,689 H3K27ac peaks, 38,247 H3K4me1 peaks, and 26,585 CTCF peaks. Preliminary chromatin state models built using this data show good correlation with gene expression and chromatin states representative of promoters, enhancers, and insulators. Work is ongoing to improve existing data and include the remaining brain, adipose, and muscle tissues, leading to the final genome-wide chromatin state predictions and comprehensive catalogs of regulatory elements for these three species.

W441: Functional Annotations of Animal Genomes (FAANG)
An Update on FR-Agencode
Sylvain Foissac[1], Sarah Djebali[1], **Andrea Rau**[2], Sandrine Lagarrigue[3], Herve Acloque[1], Elisabetta Giuffra[2] and Fr-AgENCODE group, (1)INRA-GenPhySE, Castanet-Tolosan, France, (2)UMR GABI, INRA, Université Paris Saclay, Jouy en Josas, France, (3)Agrocampus Ouest - INRA, UMR1348 PEGASE, Rennes, France
The Fr-AgENCODE project aims to generate multi-species functional genome annotations by applying high-throughput molecular assays on tissues/cells relevant to immune and metabolic traits. From two males and two females per species (pig, cattle, goat, chicken), strand-oriented RNA-seq gene expression and ATAC-seq chromatin accessibility assays were performed on liver and two T-cell types (CD3+CD4+, CD3+CD8+) sorted from blood (mammals) or spleen (chicken). Chromosome Conformation Capture (in situ Hi-C) was also carried out on liver. While most (50–80%) RNA-seq reads mapped to annotated exons, thousands of novel transcripts, extensions of annotated protein-coding genes and new lncRNAs were found. A significant proportion of called peaks in the ATAC-seq reads were found in promoter regions (36–66%), and peaks were found to preferentially accumulate around gene starts (TSS) compared to gene ends (TTS). Principal component analyses for both RNA-seq and ATAC-seq highlighted clusters characterized by cell type and sex in all species. Using 40kb-resolution interaction maps generated with the Hi-C data, we identified topologically-associating domains (TADs) and active "A" versus inactive "B" compartments, which were characterized by significantly different gene density and chromatin accessibility. Differentially expressed genes

between cell types were enriched for immunity- or metabolism-related terms, and differentially accessible chromatin regions were identified as potential regulatory sites. Interestingly, correlations between gene expression and promoter accessibility across samples were skewed towards both positive and negative values, suggesting distinct regulatory mechanisms of gene expression. Finally, candidate enhancers and repressors were identified by comparing ATAC-seq regions with predicted binding sites of 500+ transcription factors.

W442: Functional Annotations of Animal Genomes (FAANG)
The Functional Annotation of Animal Genomes (FAANG): Establishing Metadata Standards, Validation and the FAANG Data Portal
Peter W Harrison, Jun Fan, Daniel R. Zerbino, Guy Cochrane and Paul Flicek, 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. Through its working groups and collaborations it standardises core assays, experimental protocols and analysis methods to maximise effectiveness and inter-comparability of assay data, supporting 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 animals, specimens, cell cultures and experimental assays. This is achieved through provision of metadata validation tools (http://www.ebi.ac.uk/vg/faang/validate/) to the community that ensure submissions meet the FAANG standards (http://www.ebi.ac.uk/vg/faang/rule_sets/FAANG). FAANG encourages the pre-publication of data in the appropriate archives in line with FAANG's data release policy (http://www.faang.org/data-share-principle). The FAANG data portal (http://data.faang.org/home), created and hosted by the Data Coordination Centre (DCC) at EMBL-EBI, 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. We are continually improving the filtering and search mechanisms to aid researchers in identifying appropriate data for their research. The portal provides direct links for downloading all associated assay data, and has an associated API for programmatic access. Through effective standard driven metadata validation, a powerful search driven data portal and promotion of best practice in metadata implementation, FAANG aims to maximise effectiveness and inter-comparability of assay data, supporting the community to create a rich genome to phenome resource.

W443: Functional Annotations of Animal Genomes (FAANG)
FAANG Bioinformatics and Data Analysis Update
James M. Reecy, Department of Animal Science, Iowa State University, Ames, IA

W444: Functional Annotations of Animal Genomes (FAANG)
Protocol Update on ChipSeq in Cattle, Pig, Chicken and Horse
Richard Crooijmans, Wageningen University & Research, Wageningen, Netherlands

W445: Functional Annotations of Animal Genomes (FAANG)
Identification of Genetic Variation Regulating Gene Expression in Dairy Cattle with RNA Sequence Data
Amanda J. Chamberlain[1], Ben J. Hayes[1,2], Ruidong Xiang[3], Christy J Vander Jagt[1], Coralie M. Reich[1], Iona M MacLeod[1], Claire P. Prowse-Wilkins[1], Brett A. Mason[1], Hans D. Daetwyler[4,5] and Michael E. Goddard[1,3], (1)Agriculture Victoria, Bundoora, Australia, (2)The University of Queensland, Brisbane, Australia, (3)The University of Melbourne, Parkville, Australia, (4)School of Applied Systems Biology, La Trobe University, Bundoora, Australia, (5)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Australia
There is increasing evidence to suggest many mutations affecting complex traits may be regulatory, that is they affect the expression of genes. The identification of regulatory variants could lead to increases in the accuracy of genomic breeding values. With the aim of identifying this type of variant, we performed three different analyses that tested variants for an association with changes in gene expression via three regulatory mechanisms: 1) By changing the total gene expression (expression QTL, eQTL), 2) changing the balance of the two parental alleles expressed (allele specific expression QTL, aseQTL), 3) changing the isoforms that are expressed (splice variant QTL, sQTL). We utilised RNA sequence data from milk and white blood cells collected from 141 lactating cows with imputed whole genome sequence data for all three analyses. Many variants were detected with significant eQTL, aseQTL and sQTL effects, with low false discovery rates. There was significant overlap in genes with significant eQTL, aseQTL and sQTL. eQTL with a large effect in white blood cells were likely to have a large effect, in the same direction, in milk cells as well. sQTL significant in both milk and white blood cells more often caused expression of the same isoform. There was a trend for the most significant variant to be < 100 kb from the transcription start site of the gene they were affecting for all three QTL types. The putative regulatory mutations affecting gene and isoform expression identified here are candidates for mutations affecting complex traits. Future work will refine these variants to those likely to be involved in regulating genes associated with traits important to the dairy industry and will test their impact on genomic prediction accuracy.

W446: Functional Annotations of Animal Genomes (FAANG)
FAANG: An Update on Equine Studies
Carrie J. Finno, University of California - Davis, Davis, CA

W447: Functional Annotations of Animal Genomes (FAANG)
Developing the Functional Annotation of the Sheep Genome
Brenda M. Murdoch[1], Stephen N. White[2], Michelle R. Mousel[2], Alisha T. Massa[3], Kim C. Worley[4], Alan L. Archibald[5], Emily L. Clark[5], Brian Dalrymple[6], James W. Kijas[7], Shannon Clarke[8], Rudiger Brauning[8], Timothy P.L. Smith[9], Tracy Hadfield[10] and Noelle Cockett[10], (1)University of Idaho, Moscow, ID, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA,

(3)Washington State University, Pullman, WA, (4)Human Genome Sequencing Center, Department of Molecular and Human Genetics, Houstin, TX, (5)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (6)Institute of Agriculture, The University of Western Australia, CRAWLEY, Australia, (7)CSIRO Animal, Health and Food Science, St Lucia, Australia, (8)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (9)USDA, ARS, USMARC, Clay Center, NE, (10)Utah State University, Logan, UT
The functional annotation of the sheep genome will facilitate a better understanding of the complex nature of gene regulation within this globally important food and fiber species. Through the generation and compilation of transcription and genomic data, we will characterize and define the multifaceted biological mechanisms that contribute to gene regulation. This research examines the genome for coding and non-coding transcript isoforms and alternative splicing, promoters and cis-acting regulatory elements, open chromatin, histone modifications, and DNA methylation across a wide range of sheep tissues. Members of International Sheep Genomics Consortium (ISGC) and other researchers used protocols compliant with the FAANG Consortium assays to collect approximately 100 tissues from the same Rambouillet female used for *de novo* genome assembly. To fully understand the transcriptome, we performed three types of RNA sequencing including long read PacBio IsoSeq, and short read Illumina mRNA and miRNA sequencing. Furthermore, to complement gene expression data and to identify active promoters and confirm transcription start sites, cap analysis of gene expression (CAGE), is being performed on all tissues. Further, to assess chromatin accessibility, ATAC-seq is being performed for all tissues. Nuclear DNA was collected from 10 tissues at slaughter for library preparation and sequencing. Individual cells were collected from 18 additional tissues, and intact samples were taken from remaining tissues and slow-frozen for ATAC-seq analyses. Overall, this project will provide tissue-specific detailed understanding of gene regulatory signals and gene products that contribute to evolution, development and functional phenotype of sheep.

W448: Functional Annotations of Animal Genomes (FAANG)
An Overview of the Sheep and Water Buffalo Gene Expression Atlas Projects
Emily L. Clark, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom
We have produced a high-resolution gene expression atlas for both sheep and water buffalo to support functional annotation of the genome of these two economically important livestock species. RNA-Seq libraries were generated by Edinburgh Genomics (http://genomics.ed.ac.uk) from tissues and cells representing all major organ systems from adult Texel x Scottish Blackface sheep and multiple juvenile, neonatal and prenatal developmental time points. The dataset includes 367 medium depth and 74 high depth 125bp stranded Illumina RNA-Seq libraries. Similarly, for water buffalo we have generated a fine scale gene expression atlas of 220 tissue and cell types collected from adult riverine water buffalo (Mediterranean, Pandharpuri and Bhadawari breeds). We have quantified gene expression from the RNA-Seq data using two different methods, an alignment free method, Kallisto, and a conventional alignment based HiSat2-Stringtie pipeline. Co-expression patterns across tissues were visualised using the network analysis tool Miru (https://kajeka.com/). Both datasets have been deposited in the Short Read Archive (SRA) and sample metadata is available in BioSamples. These comprehensive atlases of gene expression provide model transcriptomes for ruminants, have the potential to inform future improvements in livestock productivity, efficiency and health and are a valuable resource for the international Functional Annotation of Animal Genomes (FAANG) initiative.

W449: Functional Annotations of Animal Genomes (FAANG)
Update on Cattle FAANG Efforts in Canada
Graham S Plastow, University of Alberta, Edmonton, AB, Canada
The FAANG initiative is an important effort to help refine and extend the application of genomics in cattle in Canada. Priorities include feed efficiency and health and welfare as well as improving genomic selection across breeds, particularly for beef production. Different phenomics projects are underway to identify animal resources for collection of sets of tissues for FAANG assays as described in the FAANG white paper (Andersson *et al*. 2015, *Genome Biology* 16:57). For example, an Agriculture and Agri-Food Canada (AAFC) team is studying Johne's disease in Holstein with ChiPSeq, metabisulphite sequencing and RNA analyses. Tissues from beef animals divergent for feed efficiency and methane emission will be harvested at different developmental stages. Other studies are underway on beef quality and disease susceptibility. Collaboration with international FAANG partners is a key aspect for success. Inputs provided by Eveline Ibeagha-Awemu (AAFC, Sherbrooke) and Angela Canovas (University of Guelph).

W450: Functional Annotations of Animal Genomes (FAANG)
Integrative Alignments of DNA Elements for Transcriptional Regulation in Swine Epigenome
Jianhua Cao, Huazhong Agricultural University, Wuhan, China

W451: Functional Annotations of Animal Genomes (FAANG)
Functional Annotation of All Salmonid Genomes (FAASG)
Ben F. Koop, University of Victoria, Victoria, BC, Canada

W452: Functional Genomics
The Wild Olive Genome
Oussama Badad, SIUC, Carbondale

W453: Functional Genomics
Soybean Mutations Mapping: Applications in Functional Gene Analysis and Soybean Seed Improvement
Naoufal Lakhssassi, Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL
Soybean [*Glycine max* (L.) Merr.] is the most widely consumed legume crop in the world, providing 56% of the world's oilseed production. Soybean cultivars contain between 3-4% seed stearic acid. Increasing stearic acid confers a higher melting temperature and oxidative stability

necessary for solid fat application. Highly-saturated soybean seed oil would be suitable for this end use. Stearoyl-acyl carrier protein desaturase *SACPD-C* has been reported to control the accumulation of seed stearic acid, however, no study has previously reported its involvement in leaf stearic acid content and impact on leaf structure and morphology. A subset of an EMS mutagenized soybean population was screened to identify mutants within *GmSACPD* genes. Using a forward genetics approach, a nonsense and four missense *Gmsacpd-c* mutants were identified to contain not only high levels of seed, but also increased leaf and nodule stearic acid content. Homology modeling and *in silico* analysis of the GmSACPD-C enzyme reveals that most of these mutations were localized near or at conserved residues essential for di-iron ion coordination. Furthermore, mutations at conserved residues cause the highest stearic acid content and correlate with the presence of cell senescence and a necrotic cavity in the nitrogen fixing nodules. Interestingly, soybean plants with GmSACPD-C mutations in non-conserved residues show an increase in stearic acid content while conserving healthy nodules. Interestingly, mutational analysis uncovers the impact of GmSACPD-C mutations in leaf and nodule structure and morphology. Random mutagenesis coupled with mutational analysis allows for the achievement of high stearic acid content with no associated negative agronomic characteristics.

W454: Functional Genomics
The Phomopsis Genome
Shuxian Li, USDA, Stoneville, MS

W455: Functional Genomics
Transcriptomic Analysis of Fine Mapped Soybean Aphid (*Aphis glycines*) Resistance Gene, *rag1c* in Soybean (*Glycine max* L.)
Jiazheng (John) Yuan, Fayetteville University, Fayetteville, NC

W456: Functional Genomics
Comparative Omics Analysis of Stem Solidness in Wheat and its Synteny with Closely Related Species
Hikmet Budak, Montana State University, Bozeman, MT

W457: Functional Genomics of C_4 and CAM photosynthesis
Engineering C4 Photosynthesis Traits into Rice
Tammy L. Sage, Department of Ecology and Evolutionary Biology University of Toronto, Toronto, ON, Canada

W458: Functional Genomics of C_4 and CAM photosynthesis
Crosstalk between the Circadian Clock and Metabolism during the Daily Cycle of Crassulacean Acid Metabolism
Suzanna Boxall, University of Liverpool, Liverpool, United Kingdom

W459: Functional Genomics of C_4 and CAM photosynthesis
Molecular Evolution of C_4 Photosynthesis Sub-Types in the Paniceae Grasses
Jacob D. Washburn[1], Satya S. Kothapalli[2], Julia M. Brose[2], Sarah Covshoff[3], Julian M Hibberd[4], Gavin C. Conant[5] and J. Chris Pires[6], (1)Cornell University, Ithaca, NY, (2)Division of Biological Sciences/University of Missouri, Columbia, MO, (3)Department of Plant Sciences/University of Cambridge, Cambridge, United Kingdom, (4)University of Cambridge, Cambridge, United Kingdom, (5)Department of Biological Sciences/North Carolina State University, Raleigh, NC, (6)Division of Biological Sciences, University of Missouri, Columbia, MO
In C_4 plants the enzymatic machinery underpinning photosynthesis can vary, with for example, three distinct C_4 acid decarboxylases being used to release CO_2 in the vicinity of RuBisCO. For decades, these decarboxylases have been used to classify C_4 species into three biochemical sub-types. However, more recently the notion that C_4 species mix and match C_4 acid decarboxylases has increased in popularity and, as a consequence, the validity of specific biochemical sub-types has been questioned. Using species from the grass tribe Paniceae we show that whilst transcripts encoding multiple C_4 acid decarboxylases accumulate in bundle sheath cells in some species, in others, transcripts encoding only one enzyme are detected. In addition, a method that allows isolation of bundle sheath cells from a C_3 species within the Paniceae, *Sacciolepis indica*, was developed. Deep sequencing of bundle sheath preparations from four species, combined with ancestral state reconstruction, support the notion that the three biochemical C_4 sub-types found in the Paniceae existed together in their most recent common ancestor. Thus, these species likely inherited the functional building blocks of all three C_4 pathways. We conclude that classification of C_4 plants into the classical biochemical sub-types is still appropriate for some species, and that evolution of this trait has been facilitated by characteristics of the ancestral C_3 bundle sheath and made use of multiple convergent routes involving either one or multiple C_4 acid decarboxylases.

W460: Functional Genomics of C_4 and CAM photosynthesis
Evolution of CAM Photosynthesis in *Tillandsia* spp. (Bromeliaceae): Emerging Evidence from Genomics, Transcriptomics, and Targeted Metabolite Profiling
Marylaure de la Harpe[1], Margot Paris[2], Jaqueline Hess[1], Martha Serrano[3], Michael Barfuss[1], Walter Till[1], Nicolas Salamin[3], Michael Kessler[4], Wolfram Weckwerth[1], Ray Ming[5] and **Christian Lexer**[1], (1)University of Vienna, Vienna, Austria, (2)University of Fribourg, Fribourg, Switzerland, (3)University of Lausanne, Lausanne, Switzerland, (4)University of Zurich, Zürich, Switzerland, (5)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, China
The pineapple family (Bromeliaceae; bromeliads) represents one of the most diverse and enigmatic adaptive radiations of the Neotropics. Diversification in this group was facilitated by several ´key innovations´ or adaptive trait shifts, one of which is the transition from C3 to CAM

photosynthesis. We have used a phylogenomic approach complemented by differential gene expression analysis (RNA-seq) and targeted metabolite profiling to address the patterns and mechanisms of C3 / CAM evolution in the extremely species-rich bromeliad genus *Tillandsia* and related taxa. Results from whole-genome phylogenomics indicate the presence of several independent C3 / CAM transitions in this group, and these patterns are unlikely to be strongly affected by hybridization and gene flow. Temporal (day / night) and interspecific (CAM / C3) patterns of gene expression reflect these phylogenomic relationships and suggest that C3 taxa in this group are pre-adapted for the evolution of CAM. An analysis of Copy Number Variants (CNV´s) revealed extensive duplications and gene family expansion including many genes involved in CAM and its regulation in subgenus *Tillandsia*, a species-rich group that radiated upon colonization of ecologically diverse and extreme niches in Central- and North America. Distributions of mutations in protein-coding regions point to genes affected by positive selection in branches leading to species with CAM metabolism. We will discuss genomic and transcriptomic patterns and potential mechanisms of CAM evolution in this ecologically important and species-rich group of Bromeliaceae. To make this possible, we will put the results into the context of targeted metabolite profiles and known physiological phenotypes of these species, and we will compare the results to recently published findings on the likely mechanisms of CAM in the distantly related bromeliad genus *Ananas*, the cultivated pineapple.

W461: Functional Genomics of C_4 and CAM photosynthesis
Characterization of C4 Photosynthesis Traits in Setaria
Patrick Ellsworth, Washington State University, Pullman, WA

W462: Functional Genomics of C_4 and CAM photosynthesis
Temporal and Spatial Transcriptome of CAM Photosynthesis in Pineapple
Jennifer Wai, University of Illinois - Champaign-Urbana, Urbana, IL

W463: Fungal Genomics
Draft Genome Sequence of *Tilletia indica*, Karnal Bunt Pathogen of Wheat
Soma Marla, ICAR.NBPGR, New Delhi, India

W464: Fungal Genomics
A Scalable Platform to Identify Fungal Secondary Metabolites and their Gene Clusters
Kenneth Clevenger[1], Jin Woo Bok[2], Rosa Ye[3], Galen Miley[1], Maria Verda[1], Thomas Velk[2], Cynthia Chen[4], Kahoua Yang[2], Matt Robey[1], Peng Gao[1], Matt Lamprecht[4], Paul Thomas[1], Md Nurul Islam[4], Jon Palmer[2], Nancy P Keller[2], Neil L Kelleher[1] and **Cheng-Cang Charles Wu**[4], (1)Northwestern University, Evanston, IL, (2)University of Wisconsin at Madison, Madison, WI, (3)Intact Genomics, Inc., St. Louis, MO, (4)Intact Genomics, Inc., St Louis, MO
The genomes of filamentous fungi contain up to 90 biosynthetic gene clusters (BGCs) encoding diverse secondary metabolites—an enormous reservoir of untapped chemical potential. However, the recalcitrant genetics, cryptic expression, and unculturability of these fungi prevent scientists from systematically exploiting these gene clusters and harvesting their products. As heterologous expression of fungal BGCs is largely limited to the expression of single or partial clusters, we established a scalable process for the expression of large numbers of full-length gene clusters, called FAC-MS. Using fungal artificial chromosomes (FACs) and metabolomic scoring (MS), we screened 56 secondary metabolite BGCs from diverse fungal species for expression in *Aspergillus nidulans*. We discovered 15 new metabolites and assigned them with confidence to their BGCs. Using the FAC-MS platform, we extensively characterized a new macrolactone, valactamide A, and its hybrid nonribosomal peptide synthetase–polyketide synthase (NRPS–PKS). The ability to regularize access to fungal secondary metabolites at an unprecedented scale stands to revitalize drug discovery platforms with renewable sources of natural products.

W465: Fungal Genomics
Nuclear and Mitochondrial Genomes Sequenced by Genome Skimming and Reference-Based RadSeq Resolve Robust Relationships among Closely Related Lichen-Forming Fungi
Felix Grewe[1], Steven Leavitt[2], Jen Pan Huang[1], Todd Widhelm[1] and H. Thorsten Lumbsch[1], (1)The Field Museum of Natural History, Chicago, IL, (2)Brigham Young University, Provo, UT
Despite increasing availability of phylogenomic datasets, strategies to generate genome-scale data from organisms involved in symbiotic relationships remains challenging. Using the reduced genome representation approaches of genome skimming and reference-based RADseq, we recently circumscribed seven new species within the nominal species *Rhizoplaca melanophthalma* (black rock-posy lichen). In addition, we used the data to assemble 31 whole mitochondrial genomes of diverse *Rhizoplaca* species. The fungal mitochondrial genomes feature a consistent gene content and syntenic gene order, in addition to a unidirectional transcription of genes. The genome sizes are different largely due to a varying intron number that is consistent across *Rhizoplaca* species. A most parsimonious reconstruction of gain and loss events of these introns agrees with a maximum likelihood phylogeny based on the mitochondrial genes. However, the mitochondrial phylogeny and intron pattern differs from the nuclear phylogeny in two aspects: (1) *R. melanophthalma* sensu stricto can be further divided into two distinct species, one positioned as sister of all other *Rhizoplaca* species, and (2) *R. haydenii* and *R. parilis* form one group but are sister taxa in the nuclear phylogeny. We will examine further if these inconsistencies are based on incomplete lineage sorting and/or mitochondrial introgression. These results indicate that mitochondrial genome information, including the utilization of mitochondrial intron pattern variation as a barcoding marker, will offer more insights into long-standing questions about the evolution of closely-related obligate symbiotic organisms.

W466: Fungal Genomics
Dothideomycetes Comparative Genomics
Sajeet Haridas, US Department of Energy Joint Genome Institute, Walnut Creek, CA

W467: Fungal Genomics
Comparative Genomics of Industrial Fungi
Scott E. Baker, Environmental Molecular Sciences Laboratory/Pacific Northwest National Laboratory, Richland, WA

W468: Fungal Genomics
A Near-Finished Reference Genome for the Wheat Blast Fungus Provides Insight on Pathogenicity and Adaptation
Zhao Peng[1,2], Ely Garcia-Oliveira[1], Guifang Lin[1], David Cook[1], Mark Farman[3], Melinda Dalby[1], Ying Hu[1], Frank F. White[2], Barbara Valent[1] and **Sanzhen Liu**[1], (1)Kansas State University, Manhattan, KS, (2)University of Florida, Gainesville, FL, (3)University of Kentucky, Lexington, KY
Wheat blast, a devastating new fungal disease caused by the haploid fungus *Magnaporthe oryzae* pathotype *Triticum* (MoT), has been spreading in South America since it was first reported in Brazil in 1985. This disease jumped to Bangladesh in 2016 and now poses a major threat to wheat production in South Asia and beyond. To produce a MoT reference genome, the highly aggressive MoT isolate B71, collected in Bolivia in 2012, was sequenced and assembled with long PacBio reads, Illumina sequences, and a novel scaffold technology using Long-Insert-End-Pair (LIEP) sequences, resulting in a near telomere-to-telomere assembly. The availability of the near-finished assembly facilitated the dissection of genomic structural variation (SV) between B71 and a less aggressive MoT strain isolated in Brazil in 1988, as well as among multiple *M. oryzae* isolates adapted for infecting wheat or other crop species. A dispensable mini-chromosome in the B71 genome was identified through the SV analysis and verified, while none or two mini-chromosomes were found in two other MoT isolates. Two effector genes first characterized in the rice blast pathogen, *M. oryzae* pathotype *Oryza*, occur together in the MoT mini-chromosome, where both genes maintain characteristic *in planta* specific expression. The results imply the potential role of the mini-chromosome in the pathogenicity and adaptation of MoT.

W469: Galaxy: An Open Platform for Data Analysis and Integration
Galaxy: Introduction and Ecosystem
Dave Clements, Johns Hopkins University, Eugene, OR and Galaxy Project
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 a grant-funded open-source project that is deployed in hundreds, if not thousands, of organizations around the world. 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.
This session complements the *Galaxy Community Update* talk in the GMOD Session on Wednesday. That talk will focus on recent developments and future plans for the project.

W470: Galaxy: An Open Platform for Data Analysis and Integration
G-OnRamp: Visualizing and Annotating Eukaryotic Genomes with Galaxy
Jeremy Goecks[1], Luke Sargent[1], Yating Liu[2], Wilson Leung[2] and Sarah C R Elgin[3], (1)Oregon Health & Science University, Portland, OR, (2)Washington University in St. Louis, St. Louis, MO, (3)Washington University in St Louis, St. Louis, MO
Eukaryotic genomes are being sequenced at a rapid rate, but many bioinformatics analyses require high-quality gene annotations (*e.g.*, differential gene expression from RNA-Seq data, enriched gene ontology terms, etc.). G-OnRamp (http://gonramp.org) — a collaboration between Galaxy (https://galaxyproject.org) and the Genomics Education Partnership (GEP; http://gep.wustl.edu) — provides a web-based platform for biologists to generate a browser incorporating large genomic datasets for the interactive annotation of eukaryotic genomes. G-OnRamp is a fully-provisioned Galaxy server for data analysis connected with UCSC Assembly Hubs and GMOD's Apollo/JBrowse for visualization and genome annotation, using CyVerse for data management and access. G-OnRamp includes tools and workflows to create evidence tracks (sequence similarity, *ab initio* gene predictions, RNA-Seq, and repeats), which are then incorporated into UCSC Assembly Hubs and Apollo/JBrowse. The integration with CyVerse enables users to store generated datasets in the CyVerse Data Store for cost-effective long-term access and visualization. For researchers, G-OnRamp provides a stable, user-friendly platform for collaborative genome annotations. For educators, G-OnRamp provides a platform for introducing genomics/bioinformatics research into the undergraduate curriculum. GEP is a consortium of over 100 colleges and universities that provide classroom undergraduate research experiences (CUREs) in genomics/bioinformatics for students at all levels. GEP faculty have used G-OnRamp to construct genome browsers for four parasitoid wasp species, using these genome browsers in a CURE during Fall 2017. G-OnRamp training workshops will be held June 12-15 and July 16-19, 2018 at Washington University in St. Louis. Travel support is available for selected participants; sign up at http://gonramp.org/signup.
Supported by NIH 1R25GM119157.

W471: Galaxy: An Open Platform for Data Analysis and Integration
South Green, a Hitchhiker's Guide to Galaxy
Marilyne Summo, CIRAD, Montpellier, France
South Green is a bioinformatics platform applied to the genomic resource analysis of southern and Mediterranean plants.
The South Green web portal (http://www.southgreen.fr/) provides access to a large panel of bioinformatics resources including its own Galaxy instance which supports a large community of users in Montpellier, France and beyond. This talk discusses our experience deploying, customizing, and maintaining a Galaxy server within South Green.

In addition to the generic tools provided with the standard installation of Galaxy, the South Green Galaxy instance (http://galaxy.southgreen.fr/galaxy/) contains a large collection of exclusive tools, Galaxy wrappers and workflows designed for analyses applied to plant genomes.
It actually comprises more than 100 Galaxy wrappers, 9 pre-configured workflows designed for recurrent analyses such as NGS mapping/cleaning, RNAseq, SNP calling and filtering, Genome-Wide Association Studies, basic population genetics, structural variations, metagenomics and phylogenetics. We also developed innovative solution to graphically display outputs of each workflows.
Home-made Galaxy wrappers have been deposited in our local/central toolshed (http://galaxy.southgreen.fr/toolshed/) or in github (https://github.com/SouthGreenPlatform/galaxy-wrappers). Galaxy is extensively used to conduct capacity building activities. It is currently connected to HPC but we are also initiating use of Docker to disseminate some workflows in the IFB (Institut Français de Bioinformatique) cloud, thus facilitating training activities worldwide.

W472: Galaxy: An Open Platform for Data Analysis and Integration
Building a Bridge from the Tripal Community Database to Galaxy
Margaret Staton[1], Ming Chen[2], Connor Wytko[3], Brian Soto[3] and Stephen P. Ficklin[3], (1)University of Tennessee, Knoxville, Knoxville, TN, (2)University of Tennessee, Knoxville, TN, (3)Washington State University, Pullman, WA
Community genome websites and analysis engines have longed used many different software tools in concert to provide the best experience for users. However, seamless integration requires (or could be much improved) by the developers of tools working together to build bridges for tools to communicate, interact and exchange data. We have built a new way to integrate two popular and widely adopted GMOD tools: Tripal, a content management system for building community genome websites, and Galaxy, a web-based workflow engine for biological data analysis. The new Tripal-to-Galaxy bridge allows users of community databases to not only discover, visualize and download genomic information but to directly port it to Galaxy for analysis. This Tripal module, “Tripal Galaxy,” provides a convenient administrative back end to link to a Galaxy instance via an API key, to import workflows from Galaxy to Tripal, and to manage user data between the two. Users are able to select pre-designed workflows for common analyses, upload their own data, and utilize site data as input to the workflows. Further, we identified a need for users to be able to explore their workflow results visually through the browser. We have developed a new R Markdown tool wrapper framework that enables the execution of a software tool by Galaxy followed by the construction of HTML reports summarizing and visualizing the tool output. This HTML can be rendered for users to explore within Galaxy or it can be easily passed back and presented to users within Tripal.

W473: Gene Expression Analysis
Workshop Introduction
Gregory D. May, DuPont Pioneer, Johnston, IA

W474: Gene Expression Analysis
Host Induced Gene Silencing as an Effective Engineering Tool to Suppress Aflatoxin in Maize
Monica Schmidt, University of Arizona, Tucson, AZ
Aflatoxins are a potent carcinogenic class of secondary compounds produced by certain species of *Aspergillus* fungus. Despite decades of implementing control mechanisms, aflatoxin remain a chief agricultural problem. Reducing these losses would have a significant impact on both US economy and food security/safety. Host-inducing gene silencing (HIGS) is a biotechnology mechanism that involves a host cell expressing small interfering RNA molecules, which then migrate into a contaminating pathogenic cell and cause targeted suppression. We have successfully used HIGS to effectively silence the aflatoxin biosynthetic pathway to prevent *Aspergillus* fungus from producing aflatoxin on developing contaminated maize kernels. We targeted three sections of a key *Aspergillus* toxin biosynthetic gene to suppress toxin production in the edible portion of maize during kernel development. A known toxin-producing *Aspergillus* strain was used to infect expressing homozygous transgenic developing maize kernels. Kernels surrounding the infection site were assayed for toxin and levels over 1,000 ppb were detected in nontransgenic segregating control maize kernels and no toxin was detected in the HIGS transgenic maize kernels. Analysis of total RNA isolated from developing kernels did not reveal a single significantly differentially expressed transcript in the transgenic kernels when compared to their nontransgenic segregating counterparts. This research was the first to demonstrate that HIGS can be used in crops to suppress a fungal produced toxin without any unintended off targets at the transcription level. This technology holds promise to enhance both food security and safety in other crops.

W475: Gene Expression Analysis
Large Scale Gene Losses Underlie the Genome Evolution of *Cuscuta australis*
Guiling Sun[1], Yuxing Xu[2], Hui Liu[2], Ting Sun[3] and Jianqiang Wu[2], (1)State Key Laboratory of Cotton Biology, Henan University, Kaifeng, China, (2)Department of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, (3)Institute of Plant Stress Biology, State Key Laboratory of Cotton Biology, Henan University, Kaifeng, China
Dodders (*Cuscuta* spp., Convolvulaceae) are globally distributed root and leafless parasitic plants that parasitize a wide range of hosts. The physiology, ecology, and evolution of these holoparasites are still poorly understood. A high-quality reference genome (size 264.83 Mb and contig N50 of 3.63 Mb) of *Cuscuta australis* was assembled. Our analysis revealed that *Cuscuta* experienced accelerated evolution, and *Cuscuta* and the convolvulaceous morning glory (*Ipomoea*) shared a common whole-genome triplication event before their divergence. Importantly, *C. australis* genome harbors only 19805 protein-coding genes, and 15.4% of the conserved orthologs in autotrophic plants are lost in *C. australis*, many of which are involved in maintaining autotrophic lifestyle and resistance to stress factors, indicating that gene loss underlies the regressive evolution of *Cuscuta*. The *C. australis* genome provides important resources for studying the evolution of parasitism, regressive evolution, and evo-devo in plant parasites.

W476: Gene Expression Analysis

Single Cell Transcriptomics: Perspectives from Maize

Jack Satterlee, Plant Biology, School of Integrative Plant Science, Cornell University, Ithaca, NY and Michael Scanlon, Cornell University, Ithaca, NY

Single cell RNA-Seq (scRNA-Seq) has emerged as a technology to facilitate the high-throughput transcriptomic analysis of individual cells. Numerous scRNA-Seq studies are reported in animals; however, the technology has seen limited use in plants. Here I present preliminary single cell transcriptomic data generated from maize embryos using the 10X Genomics Chromium™ platform. The maize embryo was selected as a study system owing to its diversity of cell and tissue types, including the embryonic shoot and root apical meristems (SAM and RAM, respectively). We aim to use this technology in combination with existing *in situ* hybridization and RNA-Seq data to examine the contributions of cell-type heterogeneity, signaling, and differentiation programs to SAM patterning and development, and to resolve single-cell gene co-expression networks.

Analysis of the dataset identifies embryo-specific marker genes, including previously-described lowly-expressed genes and transcripts accumulating in small cell populations. A preponderance of cells expressing epidermal cell marker genes were identified, suggesting biases in cell isolation. Dimensionality reduction resolves distinct cell clusters, but differential gene expression analysis indicates that cell variation in highly expressed housekeeping genes may explain some of these patterns. Nonetheless, marker gene expression analysis and pseudotemporal ordering of cells along differentiation trajectories reveals potentially pertinent developmental biology.

Future work will attempt to circumvent the unique challenges of unbiased cell-type isolation in plants by using nuclei as a source of RNA for transcriptomic profiling, and utilizing higher cell/nuclei populations to enhance statistical power.

W477: Gene Expression Analysis

Single Cell Methylomes Reveal Neuronal Populations and Regulatory Elements in the Mammalian Brain

Chongyuan Luo, Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA

Human and mouse brain cell types show pervasive CG and non-CG methylation variation throughout the genome, suggesting the possibility of unbiased classification of neurons types using sparse single cell DNA methylation profiles. We developed a new method for high-throughput single nucleus methylcytosine sequencing (snmC-seq) that allows multiplexing and shows greater reads mapping rate than existing methods. Using >6,000 single cell methylomes, we identified 16 mouse and 21 human neuron types in the frontal cortex. The results suggest expanded neuronal diversity in the human cortex, which is consistent with the finding of a human-specific inhibitory neuron subpopulation. Our epigenomic analysis allowed the prediction of approximately ~500,000 cell type specific regulatory elements for mouse or human neuron types.

We are generating single cell methylomes for cell type classification of the whole mouse brain. Over 4,000 single cell methylomes are being produced from mouse motor cortex and are being used to compare cell type composition and the usage of regulatory elements between two cortical regions (frontal and motor). In addition, we are studying the contribution of genetic variants to cell-type specific gene regulation using single-cell methylomes generated from multiple human individuals and brain regions. The preliminary results suggest extensive cell type differences between human cortical regions. Our single nucleus methylome approach can be applied to all human tissues for producing epigenomic profiles to inform the human cell atlas.

W478: Gene Expression Analysis

The Contribution of *cis*- and *trans*-Acting Variants to Gene Regulation in Wild and Domesticated Barley

Matthew Haas, Leibniz Institute of Plant Genetics and Crop Plant Research, OT Gatersleben, Germany

Crop domestication and subsequent artificial selection have resulted in plants with reduced genetic diversity that differ from their wild ancestors in key traits including plant architecture and growth habit. These changes, collectively termed the 'domestication syndrome', are often the result of altered gene regulation rather than altered coding sequence variation. Domestication syndrome has negative consequences for crop improvement, including limiting the ability to adapt crops to climate change. An improved understanding of the regulation of gene expression in barley (*Hordeum vulgare* L.) could accelerate the use of wild genetic resources in barley breeding. Regulation of gene expression is challenging to measure, but can be accomplished by matching transcripts to their specific allele of origin in F_1 hybrids (allele-specific expression, ASE). To assess ASE in barley we utilized nine diverse accessions for which we possess genomic data and F_1 hybrids between these accessions and cv. Morex. These include three cultivars, two landraces and four wild barleys. These accessions represent the spectrum of improvement status, row type and growth habit of barley. Accessions were sown in two trays and placed in a growth chamber (one each for cold stress and control treatments) using a randomized design. When the first leaf of each accession was fully extended, one tray was moved to a cold chamber before harvesting tissue from both conditions. The experiment was repeated four times. RNA extracted from each sample was sequenced on an Illumina machine. Sequence reads were then aligned to the barley reference genome sequence and analyzed for asymmetric allele expression and differential expression between cold and control treatments and the relative contribution of *cis*- and *trans*-acting effects.

W479: Gene Expression Analysis

Low Number of Fixed Somatic Mutations in a Long-Lived Oak Tree

Emanuel Schmid-Siegert[1], Christian Iseli[2], Namrata Sarkar[3], Sandra Calderon[4], Caroline Darimont-Nicolau[3], Jacqueline Chrast[3], Pietro Cattaneo[3], Frederic Schutz[5], Laurent Farinelli[6], Marco Pagni[2], Michel Schneider[7], Jeremie Voumard[3], Michel Jaboyedoff[3], Christian Fankhauser[3], Christian S. Hardtke[3], Laurent Keller[3], John Pannell[3], Alexandre Reymond[3], Mark R. Rechavi[3], Ioannis Xenarios[2] and Philippe Reymond[3], (1)Vital-IT Competence Center, Swiss Institute of Bioinformatics, Lausanne, Switzerland, (2)VITAL-IT, Lausanne, Switzerland, (3)University of Lausanne, Lausanne, Switzerland, (4)GTF Lausanne, Lausanne, Switzerland, (5)Swiss Institute of Bioinformatics, Lausann, Switzerland, (6)Fasteris SA, Plan-les-Ouates, Switzerland, (7)UniProt, Geneva, Switzerland

Because plants do not possess a defined germline, deleterious somatic mutations can be passed to gametes and a large number of cell divisions separating zygote from gamete formation in long-lived plants may lead to many mutations. We sequenced the genome of a 234-year-old oak

tree to study the number and nature of fixed somatic single-nucleotide variants (SNVs). The highly repetitive 720 Mb genome of *Quercus robur* was sequenced and assembled with a combination of Illumina short-reads and PacBio long-reads to generate a reference genome. A special focus was set on accurate prediction of coding genes, in order to evaluate the potential effect of SNVs. We then compared the genome from leaves sampled on terminal ramets of one lower and one upper branch of the tree. Unexpectedly, only few fixed somatic SNVs were found, whose sequential appearance in the tree could be traced along nested sectors of younger branches. Our data suggest that stem cells of shoot meristems are robustly protected from accumulation of mutations in trees.

W480: Gene Introgression
The Patchwork Genome: Assembling Broad Spectrum Rust Resistance in Wheat
Mary J. Guttieri, USDA Agricultural Research Service, Manhattan, KS

W481: Gene Introgression
Mobilizing useful Genes from Wheat Crop Wild Relatives
Ahmed Amri, ICARDA, Rabat, Morocco

W482: Gene Introgression
Developing Broad-Range Disease Resistance Traits for Greenhouse Vegetables
Michael Pautler, Vineland Research and Innovation Centre, Vineland Station, ON, Canada

W483: Gene Introgression
Introgression of a Newly Discovered Xanthophyll Acyl Transferase into Durum Wheat
Diane Mather, The University of Adelaide, Adelaide, Australia
The carotenoid lutein is the main yellow pigment in the grain of both bread wheat (AABBDD) and durum wheat (AABB). Lutein confers health benefits and is important in determining the colour of wheat-based food products. During postharvest storage of bread wheat grain, lutein can be converted into lutein esters, which are more stable than free lutein. This esterification does not occur in durum wheat. Recently, we mapped the esterification trait to a locus on chromosome 7D. We have now confirmed that a gene at that locus encodes a xanthophyll acyl transferase (XAT). This is the first such enzyme to be functionally confirmed in any plant species.
Given the importance of stable yellow pigment for pasta and other durum-based products, we wanted to transfer the *Xat1* gene into durum wheat. We crossed a 7D(7A) disomic chromosome substitution line of Langdon durum with DBA-Aurora, an Australian durum cultivar. This provided opportunities for 7D-7A pairing and recombination in the F_1 generation. Using marker assays designed for 7D-7A SNPs, we screened F_2 progeny for evidence of spontaneous 7D-7A recombination. Plants with recombinant chromosomes were discovered, including one with only a short distal 7D segment that includes the *Xat1* gene. Homozygous recombinant progeny were developed and are being backcrossed to DBA-Aurora to develop a high-performing durum with improved lutein stability. This work demonstrates that with the use of appropriate crosses and markers, plants with intergenomic recombinations can be discovered for use in breeding.

W484: Gene Introgression
Harnessing Robust Resistance of Wild Species, *Arachis cardenasii,* for Improving Resistance to Foliar Fungal Diseases through Genomics in Cultivated Groundnut
Manish K Pandey, Rajeev K Varshney and Pasupuleti Janila, ICRISAT, Patancheru, India
Foliar fungal diseases namely leaf rust and late leaf spot (LLS) cause significant yield loss worldwide and also deteriorate fodder quality in groundnut (*Arachis hypogaea*). As compared to cultivated gene pool, the high level of resistance was detected in wild species and its interspecific derivatives (*A. hypogaea* × *A. cardenasii)*. A resistant variety 'GPBD 4', developed using one of these interspecific derivative namely ICGV 86855 (CS 16), was used to develop recombinant inbred line population (TAG 24 × GPBD 4). The linked SSR markers identified through genetic mapping were deployed to improve resistance of three popular groundnut varieties (ICGV 91114, JL 24 and TAG 24). The introgression lines with stable resistance, added advantage of early maturity, higher pod yield (39-79%) and higher haulm yield (25-89%) over their respective recurrent parents were nominated to the Indian national trial for evaluation and release. Based on the yield trials at 8 locations of India during rainy 2017, three lines (ICGV 14421, ICGV 13189 and ICGV 13207) were promoted to final year of testing and release. The efforts are also in progress for pyramiding foliar disease resistance and high oleic acid using genomics-assisted breeding. The sequencing and high throughput genotyping based trait mapping helped in precise mapping of genomic regions on pseudomolecules A03 and A02 which is now guiding for gene discovery and cloning. Most importantly, these efforts also led to identification of more precise markers amenable to low-throughput (SSRs) and high throughput (SNPs) genotyping including a 10-SNP panel costing just 1.5 USD/sample including DNA isolation.

W485: Gene Introgression
Genetic Regulation of Spikelet Development in Bread Wheat
Laura E Dixon, Nikolai M. Adamski, Clare Lister, Luzie U. Wingen and **Scott Boden**, John Innes Centre, Norwich, United Kingdom
Inflorescence architecture contributes significantly to grain production in cereals, and it has been modified during domestication to increase yields and facilitate harvesting. A unique attribute of grass inflorescences is the arrangement of flowers (or florets) on reproductive branches known as spikelets. In wheat, spikelets are arranged in an alternating phyllotaxy on opposite sides of the central rachis; despite their importance for grain production, little is known about the genes that underpin spikelet development in wheat. We have investigated the genetic regulation of inflorescence architecture by studying a trait known as paired spikelets, which are characterized by the formation of two spikelets at a single node, rather than the typical single spikelet. Here, we have investigated two double haploid populations (Avalon x Cadenza; Buster x Charger)

to identify a region on chromosome 2DS that positively regulates paired spikelet formation, which is a robust QTL that was detected under multiple environmental conditions. Genetic analysis of near-isogenic lines, exome capture analysis and expression data has been used to delimit this region that contains 26 genes. We demonstrate that paired spikelet development is enhanced in NILs for this 2DS locus by short-day photoperiods and increased ambient growth temperatures, suggesting that the genetic regulation of this trait by the 2DS is facilitated by conditions that delay inflorescence development, which is supported by analysis showing that expression of meristem identity gene expression. This data therefore provides evidence for a new genetic region that regulates inflorescence architecture in cereals, which includes functional characterization of alleles that are present in UK breeding programs.

W486: Gene Mapping by Segregation

Gene Mapping By Segregation: From Mendel to BSA, and Beyond

Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

The observations on segregation of pea flowers by George Mendel led to the discoveries of the first two inheritance laws: segregation and independent assortment. If the connectional mapping from phenotypes to genetics is considered as the first-generation of gene mapping, the second generation gene mapping should be the Morgan law that the degree between complete co-segregation and independent assortment is proportional to genetic distance. The Morgan law has been widely used to map major genes through pedigree segregant analysis and Quantitative Trait Loci (QTLs) for complex traits. The major challenge of QTL mapping is the resolution caped the recombination events occurred during recent generation. The reverse strategy was to use the Linkage Disequilibrium (LD) that could be remained after many generation of random matting if the genetic loci are close enough. This third-generation gene mapping, Genome-Wide Association Study (GWAS), plays the major role even today. To exclude the possibility of LD due to non-physically linked reason, homogenous populations, such as F2, were suggested as alternatives. To gain enough meiosis in the homogenous populations, large number of individuals are required. Therefore, the fourth-generation gene mapping pools individuals at with extreme phenotype distribution for deep sequencing to reveal the allele frequencies. This Bulked Segregant Analysis (BSA) gain both on false positive control and mapping resolution at the cost of making crosses. Based on reviewing the first four generation gene mapping methods, the prospective properties are discussed for the next generation gene mapping.

W487: Gene Mapping by Segregation

Distinct Genetic Architectures for Phenotype Means and Plasticities in *Zea mays*

Aaron M Kusmec, Srikant Srinivasan, Dan Nettleton and **Patrick S. Schnable**, Iowa State University, Ames, IA

Phenotypic plasticity describes the phenotypic variation of a trait when a genotype is exposed to different environments. Understanding the genetic control of phenotypic plasticity in crops such as maize is of paramount importance for maintaining and increasing yields in a world experiencing climate change. Here, we report the results of genome-wide association analyses of multiple phenotypes and two measures of phenotypic plasticity in a maize nested association mapping (US-NAM) population grown in multiple environments and genotyped with ~2.5 million single-nucleotide polymorphisms. We show that across all traits the candidate genes for mean phenotype values and plasticity measures form structurally and functionally distinct groups. Such independent genetic control suggests that breeders will be able to select semi-independently for mean phenotype values and plasticity, thereby generating varieties with both high mean phenotype values and levels of plasticity that are appropriate for the target performance environments.

W488: Gene Mapping by Segregation

Discovery, Validation and Deployment of Major QTL for Resistance to Maize Lethal Necrosis in Tropical Maize

Michael Olsen[1], Yoseph Beyene Aydagn[1], Prasanna Boddupalli[1] and Dan Makumbi, Biswanath Das, MacDonald Jumbo, Kassa Semagn, Stephen Mugo, Manje Gowda, , (1)CIMMYT, NAIROBI, Kenya

Maize lethal necrosis (MLN) caused by synergistic interaction between *Maize Chlorotic mottle virus* and S*ugarcane mosaic virus,* is seriously threatening maize production in east Africa. We applied GWAS, bi-parental mapping, and selective genotyping approaches to identify QTL for deployment to accelerate breeding for improved MLN tolerance. Major QTL regions on chromosome 3, 6, and 9, as well as several minor QTL on other chromosomes have been identified. Results have been validated in four independent bi-parental populations through QTL mapping and joint-linkage association mapping. Validation results confirm major effect QTL on chromosome 3 and 6 are consistently found across different genetic backgrounds as well as across environments. Markers linked to MLN tolerance QTL leads were deployed through MABC to improve MLN tolerance of commercially important CIMMYT lines. Efficacy and equivalency results indicate that testcross yield under severe MLN pressure has been improved for several important CIMMYT lines while maintaining comparable yield in the absence of MLN. Further a ten SNP panel including a large effect MLN tolerance haplotype together with a three SNP haplotype for maize streak virus has been successfully used for population enrichment and pre-selection of DH lines in early stages of east African maize breeding programs.

W489: Gene Mapping by Segregation

CandiSNP and CHERIPIC: Tools for Rapid Fine Mapping of Causative Mutations using Draft or Completed Genome Sequences

Ghanasyam Rallapalli, University of East Anglia, Norwich, United Kingdom

Traditional Map Based Cloning approaches, used for the identification of desirable alleles, are extremely labor intensive and years can elapse between the mutagenesis and the detection of the polymorphism. Combining bulk segregant analysis with high throughput sequencing (HTS), referred to as Mapping-By-Sequencing (MBS) has accelerated the identification of causative mutations for genomes with good quality reference sequences. Our group's CandiSNP analysis and visualisation application, generates density plots and table with a list of candidate causative mutations, defined as SNPs causing non-synonymous changes in annotated coding regions using SNPs obtained from bulk segregant sequencing (BSS) data. CandiSNP is a user-friendly application that will aid in novel discoveries from forward-genetic mutant screens. The web-application is freely available online at http://candisnp.tsl.ac.uk/ and source code at http://github.com/TeamMacLean/candisnp/.

Although MBS studies are effective, require an ordered genome assembly and cannot be used with fragmented, un-scaffolded draft genomes, limiting their application to model species and precluding many important organisms. We developed a computational method and software implementations to address this gap in resource. Called CHERIPIC (Computing Homozygosity Enriched Regions In genomes to Prioritise Identification of Candidate variants - http://cheripic.tsl.ac.uk/ and https://github.com/shyamrallapalli/cheripic/), it makes use of fragmented genome assemblies resulting from BSS data to call variants and identifies a causative mutation or a few closely linked variants that help narrow down the region harbouring the trait of interest. CHERIPIC has been applied to assemblies of bulked whole genome sequence data from Arabidopsis, bulked RNA-seq data from maize, bulked exome data of barley and identified variants that are very closely linked to the region of the causative mutation.

W490: Gene Mapping by Segregation

A Whole Genome Association Study in Bailinggu (*Pleurotus tuoliensis*)

Yongping Fu, Engineering Research Center of Chinese Ministry of Education for Edible and Medicinal Fungi/Jilin Agricultural University, Changchun, China

Bailinggu (*Pleurotus tuoliensis*) is a major, commercially cultivated mushroom and widely used for nutritional, medicinal, and industrial applications. Here we report a high-quality genome of domesticated *P. tuoliensis* (Bailinggu) obtained by PacBio SMRT sequencing. The assembled genome size was 45 M, and 95% of the genome assembly was contained in 24 contigs (N50 = 2.8 M). A phylogenetic tree was constructed including *P. tuoliensis* and ten other fungal species. The estimated divergence times between *P. tuoliensis* and *P. eryngii*, *P. tuoliensis* and *P. ostreatus* are 6.6 and 20.1 million years ago (Mya), respectively. We then investigated genomic variation between domesticated and wild Bailinggu based on the genome-wide resequencing of 54 strains. Evidence for the molecular footprints of artificial selection in Bailinggu was found, and it revealed signals of selection in 170 genes of domesticated strains, several of which relate to cellulose catabolic process, nitrogen compound metabolic process and environmental adaptation. In addition, the integrative genomic and transcriptomic analyses of domesticated Bailinggu were used to identify genes crucial to fruiting body formation and response to cold stimulation. Our results provide insights into the genetic basis of Bailinggu diversity and domestication. These findings will improve future research on Bailinggu breeding and other closely related species.

W491: Gene Mapping by Segregation

Mapping QTLs for Kernel Row Number and *Fasciated* Ear by SNP-Based Bulk Segregant Analysis in Maize

Silvia Giuliani, Roberto Tuberosa and **Silvio Salvi**, DipSA - University of Bologna, Bologna, Italy

In maize, the number of kernel rows (KRN) of the ear is one of the most important grain yield components. Both QTLs and Mendelian mutations (such as abnormally shaped - *Fasciated* - ear mutants) have been discovered for this trait and utilized to gain information on the molecular genetic control of ear development. In this study, two Italian maize inbred lines were identified to show extreme phenotypes in terms of ear fasciation and low KRN, respectively and utilized to develop three recombinant inbred line (RIL) populations. Two of the populations (A and B) had the fasciated ear type inbred line as parent, while the third population (C) was generated by crossing the elite line B73 with the low KRN line. The three populations were thoroughly phenotyped for ear morphology and KRN in F5 and F6 generations and showed an overall continuous type of variation for ear traits. We next attempted to map QTLs for fasciated ear and KRN using bulk segregant analysis (BSA) based on a high-density maize SNP array (15k Illumina Infinium) in two successive years. Bulks included 15 plants (extremely fasciated ear plants or wild-type ear plants for populations A and B, and plants with highest or lowest KRN for population C). Preliminary results showed the presence of major QTLs segregating and affecting both ear fasciation and KRN.

W492: Gene Mapping by Segregation

Bulked Sample Analysis in Genetics, Genomics and Crop Improvement

Cheng Zou[1], Fengchao Cui[2], Dan Jeffers[3], Pingxi Wang[2], Huihui LI[4] and Yunbi Xu[5], (1)Chinese Cademy of Agricultural Sciences, Beijing, China, (2)Chinese Academy of Agricultural Sciences., Beijing, China, (3)CIMMYT International Maize and Wheat Improvement Center, Edo. de Méxic, Mexico, (4)Chinese Academy of Agricultral Sciences, Beijing, China, (5)CIMMYT/CAAS, Beijing, China

Biological assay has been based on analysis of all individuals collected from sample populations. Bulked sample analysis (BSA), which works with selected and pooled individuals, has been extensively used in gene mapping through bulked segregant analysis with biparental populations, mapping by sequencing with major gene mutants and pooled genomewide association study using extreme variants. Compared to conventional entire population analysis, BSA significantly reduces the scale and cost by simplifying the procedure. The bulks can be built by selection of extremes or representative samples from any populations and all types of segregants and variants that represent wide ranges of phenotypic variation for the target trait. Methods and procedures for sampling, bulking and multiplexing are described. The samples can be analysed using individual markers, microarrays and high-throughput sequencing at all levels of DNA, RNA and protein. The power of BSA is affected by population size, selection of extreme individuals, sequencing strategies, genetic architecture of the trait and marker density. Here, we also present a new optimized statistical method Ext-BSA that take advantage of high density genetic markers generated by the next generation sequencing and conducted inclusive statistical testing to find candidate loci that significantly different in two bulks. BSA will facilitate plant breeding through development of diagnostic and constitutive markers, agronomic genomics, marker-assisted selection and selective phenotyping.

W493: Gene Mapping by Segregation

Harnessing the Power of GWAS to Pool-Sequencing for Real-Time Studies in Natural Populations

Alexandre Fournier-Level, The University of Melbourne, Parkville, Australia

W494: Gene Mapping by Segregation

iPat: Intelligent Prediction and Association Tool for Genomic Research

Chun-Peng Chen, Washington State University, Pullman, WA
The ultimate goal of genomic research is to effectively predict phenotypes from genotypes so that medical management can improve human health and molecular breeding can increase agricultural production. Genomic prediction or selection (GS) plays a complementary role to genome-wide association studies (GWAS), which is the primary method to identify genes underlying phenotypes. Unfortunately, most computing tools cannot perform data analyses for both GWAS and GS. Furthermore, the majority of these tools are executed through a command-line interface (CLI), which requires programming skills. Non-programmers struggle to use them efficiently because of the steep learning curves and zero tolerance for data formats and mistakes when inputting keywords and parameters. To address these problems, this study developed a software package, named the Intelligent Prediction and Association Tool (iPat), with a user-friendly graphical user interface (GUI). With iPat, GWAS or GS can be performed using a pointing device to simply drag and/or click on graphical elements to specify input data files, choose input parameters, and select analytical models. Models available to users include those implemented in third party CLI packages such as GAPIT, PLINK, FarmCPU, BLINK, rrBLUP, and BGLR. Users can choose any data format and conduct analyses with any of these packages. File conversions are automatically conducted for specified input data and selected packages. A GWAS-assisted genomic prediction method was implemented to perform genomic prediction using any GWAS method such as FarmCPU. iPat was written in Java for adaptation to multiple operating systems including Windows, Mac, and Linux.

W495: Genome annotation resources at the EBI
Ensembl and Ensemblgenomes

Benjamin Moore, EMBL-EBI, Hinxton, United Kingdom
'Browsing Genes and Genomes with Ensembl and Ensembl Genomes' 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 VEP.
Ensembl (www.ensembl.org) provides an interface and an infrastructure for accessing genomic information covering over 100 vertebrate species, including cow, pig, sheep, and chicken. Its sister project, Ensembl Genomes (www.ensemblgenomes.org), consists of five sub-portals (bacteria, protists, fungi, plants, and invertebrate metazoa) which contain data for over 700 eukaryotic (including wheat, barley, tomato and brassicas) 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 (for protists, fungi, plants, and animals), FTP, Perl APIs, REST API, and MySQL. Furthermore, the Variant Effect Predictor (VEP) is a powerful tool for analysing sets of genomic variants, available for all species in Ensembl and Ensembl Genomes.
Highlights of the past year include;

- Over 30 new and updated rodent and primate genomes now available, as well as new assemblies and genebuilds for pig and cat.
- New plant species: jute, cassava, yam, sunflower, cotton, bean and cucumber. New assembly for barley, and updated assemblies and annotation for sorghum bicolor, soybean, peach, rice and maize.
-

W496: Genome annotation resources at the EBI
Manual Genome Annotation in Ensembl

Jane Loveland and Adam Frankish, EMBL-EBI, Hinxton, United Kingdom
The Human and Vertebrate Analysis and Annotation (HAVANA) team have recently moved to the European Bioinformatics Institute (EBI) as 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. We have also annotated across whole genomes for zebrafish, pig and rat and specific regions of interest, such as the Major Histocompatibility Complex (MHC), for selected organisms. Our manual annotation is merged with the automated Ensembl annotation to produce the GENCODE gene set for human and mouse. The annotation can be downloaded from gencodegenes.org and is available from the Ensembl and UCSC genome browsers.
We are working to integrate our annotation software (Zmap/Otter) further within the Ensembl pipeline to streamline the merge process and enable faster annotation updates.
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:
ftp://ngs.sanger.ac.uk/production/gencode/update_trackhub/hub.txt

W497: Genome annotation resources at the EBI
European Nucleotide Archive (ENA): Data Coordination, Archiving and Retrieval

Peter W Harrison, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom
The European Nucleotide Archive (ENA) is a globally comprehensive data resource for nucleotide sequence, spanning raw data, alignments and assemblies, functional and taxonomic annotation and rich contextual data relating to sequenced samples and experimental design, and the European node of the International Nucleotide Sequence Database Collaboration (www.insdc.org). The ENA provides a portfolio of services for submission, data management, search and retrieval of sequencing data, for the plant and animal communities. The ENA also plays a central role in livestock initiatives such as providing the Data Coordination Centre for the Functional Annotation of Animal Genomes (FAANG) project. In this capacity, the ENA provides bespoke submission tools to support provision of high quality and rich supporting metadata to maximise effectiveness and inter-comparability of assay data, supporting the community to create a rich genome to phenome resource. We are striving to simplify our submission services whilst maintaining a high level of data integrity, discoverability and reusability. We have recently released a suite of tutorials that cover both interactive and programmatic submissions. I will provide an overview of our newly implemented data discovery API, that now includes the ability to share prepublication datasets with authenticated access and a suite of improvements such as improved data-sample searches and simplified programmatic interactions. We encourage plant and animal researchers to submit and source experimental data from the ENA, and look to us as a collaborator and data management platform for nascent plant and animal studies.

W498: Genome annotation resources at the EBI

The Transition of Non-Human Genetic Variation Data from dbSNP to EVA

Gary Saunders, EMBL-EBI, Cambridge, United Kingdom

The long-term commitment to provide stable identifiers for genetic variation data is of key importance so that newly discovered variants and alleles can be referenced in publications, cross-linked between databases and integrated with successive reference genome builds. In a new agreement between the European Variation Archive (EVA; www.ebi.ac.uk/eva) and dbSNP, all non-human genetic variation data submissions will be solely handled by EVA, and dbSNP will support only human variation data.

Accordingly, the EVA is now responsible for the generation and maintenance of reliable accessions for non-human genetic variation data; dbSNP is responsible for assigning new stable identifiers to human genetic variation data only.

The EVA follows the same accessioning principles as employed by dbSNP: non-human variants submitted to the EVA are issued 'Submitted SNP' (ss) accessions, and these are periodically clustered to form 'Reference SNP' (rs) accessions. The EVA is fully committed to the continuation of all existing dbSNP ss and rs accessions.

In this presentation I will give an update on the current status of this transition, future plans, working timelines and the subtle differences in the submission of, and access to, genetic variation data at EVA compared with dbSNP.

W499: Genome annotation resources at the EBI

The Vertebrate Gene Nomenclature Committee (VGNC) – Standardizing Gene Names in Vertebrates

Susan Tweedie, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Standardized gene nomenclature provides an essential resource for all researchers. However an ever-increasing number of vertebrate genomes are being sequenced and the data released into the public domain without any systematic annotation or gene naming. There are currently only six vertebrate model organisms with an official gene nomenclature group (mouse, rat, chicken, Anolis, Xenopus and zebrafish), all of which base their gene names on those approved by the HUGO Gene Nomenclature Committee (HGNC) for human genes. This presentation describes the work of the Vertebrate Gene Nomenclature Committee (VGNC) – a parallel branch of the HGNC tasked with approving gene names and symbols across vertebrate species that lack their own nomenclature group. Our strategy combines an automated pipeline for naming high confidence 1:1 human orthologs in each species with careful manual curation for non-consensus orthologs and within complex gene families. Approved nomenclature is now available for chimpanzee, cow, dog and horse, with >10K protein coding genes named in each species to date. All VGNC data are available via the VGNC website (https://vertebrate.genenames.org/) and the symbols and names are displayed on key resources such as Ensembl and NCBI Gene. As well as continuing to name the remaining genes in these four species, we will add new species to VGNC based on the quality of genome assembly and annotations, perceived importance as a model for humans and demand from the research community. Please email us at vgnc@genenames.org if you have expertise in a particular species or gene family that you could help us to name.

W500: Genome annotation resources at the EBI

Expert Curation of Proteins in UniProtKB/Swiss-Prot

Damien Lieberherr[1], Sylvain Poux[1] and UniProt Consortium, (1)SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

The UniProt KnowledgeBase (UniProtKB) provides a single, centralized, freely available resource for protein sequences and functional information. For Arabidopsis, our main targets for expert annotation are proteins with some functional characterization and most of them are now included in UniProtKB/Swiss-Prot. Expert curation 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.

Our annotation program has been actively collaborating with other resources, such as Araport, the Arabidopsis portal. We provide Araport with all the gene model corrections that we introduced on the bases of our trans-species family annotation. We are also completing the knowledgebase by importing missing information from EnsemblPlants.

The UniProt consortium is also actively involved in GO annotation and 19,000 manual annotations has been added to more than 4700 plant proteins. Experimental peptides from high-throughput proteomics experiments that uniquely match the product of a single gene are used to generate annotations describing post-translational modifications and protein processing events. UniProtKB serves as a central hub for biomolecular information with access to more than 100 other resources, such as nucleotide sequence database, 3D protein structure databases, InterPro or MODs.

W501: Genome annotation resources at the EBI

Accessing the Gene Ontology, a Resource for Whole Genome Analysis

Sandra Orchard, EMBL-EBI, Hinxton, United Kingdom

The Gene Ontology (GO) is used to describe gene products across the entire taxonomic range of species, and an association between a GO term and a gene product gives computationally-accessible information on the temporal and spatial expression pattern and localization of the molecule, its physiological functions and the pathways and processes it plays a role in. These data enable researchers to easily explore the composition of a whole genome or to extract biological meaning from large-scale datasets. Experimental data is added to proteins and ncRNA molecules in well studied organisms and this information is propagated to other proteins in the UniProtKB database by sequence homology or by recognising similar proteins using protein signatures collated by the InterPro resourcedatabase. The EMBL-EBI provides the user with access to both the Gene Ontology and Gene Ontology annotation data via QuickGO (https://www.ebi.ac.uk/QuickGO), a fast web-based browser. Gene Association files with consistent mappings of GO terms to UniProt can be downloaded from this site, or accessed from https://www.ebi.ac.uk/GOA.

W502: Genome annotation resources at the EBI

Adding Information to Plant and Animal Proteomes in Uniprot through Automatic Annotation

Kate Warner, EMBL-EBI, Hinxton, United Kingdom

The mission of UniProt is to provide a comprehensive and thoroughly annotated protein resource to the scientific community, most notably through the UniProt Knowledgebase (UniProtKB). Within UniProtKB, the reviewed section (Swiss-Prot) contains high quality, manually curated, richly-annotated protein records. In contrast, the unreviewed section (TrEMBL) which makes up 99% of UniProtKB, depends for its annotation on automatically extracted experimental data from 3D structures, links to other databases and rule-based annotation when there is no experimental data available.

Rules are a formalized way of expressing an association between conditions, which have to be met, and annotations, which are then propagated. UniProtKB employs two approaches to automatically annotate TrEMBL entries in an efficient and scalable manner with a high degree of accuracy:

The Statistical Automatic Annotation System (SAAS) formed of system generated rules, and UniRule which is composed of expertly generated rules.

Currently the UniRule system contains over 5,700 rules, which provide annotation for approximately 30% of unreviewed entries. In UniRule InterPro signatures, predictive models for the functional classification of protein sequences, and taxonomic constraints are the fundamental conditions, but others are used too to provide annotation types such as nomenclature, function, Gene Ontology (GO) terms and sequence features. Rules are evaluated against reviewed entries to ensure accuracy, and are re-evaluated and updated at every release to make sure that the propagated annotation in the unreviewed entries is kept up to date.

All aspects of the UniRule system including how users can view and explore UniRule will be explained at the workshop.

W503: Genome annotation resources at the EBI

ArrayExpress and Expression Atlas: Tools for Archiving, Searching and Visualizing Functional Genomics Data at EMBL-EBI

Suhaib Mohammed, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Expression Atlas at EMBL-EBI (https://www.ebi.ac.uk/gxa) is a database and web-service that provides information about gene and protein expression patterns in plant and animal species across different tissues, developmental stages and diseases. All datasets are curated with a semi-automatic process of identifying the experimental factors, such as diseases or perturbations, annotating metadata with Experimental Factor Ontology terms (EFO) and describing the experimental comparisons for further processing. Analyses of RNA-seq datasets are performed using our standardized pipeline iRAP (https://nunofonseca.github.io/irap) while our microarray pipeline use standard open source tools. Expression Atlas provides baseline and differential studies. Baseline studies report transcript abundance within tissues, developmental stages or cell lines while differential studies report changes in expression across two different conditions, for example, healthy versus disease. Presently, we provide results on more than 3,100 experiments from more than 30 species. Large studies include BluePrint, GTEx, Encode, CCLE, HipSci and Pan-Cancer. A quarter of all studies are plant experiments and 11% are relevant to diseases. All data in Expression Atlas are free to browse, download, and reuse.

ArrayExpress at EMBL-EBI (https://www.ebi.ac.uk/arrayexpress) is the source of most data in Expression Atlas. It is an archive for functional genomics experiments, supporting their reuse by the research community. Experiments in ArrayExpress are directly submitted by scientists through Annotare (https://www.ebi.ac.uk/fg/annotare/login/), our webform-based tool. Accession numbers are generated within 15 minutes of submission, pre-published data sets can be kept private, and submitter's identity can be hidden for double-blind review.

W504: Genomic features and chromosome functionality

Evolution of microRNAs and their Targets after Ancient Polyploidizations in Plants

Liuyu Qin and **Yuannian Jiao**, Institute of Botany, The Chinese Academy of Sciences, Beijing, China

MicroRNAs (miRNAs) are a specialized class of small silencing RNAs that regulate gene expression in animals and plants. The evolutionary dynamics of duplicated protein-encoding genes after polyploidizations, has been studied extensively. However, the evolutionary patterns and consequences of duplicated *MIRNAs* and the potential influence on the evolution of their targets are poorly understood. Here, we examined the evolutionary pattern of miRNAs and their targets subsequent to five independent ancient polyploidization events across the phylogeny of flowering plants. Overall, the retention of duplicated *MIRNAs* appears to be associated with the retention of their corresponding duplicated target genes. In addition, conserved miRNAs are more likely to be retained as duplicates than young miRNAs, while the mature miRNA region of conserved *MIRNAs* show higher sequence conservation than that of young miRNAs. Finally, we found that the duplicated miRNAs with sequence variations could potentially regulate more target genes and thereafter fit into novel biological regulation networks. Together, these results shed light on evolutionary patterns of *MIRNAs* following genome duplications and evolutionary interplay between *MIRNAs* and their target genes.

W505: Genomic features and chromosome functionality

The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis

Li-zhi Gao, Kunming University of Science and Technology, Kunming, China

W506: Genomic features and chromosome functionality

Fertility and Meiotic Stability in Novel *Brassica* Crop Types

Margaret W. Mwathi, University of Queensland, Brisbane, Australia, Jacqueline Batley, University of Western Australia, Perth, Australia and Annaliese S Mason, Justus Liebig University, Giessen, Germany

The *Brassica* genus contains a large number of food crops, including oilseeds, vegetables and condiments. Species in this genus are closely related, enabling them to be easily crossed to each other for agricultural improvement. Three diploid species (*B. rapa* - Chinese cabbage; *B. nigra* – black mustard; *B. oleracea* - cauliflower, cabbage,) have the A, B and C genomes, while three allotetraploids (two-genome species: *B. juncea* – Indian mustard; *B. napus* – rapeseed; *B. carinata* – Ethiopian mustard) have genome complements AB, AC and BC. We aim to

produce a *Brassica* crop type with three different genomes, i.e. ABC, which will contain genetic diversity and important agronomic traits from all six related *Brassica* crop species. The main challenge facing production and agronomic utilisation of these hybrids is the problem of meiotic instability, which results in subsequent loss of chromosomes and poor fertility in early generations.
We tested allohexaploid *Brassica* genotypes derived from the crosses *B. rapa* × *B. carinata* and (*B. napus* × *B. carinata*) × *B. juncea* for fertility and meiotic chromosome behaviour. Meiotic behaviour, plant fertility and agronomic traits were assessed in two large populations derived from these two hybrid types. Interspecific hybridization between *Brassica juncea* × *Brassica oleracea* followed by embryo rescue was also carried out and confirmed triploid hybrids produced, followed by chromosome doubling to create allohexaploid plants.
Parent and hybrid genotype was found to influence fertility, agronomic and meiotic traits, including the frequency of chromosome loss and abnormal chromosome pairing behaviour during meiosis. Statistical analysis found significant differences between genotypes and the mean of plant height, as well as between progeny sets and pollen fertility, plant height and seed set. Plant height in the allohexaploids was found to be higher than in parent controls, indicative of heterosis for growth traits. Further SNP genotype analysis using the Illumina Infinium Brassica 90K array will be carried out to determine A/B/C chromosome interactions in subsequent generations and to provide insight into genetic stability in *Brassica* allohexaploid species, critical in the formation of a new stable crop species for agricultural benefit.

W507: Genomic features and chromosome functionality
Efforts to Construct Comparative Genomics Database of Legumes and other Plant Groups
Xiyin Wang, UGA and NCST, Watkinsville, GA
Mainly due to their economic importance, genomes of 10 legumes, including soybean, wild peanuts, barrel medic, etc, have been sequenced. However, a family-level comparative genomics analysis has been unavailable. With grape and selected legume genomes as outgroups, we managed to perform a hierarchical and event-related alignment of these genomes and deconvoluted layers of homologous regions produced by ancestral polyploidizations or speciations. Consequently, we illustrated genomic fractionation characterized by wide-spread gene losses after the polyploidizations. Notably, high similarity in gene retention between recently duplicated chromosomes in soybean supported a likely autopolypoidy nature of its tetraploid ancestor. Moreover, though mostly gene losses were nearly random, largely but not fully described by geometric distribution, we showed that polyploidization contributed divergently to copy number variation of important gene families. Besides, we showed significantly divergent evolutionary levels among legumes, and by performing Ks correction, re-dated major evolutionary events during their expansion. The present effort laid a solid foundation further genomics exploration in the legume research community and beyond. We described only a tiny fraction of legume comparative genomics analysis that we performed, and more information was stored in the newly constructed Legume Comparative Genomics Research Platform (www.legumegrp.org). The present effort and recent ones in grasses have been extended into other plant families, aiming at constructing user-friendly comparative genomics platforms.

W508: Genomic features and chromosome functionality
An Overlooked Paleo-Tetraploidization in Cucurbitaceae and a Gold Standard to Decipher Complex Genomes
Jinpeng Wang[1], Yuxian Li[1] and Xiyin Wang[2], (1)North China University of Science and Technology, Tangshan, China, (2)UGA and NCST, Watkinsville, GA
Cucurbitaceae plants are of considerable biological and economic importance, and genomes of cucumber, watermelon, and melon have been sequenced. However, a comparative genomics exploration of their genome structures and evolution has not been available. Here, we aimed at performing a hierarchical inference of genomic homology resulted from recursive paleo-polyploidizations. Unexpectedly, we found that, shortly after a core-eudicot-common hexaploidy (ECH), a cucurbit-common tetraploidization (CCT) occurred, overlooked by previous reports. Moreover, we characterized gene loss (and retention) after these respective events, which were significantly unbalanced between inferred subgenomes, and between plants after their split. The inference of a dominant subgenome and a sensitive one suggested an allotetraploid nature of the CCT. Besides, we found divergent evolutionary rates among cucurbits, and after doing rate correction, we dated the CCT to be 90-102 million years ago, likely common to all Cucurbitaceae plants, showing its important role in the establishment of the plant family. Furthermore, we found that polyploidizations contributed to the expansion of key functional gene families. The present efforts laid a comparative genomics platform to support various researches in the Cucurbitaceae community and beyond.

W509: Genomics-Assisted Breeding
Welcome & Introduction
Rajeev Varshney, ICRISAT, Hyderabad, India

W510: Genomics-Assisted Breeding
From 1 to Thousands Barley Genomes – Ways to Unlock the Global Barley Genome Diversity for Breeding and Research
Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany
One reference genome sequence is very useful; however, it is telling only part of the story of genome diversity of a given species. In barley we know that up to 10% of the coding sequence may be affected by any kind of structural variation. Establishing additional genome assemblies of diverse haplotypes is one activity required to describe the barley pan-genome. Sequencing of thousands of 5 Gbp genomes to reference quality is not exactly needed and would still be prohibitively expensive. From a breeders or seed bank manager`s perspective, however, there is a desire to fully understand the species genome complexity and, in the future, maybe re-sequence genomes as a routine at real-time in the field. To get a first step into this direction we characterized the genomic complexity of all 20,000 barley seed accessions at IPK Gatersleben, Germany, hosting the largest international ex situ seed bank of crop plants in the EU28. This unprecedented level of resolution about barley genome diversity is facilitating diversity studies, allele mining and GWAS but also potentially changing ex situ seed banks in the future from storage facilities into general information hubs for research and breeding.

W511: Genomics-Assisted Breeding
Unlocking Crop Genomic Diversity for Breeding: Lesson from the Pearl Millet Genome

Yves P. Vigouroux, IRD, Montpellier, France
Plant breeding uses more and more genomic diversity to accelerate the development of new varieties. Several new approaches have been developed to identify functional diversity variation using such genomic data. Past evolutionary history of the crop shapes both its neutral and functional diversity. Moreover, wild diversity, still largely unexplored, could also now be more efficiently studied and used. Finally, new phenotypes could be used for breeding from root architecture to variety shaped bacterial community. We will illustrated these different strategies and studies in an important Sahelian cereal, pearl millet. Using whole genome sequence and new breeding traits, such advances could considerably enhance our ability to breed better crop in a hotter climate.

W512: Genomics-Assisted Breeding
Speed GS: Accelerating Genetic Gain in Wheat
Ben Hayes, University of Queensland, Brisbane, Australia
The genetic improvement of modern wheat varieties has been very successful. However annual yield increases need to be doubled over the next few decades and the global production trends in all major wheat growing regions indicate a yield plateau. To overcome this, innovative strategies that efficiently integrate modern technologies in breeding programs are required. Using simulations based on real wheat data sets, we demonstrate how genomic selection and “speed breeding”, a novel rapid generation advancement technology, can be combined to substantially reduce the length of the breeding cycle and maximise genetic gain per unit time. We outline the opportunities and challenges associated with the fusion of these breeding tools and reinforce the importance of integrating novel genetic diversity in breeding programs to achieve sustainable long-term genetic gain.

W513: Genomics-Assisted Breeding
Searching for an Accurate Marker-Based Prediction of an Individual Quantitative Trait in Molecular Plant Breeding
Yong-Bi Fu[1], Mo-Hua Yang[1,2], Fangqin Zeng[1] and Bill Biligetu[3], (1)Plant Gene Resources of Canada, Saskatoon, SK, Canada, (2)College of Forestry, Central South University of Forestry & Technology, Changsha, Hunan, China, Changsha, Hunan, China, (3)Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada
Molecular plant breeding with the aid of molecular markers has played an important role in modern plant breeding over the last two decades. Many marker-based predictions for quantitative traits have been made to enhance parental selection, but the trait prediction accuracy remains generally low, even with the aid of dense, genome-wide SNP markers. To search for more accurate trait-specific prediction with informative SNP markers, we conducted a literature review on the prediction issues in molecular plant breeding and on the applicability of an RNA-Seq technique for developing function-associated specific trait (FAST) SNP markers. To understand whether and how FAST SNP markers could enhance trait prediction, we also performed a theoretical reasoning on the effectiveness of these markers in a trait-specific prediction, and verified the reasoning through computer simulation. To the end, the search yielded an alternative to regular genomic selection with FAST SNP markers that could be explored to achieve more accurate trait-specific prediction. Continuous search for better alternatives is encouraged to enhance marker-based predictions for an individual quantitative trait in molecular plant breeding.

W514: Genomics-Assisted Breeding
The Two-Layer Multiple-for-Multiple Genetic Interaction System between R-Genes and Virulence Genes underlying the Arms-Race between Rice and *Xanthomonas oryzae* pv. *oryzae* Revealed by Two-Dimensional GWAS
Fan Zhang[1], Zhi-Qiang Hu[1], Fan Zhang[1,2], Zhi-Chao Wu[1], Xi-Yin Wang[3], Jing-Peng Wang[3], Ming-Ming Wang[1], Yan-Ru Cui[1], Ying-Yao Shi[4], Da-Long Zhou[1,4], Casiana Vera Cruz[5], Dan-Dan Hu[1,4], Zheng-Lin Du[5], Wen-Sehng Wang[1], Xiu-Qin Zhao[1], Jian-Long Xu[1,6], Bin-Ying Fu[1], Jauhar Ali[5], Yong-Li Zhou[1,6] and **Zhi-Kang Li**[1,6], (1)Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China, (2)Graduate School of Chinese Academy of Agricultural Sciences, Beijing, China, (3)School of Life Sciences, North China University of Science and Technology, Hebei, China, (4)Anhui Agricultural University, Hefei, China, (5)International Rice Research Institute, Metro Manila, Philippines, (6)Shenzhen Institute of Breeding and Innovation, Chinese Academy of Agricultural Sciences, Shenzhen, China

W515: Genomics-Assisted Breeding
The GOBII Project: Transforming Breeding through Genomic Data Management and Decision Support
Elizabeth Jones[1], Star Yanxin Gao[1], Yaw A. Nti-Addae[1], David Matthews[2], Kate A. Dreher[3], Susanne Dreisigacker[3], Michael S. Olsen[4], Tobias Kretzschmar[5], David Marshall[6], Iain Milne[6], Andrzej Kilian[7], Lukas A. Mueller[8], Mark E Sorrells[2], Qi Sun[2], Edward S. Buckler[9], Jean-Luc Jannink[10], Kelly Robbins[2] 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, MINNEAPOLIS, MN, Kenya, (5)International Rice Research Institute, Los Baños, Philippines, (6)The James Hutton Institute, Dundee, United Kingdom, (7)Diversity Arrays Technology Pty Ltd (DArT PL), Canberra, Australia, (8)Boyce Thompson Institute, Ithaca, NY, (9)School of Integrative Plant Sciences, Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY, (10)USDA-ARS, Cornell University, Ithaca, NY
The Genomic and 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 in the developing world. Cornell University and the Boyce Thompson Institute in Ithaca, NY are partnering with the CGIAR research centers CIMMYT, ICRISAT and IRRI to develop open-source genomic data management, analysis and decision support systems that are critical to realizing these goals. 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. The source-code for this system is available at: https://github.com/gobiiproject. We are now working to connect information across adjacent data management systems using breeding specific

APIs (BrAPIs: http://docs.brapi.apiary.io) so that breeders will be able to access data from a single entry point. We are working with the James Hutton Institute to adapt the tool Flapjack for breeding-specific use cases relating to genomic data visualization, Diversity Arrays Technologies for data QC using KDCompute, and building genomic selection pipelines in Galaxy. We plan to house all open-source tools in a web-based portal to allow for easy access by breeders. Our goal is to support breeders in the developing world in increasing the rates of genetic gain over time through adoption of marker-assisted and genomic selection techniques.

W516: Genomics-Assisted Breeding
Summary & Wrap-up
Rajeev K Varshney, ICRISAT, Hyderabad, India

W517: Genomic Selection and Genome-Wide Association Studies
Substantial Contribution of Genetic Variation in the Expression of Transcription Factors to Phenotypic Variation Revealed by eRD-GWAS
Hung-Ying Lin[1], Qiang Liu[1], Xiao Li[2], Jinliang Yang[3], Sanzhen Liu[4], Yinlian Huang[5], Michael Scanlon[6], Dan Nettleton[1] and **Patrick S. Schnable**[1], (1)Iowa State University, Ames, IA, (2)The Broad Institute, Cambridge, MA, (3)University of California-Davis, Davis, CA, (4)Kansas State University, Manhattan, KS, (5)DATA Biotechnology (Beijing) LTD., Beijing, China, (6)Cornell University, Ithaca, NY
There are significant limitations in existing methods for the genome-wide identification of genes whose expression patterns affect traits. The transcriptomes of five tissues from 27 genetically diverse maize inbred lines were deeply sequenced to identify genes exhibiting high and low levels of expression variation across tissues or genotypes. Transcription factors are enriched among genes with the most variation in expression across tissues, as well as among genes with higher-than-median levels of variation in expression across genotypes. In contrast, transcription factors are depleted among genes whose expression is either highly stable or highly variable across genotypes. We developed a Bayesian-based method for genome-wide association studies (GWAS) in which RNA-seq-based measures of transcript accumulation are used as explanatory variables (eRD-GWAS). The ability of eRD-GWAS to identify true associations between gene expression variation and phenotypic diversity is supported by analyses of RNA co-expression networks, protein–protein interaction networks, and gene regulatory networks. Genes associated with 13 traits were identified using eRD-GWAS on a panel of 369 maize inbred lines. Predicted functions of many of the resulting trait-associated genes are consistent with the analyzed traits. Importantly, transcription factors are significantly enriched among trait-associated genes identified with eRD-GWAS. eRD-GWAS is a powerful tool for associating genes with traits and is complementary to SNP-based GWAS. Our eRD-GWAS results are consistent with the hypothesis that genetic variation in transcription factor expression contributes substantially to phenotypic diversity.

W518: Genomic Selection and Genome-Wide Association Studies
Understanding the Regulatory Mechanisms in Complex Trait Variation
Jian Yang, Institute for Molecular Bioscience, Brisbane, Queensland, Australia
Genome-wide association studies (GWAS) have identified a large number of genetic variants associated with human complex traits. However, genes or functional DNA elements through which these variants exert their effects on the traits are often unknown. I will present a method (called SMR) that integrates summary-level data from GWAS and expression quantitative trait locus (eQTL) studies to identify genes whose expression levels are associated with a trait because of pleiotropy. Applying the method to five human complex traits using GWAS data on up to 339,224 individuals and eQTL data on 5,311 individuals identifies 126 putative functional genes, of which 25 genes are novel and 77 genes (~61%) are not the nearest genes of the GWAS top SNPs. I will show how the SMR method can be extended to include other sources of omics data to understand the regulatory mechanisms underpinning polygenic variation in complex traits.

W519: Genomic Selection and Genome-Wide Association Studies
Do Stronger Measures of Genomic Connectedness Enhance Prediction Accuracies across Management Units?
Gota Morota, University of Nebraska Lincoln, Lincoln, NE
Genetic connectedness assesses the extent to which estimated breeding values can be fairly compared across management units. Ranking of individuals across units based on best linear unbiased prediction (BLUP) is reliable when there is a sufficient level of connectedness due to a better disentangling of genetic signal from noise. Although a recent study showed that genomic relatedness strengthens the estimates of connectedness across management units compared to that of pedigree, the relationship between connectedness measures and prediction accuracies has been explored only to a limited extent. In this study, we examined whether increased measures of connectedness led to higher prediction accuracies evaluated by a cross-validation based on computer simulations. We applied prediction error variance of difference, coefficient of determination, and BLUP-type prediction models to data simulated under various scenarios. We found that the greater extent of connectedness enhanced accuracy of whole-genome prediction. Connectedness across units increased with the proportion of connecting individuals and this increase was associated with improved accuracy of prediction. The use of genomic information resulted in increased estimates of connectedness and improved prediction accuracies compared to those of pedigree-based models, especially when the numbers of markers and causal loci were large. We also demonstrate the potential of non-parametric relationship matrices to quantify genomic connectedness and prediction accuracy in the presence of non-additive gene actions.

W520: Genomic Selection and Genome-Wide Association Studies
GWAS Models for Multiple Traits and Environments
Fred E. van Eeuwijk[1], Martin P. Boer[2], Willem Kruijer[2] and Marcos Malosetti[2], (1)Wageningen UR - Biometris, Wageningen, Netherlands, (2)Wageningen University & Research - Biometris, Wageningen, Netherlands

W521: Genomic Selection and Genome-Wide Association Studies
Genomic Feature Prediction Models
Peter Sørensen, Aarhus University, Tjele, Denmark
We have developed predictive models that use prior biological information on genomic features. Genomic features, consisting of a set of genetic markers, are regions of the genome that are linked to some external information (e.g. genes, biological pathways, gene ontologies, or sequence annotation). The main idea underlying this modeling approach is that these regions are enriched for causal variants affecting a specific trait or disease. We have demonstrated that genomic feature based prediction models can improve prediction accuracy for a range of complex traits such as feed efficiency in pigs, milk yield and mastitis in dairy cattle, various treatment responses in fruit fly, and body mass index and height in humans. Predictive models based on genomic features can be computationally intensive. We have therefore developed and evaluated a number of genetic marker set tests. These approaches are computationally very fast allowing us to rapidly analyze different layers of genomic feature classes to discover genomic features potentially enriched for causal variants. Information from genetic marker set test can be used to build more accurate prediction models. Our simulation studies have illustrated the impact of trait and genomic feature-specific factors on prediction accuracy and power to detect genomic features enriched for causal variants. Furthermore genomic feature models provide a formal statistical framework for leveraging and evaluating information across multiple experimental studies to provide novel insights into the genetic architecture of complex traits and diseases.

W522: Genomics of Crop Ecosystem Services
Genomics of Crop Ecosystem Services
Andrew H. Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

W523: Genomics of Crop Ecosystem Services
Accelerating the Domestication and Improvement of the Perennial Grain Crop *Thinopyrum intermedium* with Genomics
Kevin M. Dorn, Kansas State University, Manhattan, KS

W524: Genomics of Crop Ecosystem Services
Genetic Dissection of Sorghum Life History Traits Salient to Climate-Resilient Sustainable Intensification of Agriculture
Wenqian Kong, Plant Genome Mapping Laboratory, Athens, GA
Perennial crops may provide consistent food and biomass supplies while preserving ecological capital and reducing water and energy inputs. Sorghum, one of the few multi-purpose crops that provide both grain and biomass in some of the world's most adverse conditions, has two perennial relatives as well as rich morphological diversity occurring during divergent evolution both in the wild and under domestication, making it a good candidate for breeding for perenniality. We describe a quantitative study to elucidate the genetic components conferring perenniality related traits, mainly rhizomatousness and winter survival, in two novel BC1F2 populations totaling 246 genotypes derived from backcrossing two Sorghum bicolor x S. halepense F1 plants to a tetraploidized S. bicolor. Phenotyping for two years in Bogart, GA and Salina, KS permits understanding of the relationship between rhizomatousness and winter survival, providing materials for future plant breeding. Correspondence of rhizomatousness quantitative trait loci (QTLs) with two populations derived from its progenitors, S. bicolor × S. propinquum, suggests the set of QTLs inherited and genetic novelty arisen after the divergence from its progenitors, while comparisons to tillering and branching QTLs further support their developmental relationship of these two organs. An unexpected finding supported from both the BC1F2 and PQ-F2 populations suggests alleles contributing to late flowering are related to fewer rhizomatousness. Interestingly, twelve out of sixteen QTL regions conferring rhizomatousness fall in genome-duplicated regions, indicating that those genes retain their function after the duplication event 96 million years ago. This study also assists in narrowing down candidate genes for rhizomatousness from expression profiling, provide diagnostic DNA markers, and facilitate breeding for perenniality in sorghum.

W525: Genomics of Crop Ecosystem Services
Developing Perennial Maize: New Approaches Towards an Old Dream
Seth Murray, Texas A&M University, College Station, TX
Maize (*Zea mays* L., corn) has two perennial wild relatives, perennial teosinte *Z. perennis* (2n=4=40) and diploperennial teosinte *Z. diploperennis* (2n=2x=20). Interfertility between maize and *Z. perennis* was first demonstrated by Emerson and Beadle in 1930. The excitement of developing perennial maize has been ongoing ever since. Yet, still no maize that successfully overwinters in most of the country and produces a true ear has been developed; let alone an overwintering maize with competitive yield. Both perennial *Zea* species easily overwinter in sub-tropical climates but are much less aggressive in plant growth, rhizome production, and regrowth than the complementary perennial wild relatives of sorghum – especially under high temperatures, making success (and sufficient seed to test) illusive. After germplasm derived from domesticated by perennial wild crosses were further backcrossed to *Z. diploperennis* for two generations, a few plants successfully overwintered and retained some maize characteristics in Texas. Yet to date, the differences and genetic structure between the domesticated and perennial *Zea* remain too great for association studies and the populations of successful integrations too small to conduct genetic linkage mapping. Working with perennial germplasm derived from past investigations for nearly 10 years has sled to multiple new testable hypothesis and strategies. We now have some evidence to validate at least one novel hypothesis, the direction of crossing matters, likely due to the inheritance of cytoplasm and /or mitochondria genomes.

W526: Genomics of Genebanks
Validation of a Strategy for Moving Valuable, Novel Wheat Genetic Diversity from Germplasm Banks to Breeding Pipelines
Sukhwinder Singh, CIMMYT- International Maize and Wheat Improvement Centre, Texcoco, Mexico
Wheat breeding programs effectively harness the potential of elite germplasm pools to contribute to worldwide food security. However, to maintain or increase rates of genetic gains, valuable genetic diversity from un-adapted exotics (including landraces, synthetics and wild

relatives) held in germplasm banks must be channeled into breeding germplasm pools. CIMMYT's Seeds of Discovery project has been using cutting-edge DNA sequencing tools and informatics to identify and use high-value traits from exotic germplasm collections, with particular emphasis on drought and heat tolerance, and yellow rust and powdery mildew resistance. More than 90,000 accessions of CIMMYT's and ICARDA's germplasm banks have been sequenced using DArTseq technology. More than 1,000 accessions, including landraces, wild relatives and synthetics have been crossed with elite lines to generate 75 percent elite and 25 percent exotic bridging germplasm, and more than 2000 advanced-generation bridging lines have been developed from these three-way top-crosses. Evaluation of these bridging lines by partner researchers in Mexico, India and Pakistan have identified bridging lines for use in breeding for heat, drought, yellow rust and powdery mildew. Association of genomic haplotypes with phenotypic performance identified valuable contributions from the exotic parents to the bridging lines for heat, drought, yellow rust and powdery mildew. The effectiveness of the three-parent top cross strategy reported herein for channeling exotic genetic resources into elite wheat breeding programs was validated by the resulting bridging lines which are agronomically acceptable and contain novel, favorable haplotypes for the prioritized traits.

W527: Genomics of Genebanks

Germplasm Enhancement of Maize (GEM) - 24 Years of Public-Private Sector Collaboration to Increase US Maize Genetic Diversity

Candice Gardner, USDA-ARS and Iowa State University, Ames, IA, Matthew Krakowsky, USDA-ARS and North Carolina State University and David Peters, USDA-ARS and Iowa State University

The Germplasm Enhancement of Maize (GEM) Project is a mission-oriented, cooperative research effort of the USDA/ARS, land grant universities, private industry, and international agricultural research centers to broaden the germplasm base of maize cultivated within the US. The Raleigh location of the GEM Project is focused on identifying new exotic sources of maize germplasm and on developing 50% exotic/50% temperate germplasm, while the Ames location focuses on developing 25% exotic/75% temperate germplasm with high yield potential and resistance to common foliar, stalk and ear diseases which can be incorporated directly into commercial maize breeding programs. This is challenging because tropical maize is photoperiod sensitive and lacks adaptation to temperate conditions. While exotic germplasm can be defined as any germplasm that has not been sampled in a breeding program, the GEM project uses the term to cover landraces and improved germplasm from tropical and subtropical origins. The environmental conditions and geographical latitude of the Raleigh location make it feasible to work with breeding material that contains a higher percentage of exotic germplasm than is usually practical at Midwestern locations such as Ames, IA, and this enables the transfer of more genetic diversity into prebreeding germplasm.

Private sector collaborators provide proprietary germplasm and nursery and testing resources to characterize performance for a wide variety of traits.University collaborators focus on specific traits or areas of interest, and have released improved germplasm for biotic and abiotic stress resistance, for novel starch properties, and to understand the nature of diversity. Graduate students in US public universities have gained experience with introgression breeding and maize genetic resources, critical training to provide for continued crop improvement, and public breeders have released inbred lines and synthetics.

In addition to developing germplasm for use in maize breeding programs, the GEM project also develops new resources that allow the introgression of useful alleles from agronomically inferior exotic sources. This is known as the Allelic Diversity (AD) project, and involves crossing and backcrossing accessions from all of the races of maize to two formerly proprietary temperate inbreds to develop a panel of lines that represent the diversity of maize. This panel can be used to screen for alleles of interest that would otherwise be unavailable to maize researchers. To date, GEM project participants have released about 300 conventionally derived, diverse lines. Additionally, more than 200 double haploid lines have been released as part of the AD project.

W528: Genomics of Genebanks

Mining Genetic Variation in Genebank Collections

Kelly Robbins[1], Qi Sun[1], Christopher M. Richards[2], David Marshall[3] and Sarah Hearne[4], (1)Cornell University, Ithaca, NY, (2)USDA ARS National Laboratory for Genetic Resources Preservation, Fort Collins, CO, (3)The James Hutton Institute, Dundee, United Kingdom, (4)CIMMYT- International Maize and Wheat Improvement Centre, Texcoco, Mexico

Germplasm collections stored in genebanks represent a valuable source of genetic diversity that is largely underutilized in breeding programs. This aversion to use stems from challenges in deriving high performing lines from crosses with gene bank accessions, due to significant unfavorable genetic variation that accompanies any useful genetic variants. More targeted breeding approaches to backcross genomic regions with large favorable effects has shown success, but given sparse information, it is difficult for breeders to identify accessions with useful novel variants. While large-scale phenotyping of accessions will remain a challenge without high-throughput cost effective and field relevant screens, the dropping cost of next-generation sequencing (NGS) technologies makes high-density genotyping and low coverage skim sequencing a viable option for interrogating whole genebank collections. One challenge of this NGS approach is the ascertainment bias and lack of genomic position context of existing variant call approaches. The use of a single reference genome, or a narrow range of diversity, is insufficient to adequately capture the full range of genetic variation in most collections. Methods are being developed in maize and grape projects for constructing pan-genomes, calling variants, and imputing missing information using a practical haplotype graph (PHG) approach. To make use of these capabilities, an open-source information system is proposed to enable variant calling, imputation, efficient storage, and analysis tools for mining of genomic variants and curation of genebank germplasm collections. An overview of the proposed genomic information system will be presented.

W529: Genomics of Genebanks

Mango Genebank Diversity

Emily Warschefsky, Florida International University, Miami, FL and Eric Bishop-von Wettberg, University of Vermont, Burlington, VT

The maintenance of crop genetic variation in genebanks is essential to pre-breeding efforts that will ensure the future success of many of today's most important crops. Therefore, it is critical that genetic repositories are managed in a way that maximizes the amount of genetic

diversity preserved in each collection. This is all the more true for perennial crops, which require significant investments of space, time, and money to cultivate. Yet in addition to providing valuable genetic resources, diverse genebank collections can help us understand a crop's history of domestication and evolution. The mango, *Mangifera indica* L. (Anacardiaceae) is a perennial fruit tree that has been cultivated on the Indian subcontinent for 4,000 years or more. Today, the mango is one of the world's most-produced tropical fruits. We have used restriction-site associated DNA sequencing to characterize the mango genebank at Fairchild Tropical Botanic Garden, which consists of over 200 cultivars from around the world. To explore the mango's domestication history, we look for evidence of genetic bottlenecks associated with human-mediated dispersal into Africa and the Americas, characterize the geographic distribution of genetic diversity to identify regions that harbor unique diversity. While we find little evidence to indicate the presence of a secondary genetic bottleneck in the mango's past, population structure within *M. indica* suggests that mango may have been domesticated multiple times or undergone introgression with wild relatives following domestication. Collectively, this work informs the management of an important mango genebank and provides insight into the domestication history of the mango.

W530: Genomics of Non-Classical Model Animals
Epigenetic Aging Clocks for Mammals
Steve Horvath, UCLA David Geffen School of Medicine, Los Angeles, CA
In mammals, DNA methylation represents a form of genome annotation that regulates gene expression by serving as a maintainable mark whose absence marks promoters and enhancers following each round of replication.
Finding reliable biological measures of aging has been a longstanding research priority, based on the premise that these biomarkers would lead to a better understanding of how aging increases susceptibility to certain diseases, along with identifying strategies for promoting healthy aging.
We recently developed a multi-tissue epigenetic age estimation method that combines the DNA methylation levels of 353 epigenetic markers knowns as CpGs. The weighted average of these 353 epigenetic markers gives rise to an estimate of tissue age (in units of years), which is referred to as "DNA methylation age" or as "epigenetic age". DNA methylation age is highly correlated (r=0.96) with chronological age across the entire lifespan. We and others have shown that the human epigenetic clock relates to biological age (as opposed to simply being a correlate of chronological age), e.g. the DNA methylation age of blood is predictive of all-cause mortality even after adjusting for a variety of known risk factors.
We have demonstrated the human epigenetic clock applies without change to chimpanzees. However, it no longer applies to other animals due to lack of sequence conservation. We recently generated an epigenetic clock for dogs and wolves based on DNA data methylation data from the blood of 108 canids.
In this talk, I will review our ongoing effort to develop analogous epigenetic clocks for many different mammals.
Acknowledgement: This work is supported by the Paul G. Allen Foundation

W531: Genomics of Non-Classical Model Animals
Adaptation to Free Diving in the Bajau Sea Nomads of Indonesia
Rasmus Nielsen, Department of Integrative Biology and Statistics, Berkeley, CA
Understanding the physiology and genetics of human hypoxia tolerance has important medical implications. The only natural system thus far investigated is high altitude human populations. Another system, yet to be explored, is humans who engage in breath-hold diving. The indigenous Bajau people (sea nomads) of Southeast Asia live a subsistence lifestyle based on diving, and are renowned for their extraordinary breath holding abilities. However, it is unknown whether this has a genetic basis. Using a comparative genomic study, we show that natural selection on specific genetic variants in have increased the spleen size in the Bajau, providing them with a larger reservoir of oxygenated red blood cells. We also find evidence of strong natural selection specific ona gene affecting the human diving reflex. The Bajau, and possibly other natural diving populations, provide a new opportunity to study human adaptation of hypoxia tolerance.

W532: 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[1], Andrej Fabrizius[2], Alena Krueger[2], Cornelia Gessner[2] and Lars Folkow[3], (1)Universität Hamburg, Zoologisches Institut, Hamburg, Germany, (2)University of Hamburg, Hamburg, Germany, (3)The Arctic University of Norway, Tromso, Norway
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 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.

W533: Genomics of Non-Classical Model Animals
Exploiting 13-Lined Ground Squirrel Genomics to Decipher Hibernation Biology
Katharine Grabek, Department of Biomedical Data Science, Stanford University, Stanford, CA, Thomas F. Cooke, Stanford University, Stanford, CA, L. Elaine Epperson, Center for Genes, Environment and Health, National Jewish Health, Denver, CO,

Kaitlyn K. Spees, Department of Genetics, Stanford University, Stanford, CA, Gleyce F. Cabral, Brazil Scientific Mobility Program, CAPES Foundation, Brasília, Brazil, Shirley C. Sutton, Department of Cardiovascular Medicine, Stanford University, Stanford, CA, Dana K. Merriman, Department of Biology, University of Wisconsin Oshkosh, Oshkosh, WI, Sandy Martin, Cell and Developmental Biology, University of Colorado School of Medicine, Aurora, CO and Carlos D. Bustamante, Dept. of Biomedical Data Science, Stanford University, Stanford, CA
Hibernation is a highly dynamic phenotype whose timing, for many mammals, is controlled by a circannual clock and accompanied by rhythms in body mass and food intake. When housed in an animal facility, 13-lined ground squirrels exhibit individual variation in the seasonal onset of hibernation, which is not explained by environmental or biological factors, such as body mass and sex. We hypothesized that underlying genetic architecture instead drives variation in this timing. In this study, we first increased the contiguity of the draft genome assembly using a long-range scaffolding technique. We next employed a genotype-by-sequencing approach to characterize genetic variation in 153 13-lined ground squirrels. Combining this with datalogger records, we estimated high heritability for the seasonal onset of hibernation. After applying a genome-wide scan with 46,996 variants, we also identified 21 loci significantly associated with hibernation immergence, which accounted for most of the variance in the phenotype. Genes near these loci were functionally related to hibernation physiology, including insulin signaling and processing, control of food intake, and control of heart rate. Finally, we applied an expression quantitative loci (eQTL) analysis using existing transcriptome datasets to identify significant cis-eQTL associations for 9/21 variants. Our results highlight the power of applying a genetic mapping strategy to hibernation and present new insight into the genetics driving its seasonal onset.

W534: Genomics of Non-Classical Model Animals
Genomics of Speciation and Adaptation in North American Desert Tortoises (genus *Gopherus*)
Marc Tollis[1], Taylor Edwards[2], Dale F. DeNardo[3], John A. Cornelius[3], Greer A. Dolby[3], Robert W. Murphy[4] and Kenro Kusumi[3], (1)Biodesign Institute, Arizona State University, Tempe, AZ, (2)University of Arizona, Tucson, AZ, (3)Arizona State University, Tempe, AZ, (4)University of Toronto, Toronto, ON, Canada
Desert tortoises (genus *Gopherus*) are native to the Mojave, Sonoran, and Sinaloan deserts of North America. While studies of these enigmatic reptiles could shed light on the genetic basis of adaptations for longevity and aridity, modes of speciation, and prioritizing conservation efforts for the threatened Agassiz's desert tortoise (*G. agassizii*), genomic resources have not existed for this species group. To address this need, we obtained mitochondrial and nuclear sequence data and blood RNA-seq data from *G. agassizii* in Nevada and California, and G. morafkai in Arizona and Sinaloa, Mexico. Results of population genomic and demographic analyses show that *G. morafkai* consists of two evolutionarily distinct lineages, one of which has been recently elevated to species status (*G. evegoodei*), whose divergence occurred in the absence of both gene flow despite no geographic barrier. We also obtained deep transcriptomic RNA-seq from multiple tissues (lung, brain, and muscle) and 147X Illumina whole genome shotgun data to assemble and annotate a genome assembly for *G. agassizii*. We identify >20,000 protein-coding genes in the *G. agassizii* assembly and demonstrate (1) that turtles are among the slowest evolving amniotes, (2) amino acid changes in genes controlling shell development, longevity and osmoregulation, and (3) fixed variants in genes across the *Gopherus* species complex in genes related to desert adaptations such as circadian rhythm and innate immune response. Thus, we report the first genome reference and annotation for any tortoise, and demonstrate its utility for investigating genomic factors affecting tortoise health, disease and longevity.

W535: Genomics of Phytoremediators, Metal Accumulators and Relatives
Genomic Systems Approaches Towards Understanding Heavy Metal Response and Resistance Mechanisms in Plants
Julian Schroeder, Division of Biological Sciences, University of California, San Diego, La Jolla, CA

W536: Genomics of Phytoremediators, Metal Accumulators and Relatives
Multi-Ion Salt Stress Adaptation Explored using Extremophyte Genomics
Maheshi Dassanayake, Louisiana State University, Baton Rouge, LA

W537: Genomics of Phytoremediators, Metal Accumulators and Relatives
Genomics of Rapid Adaptation to Cu Mine in *Mimulus guttatus*
John H. Willis, Duke University, Durham, NC

W538: Genomics of Phytoremediators, Metal Accumulators and Relatives
Bioinformatics Approaches to Identifying Candidate Genes for Elemental Accumulation
Ivan Baxter, USDA-ARS, Danforth Plant Science Center, St. Louis, MO

W539: Genomics of Phytoremediators, Metal Accumulators and Relatives
Treating Toxins with Tree Microbes: Enhanced Phytoremediation with Plant-Microbe Partnerships
Sharon L. Doty[1], John L. Freeman[2], Andrea Firrincieli[3], Robert Tournay[1], Pierre M. Joubert[1], Christopher Cohu[4], Joel Burken[5], Zareen Khan[1] and Michael J. Blaylock[6], (1)University of Washington, Seattle, WA, (2)Intrinsyx Technologies Corporation, Moffett Field, CA, (3)University of Tuscia, Viterbo, Italy, (4)Phytoremediation and Phytomining Consultants United, Moffett Field, CA, (5)Missouri S&T, Rolla, MO, (6)Edenspace Systems Corporation, Purcellville, VA
Phytoremediation technologies can be improved with the utilization of effective microbial endophyte strains. Endophytes, the microbes living within plants, can provide multiple benefits, including pollutant detoxification, increased rooting, and improved tolerance to abiotic stresses. In a 3-year field trial at a TCE groundwater plume EPA-designated Superfund site, poplar trees inoculated with a natural TCE-degrading endophyte strain had enhanced survivability and growth compared to mock-inoculated control trees. The inoculated trees exhibited increased chloride exudation in the rhizosphere, and leaf tissue analysis confirmed enhanced TCE degradation. A TnSeq approach is underway to determine the genes required for TCE metabolism in this strain. A *Pseudomonas putida* endophyte strain effective at degrading PAH's, another

common class of environmental pollutants, reduced the phytotoxity of phenanthrene in inoculated trees. In addition to the endophytes capable of degrading organic pollutants, we isolated arsenic-tolerant endophyte strains that may reduce the phytotoxic effects of arsenic in host plants, and are investigating a potential role for extracellular polymeric substances (EPS) in the detoxification of arsenic by endophytes. Endophyte-assisted phytoremediation is a readily-deployable, important technological advance, enabling clean-up of a broad spectrum of contaminated areas.

W540: Genomics of Plant Development

Adventitious Rooting in Avocado - Unravelling the Bottleneck to Propagation

Madeleine Gleeson, **Alice Hayward**, Jayeni Hiti-Bandaralage, Stephen Fletcher and Neena Mitter, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia

Avocado (*Persea americana* Mill.) is a highly nutritious fruit of immense popularity. As orchards expand to meet consumer-driven increases in demand, the rapid propagation of elite cultivars is of the utmost importance. The clonal propagation of avocado rootstocks however, is a significant bottleneck to the industry due to the immense recalcitrance to adventitious root (AR) generation. This recalcitrance hinders both existing nursery-based propagation, in addition to the development of emerging tissue culture pipelines. We have observed that in tissue culture, shoots of some cultivars generate AR much more readily than those of others, including some of the more industrially relevant cultivars. To examine why one cultivar was AR responsive to a treatment in which the other cultivar was not, an RNAseq with qPCR validation was completed. Plant material was sequenced from tissue cultured shoots of an easy-to-root (AR+) and a difficult-to-root (AR-) cultivar 72 hours after treatment with either mock or root-inducing substrates. A selection of genes significantly differentially regulated between treatments and across cultivars were selected for validation by qPCR. These genes were also profiled across additional treatments and time points (0 hours and 24 hours) for each cultivar, to further dissect their relationship to rooting phenotype. The molecular profiles of these genes, many of which have not previously been implicated in AR, provide invaluable insight into avocado AR regulation and will guide improvements to rooting protocols into the future.

W541: Genomics of Plant Development

Arabidopsis miR858 Fine-Tunes the Expression of MYB83-Regulated Genes in Cyst Nematode Feeding Cells

Tarek Hewezi, University of Tennessee, Knoxville, TN, USA,, Knoxville, TN

MiRNA-targeted transcription factors have great potential to mediate adequate fine-tuning in cellular physiology and metabolism because of their capability to regulate the transcriptional activity of numerous genes. Here, we describe a novel control mechanism of the miR858-MYB83 regulatory module in finely balancing gene expression levels in the feeding site (syncytium) formed by the beet cyst nematode *Heterodera schachtii* in Arabidopsis roots. miR858 was found to post-transcriptionally silence MYB83 in the syncytium following BCN infection. Genetic manipulations of miR858 and MYB83 modulated Arabidopsis response to BCN infection in opposite direction. Transcriptome analysis and direct target identification revealed that MYB83 directly regulates the expression of a significant number of syncytial genes encoding key etiological factors vital for syncytium formation and function. MYB83 stabilizes its own transcript level through a feedback regulatory loop in which it positively activates the expression of its negative regulator, the miR858, to ensure optimal cellular activity during nematode parasitism of Arabidopsis.

W542: Genomics of Plant Development

Scaffolding Brassinosteroid Signaling Components at the Plasma Membrane by TTL Proteins

Miguel A. Botella, IHSM-University of Málaga-CSIC, Málaga, Spain

Brassinosteroids (BRs) form a group of steroidal hormones essential for plant growth, development and stress responses. BR perception at the plasma membrane initiates a series of phosphorylation events enabling the nuclear accumulation and activity of the key transcription factors BZR1 and BES1. We found that plant-specific Tetratricopeptide Thioredoxin-Like (TTL) proteins are positive regulators of BR signaling that function as scaffold for the BR signaling components in Arabidopsis. TTL3 associates with most core components involved in BR signaling, with the exception of the BAK1 co-receptor. TTL3 is mainly localized in the cytoplasm, and BR treatment increases its localization at the plasma membrane. In addition, the expression of *TTL3* strengthens the association of BR-signaling components BSK1 and BZR1 at the plasma membrane. Consistent with a role in BR signaling, mutations in *TTL3*, and related *TTL1* and *TTL4* genes cause reduced BR responses, and these defects that highly enhanced in a triple *ttl1 ttl3 ttl4* mutant. We propose a novel mechanistic model for BR signaling, in which cytoplasmic/nuclear BR components bound to TTL proteins are recruited to the plasma membrane upon BR perception, which in turn allows the assembly of a BR signaling complex with the goal of ensuring de-phosphorylation and nuclear accumulation of the transcription factors BZR1 and BES1. This novel TTL scaffold model for BR signaling resembles that of Wnt signaling in metazoans, in which TTL proteins would act similar to Axin1, optimizing signaling efficiency of the cascade by promoting the assembly of the signaling complex at the plasma membrane.

W543: Genomics of Plant Development

The Control of Developmental Phase Transitions in Tropical/Subtropical Tree Crops

Muhammad Umair Ahsan[1], Alice Hayward[1], Christine Beveridge[2] and Neena Mitter[1], (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia, (2)School of Biological Sciences Faculty of Science, The University of Queensland, St Lucia,, Brisbane, Australia

Unlike annual plants, horticultural tree crops are characterized by a long vegetative juvenile phase. This is an impediment to crop breeding and delays return-on-investment, as productivity depends on attaining reproductive phase. In the annual herb *Arabidopsis thaliana*, vegetative to reproductive phase transition is regulated by two microRNAs: miR156 and miR172. High abundance of miR156 is implicated in juvenility by downregulating *SQUAMOSA promoter binding protein-like (SPL)* genes as well as miR172, while miR172 promotes flowering by downregulating members of the *APETALA2-like* gene family. In horticultural trees, such as avocado (*Persea americana*), mango (*Mangifera indicia*) and macadamia *Macadamia integrifolia*, a possible role for these microRNAs in phase transition has not been comprehensively

investigated. In this study, the activity of miR156 and mir172, and their putative targets, was quantified in these tree species over development. Youngest fully expanded leaves were collected from seed-grown trees at distinct ages, from one month to >10 years old. The abundance of miR156 was highest in young seedlings of all species, and was downregulated at the mature/flowering stage. This downregulation occurred much earlier for macadamia and mango (within 6 months), than for avocado (>1.5 years). These results are consistent with findings in annual plants and suggests miR156 may play a role in vegetative identity in these crops. On the other hand, miR172 abundance remained largely unchanged over time, except in avocado where its expression was low in seedlings and high in mature/flowering trees, consistent with Arabidopsis. Target gene quantification (*SPL4, SPL9 and AP2)* revealed that *SPL4* was anticorrelated with miR156, while *SPL9* and *AP2* showed varied expression patterns with little anticorrelation to microRNAs. Taken together, this data suggests a conserved role for the miR156: *SPL4* module in phase transition between annual species and these horticultural tree crops.

W544: Genomics of Plant Development
Molecular Links between Stem Growth Habits and Responses to Water Stress in Soybean
Jianxin Ma, Purdue University, West Lafayette, IN
Soybean stem growth habit is a key adaptation and agronomic trait that directly affects plant height, flowering time and duration, node production, leaf morphology, root architecture, maturity, water use efficiency, abiotic stress tolerance, and, ultimately, soybean yield. Based on the timing of the termination of apical stem growth, most elite soybean cultivars are classified into three categories of stem architecture, commonly known as determinate, semi-determinate, and indeterminate types. Classical genetic analyses demonstrated that soybean stem growth habit was regulated by an epistatic interaction between two major genes *Dt1* and *Dt2*. We have recently cloned *Dt1*, a floral repressor that prevents terminal flowering to form indeterminate stems, and *Dt2*, a MADS-domain factor gene that directly represses the *Dt1* expression in shoot apical meristems to form semi-determinate stems. More recently, we have identified several Dt2 targets associated with responses to water stress and stomatal densities by a combination of RNA-seq and ChIP-seq analysis with the Dt2-specifc antibody, linking stem growth habit and flowering time to abiotic stress responses at the molecular level. Potential downstream implications of molecular links underlying these traits are under investigation. We anticipate our study will provide novel insights into the mechanisms modulating semi-determinacy and pave new strategies for optimizing plant architecture of soybean, and potentially other legume crops for enhanced environmental resilience and yield potential.

W545: Genomics of Tissue Regeneration in Plants and Animals
3D Platforms for Plant Regeneration - an Inspiration from Tissue Engineering
Maya M Kleiman, Agricultural Research Organization (Volcani Center), Rishon LeZion, Israel

W546: Genomics of Tissue Regeneration in Plants and Animals
Identifying the Genomic Basis of Adventitious Rooting in *Populus*
Steven H. Strauss[1], Christine Zawaski[1], Stephen DiFazio[1], Jonathan R Cumming[1], David Macaya-Sanz[1], Cathleen Ma[2], Jialin Yuan[2], Wellington Muchero[3], Fuxin Li[2], Yuan Jiang[2], Anna Carlina Magnuson[2] and Brett Pierce[2], (1)West Virginia University, Morgantown, WV, (2)Oregon State University, Corvallis, OR, (3)Oak Ridge National Laboratory, Oak Ridge, TN
Cuttings of black cottonwood (*Populus trichocarpa*) and a hybrid back-cross population ((*P. trichocarpa* x *P. deltoides*) x *P. deltoides*) were evaluated for adventitious rooting using GWAS and QTL analysis, respectively, in three experiments in Oregon (OSU) and West Virginia (WVU). Cuttings were rooted in water (OSU, WVU) or 0.5 mM $CaNO_3$ solution (WVU). At periodic intervals, root initiation and growth measurements were taken, and images were digitally analyzed using machine vision methods. Linear mixed models were used to estimate heritabilities and BLUPs of the phenotypes taking into account experimental design (blocking, replicates). Genome-wide efficient mixed model association (GEMMA), accounting for kinship, was used to correlate a panel of 13 million markers to phenotypic variation. For the backcross, R/QTL was used to derive a map with 3,800 SNP loci, and identify QTL intervals. Rooting traits had highly statistically significant variation, with heritabilities near 20% for the various measures. Significant associations among markers and phenotypic traits were found in all three experiments, with some associations shared among experiments. For example, a QTL for days to rooting on chromosome 6 was located in the same region as a significant hit for GWAS for total root length. By combining interspecifc QTL analysis with intraspecific GWAS, we gained additional insights into the genetic and physiological control of rooting in *Populus*.
We thank the NSF Plant Genome Research Program (IOS # 1546900) and the USDA National Institute for Food and Agriculture (2014-67013-21657).

W547: Genomics of Tissue Regeneration in Plants and Animals
Inducing Biotic Stress Resistance: Known Plant-Genomic Targets, and Future Research
Arye Harel, ARO Volcani Center, Rishon LeZiyyon, Israel

W548: GMOD, The Generic Model Organism Database Project
Tripal and its Benefits for Collaboration
Ethalinda Cannon, Iowa State University, Ames, IA
Tripal is a set of Drupal modules for the construction of biological database websites, using the Chado database schema as its backend. Because of its potential for customization, collaborative development and data sharing, Tripal is a key element of the Legume Federation. In this talk I will report on the use of Tripal within Legume Federation members, summarize ways that Tripal can be customized and expanded, and give an update on the anticipated release of Tripal 3.0.

W549: GMOD, The Generic Model Organism Database Project
Galaxy Community Update

Dave Clements, Johns Hopkins University, Eugene, OR
Galaxy is a widely used and deployed data integration and analysis platform for life science research (http://galaxyproject.org). This talk will briefly introduce the platform and then discuss usage, deployment options, and features with an emphasis on what's new in the past year and what's coming up this year. Recent work includes efforts focused on easier tool deployment, cloud support, and training materials.
For more on Galaxy, including an extended demonstration, please attend the *Galaxy: An Open Platform for Data Analysis and Integration* workshop on Tuesday from 4-6:10pm in the California Room.

W550: GMOD, The Generic Model Organism Database Project
G-OnRamp
Luke Sargent, Oregon Health & Science University, Portland, OR

W551: GMOD, The Generic Model Organism Database Project
Better JBrowse Vcf
Richard D Hayes, DOE Joint Genome Institute, Walnut Creek, CA

W552: GMOD, The Generic Model Organism Database Project
Integrating Biological Data for Easy Cross-Organism Queries
Joseph W. Carlson[1], Daniela Butano[2], Justin Clark-Casey[2], Sergio Contrino[2], Josh Heimbach[2], Rachel Lyne[2], Julie Sullivan[2], Yo Yehudi[2], David M. Goodstein[1] and Gos Micklem[2], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)Department of Genetics, University of Cambridge, Cambridge, United Kingdom
Strange file formats, missing columns, invisible whitespace and unintelligible filenames: every scientist who ever wrangles data has horror stories to tell. Time after time, different people download files, clean them up, and painstakingly integrate them with other sources - but we can do better than this.
InterMine is designed specifically to take away these pain points, by pre-integrating and tidying data sources, providing a single entry point for biologists and bioinformaticians, and enabling integrative analysis - a powerful approach in modern biology that allows knowledge from many sources of evidence to be analysed together. InterMine focuses on providing multiple easy ways to access data - you can query, analyse, and download data from a website, or write queries using client libraries in Python, R, Perl, Javascript, Java, or Ruby.
An important step in understanding biological processes is also the interpretation of complementary data from other organisms. There is a broad selection of InterMines worldwide, covering many organisms e.g. PhytoMine (92 plant genomes), the Legume federation InterMines (Chickpea, Soy, Legume, Peanut, Bean), MedicMine (Medicago), ThaleMine (Arabidopsis) as well as the budding yeast, rat, zebrafish, mouse and nematode model organisms. Scripts written for one InterMine can be re-used in other InterMines with little or no modification, providing a common interface across a vast range of organisms.
The InterMine registry, http://registry.intermine.org, lists available databases and more about InterMine can be found at http://www.intermine.org

W553: Graft Genetics and Genomics
The Perfect Graft: Parasitic Plants and the Role of RNA Exchange
James H. Westwood[1], So-Yon Park[1], Gunjune Kim[1], Saima Shahid[2], Nathan Johnson[2], Eric Kenneth Wafula[2], Feng Wang[2], Ceyda Coruh[2], Vivian Bernal-Galeano[1], Claude dePamphilis[2] and Michael Axtell[2], (1)Virginia Tech, Blacksburg, VA, (2)Penn State University, University Park, PA
The haustorial connection between parasitic plants and their hosts has been called “the perfect graft” because it forms a robust union between two different plants. Indeed, the parasite-host connection is so sophisticated that it allows the joining of widely differing species. It is interesting to consider whether the functional similarities of parasitism and grafting are based on shared mechanisms of interaction at the molecular level. We study parasitic plants of the genera *Phelipanche* and *Cuscuta*, with emphasis on understanding the extent and functional significance of nucleic acid exchange in the interaction. Both of these species form symplastic connections with host vascular tissues, but the haustoria of *Cuscuta* seem to be especially open to exchange of macromolecules. We have characterized the movement of mRNAs between hosts and *Cuscuta*, and found that large numbers of different transcripts are exchanged bidirectionally between the parasite and *Arabidopsis thaliana*. The potential role for mRNA exchange remains unproven, but may allow for information transfer or host manipulation by the parasite. Parasite-derived microRNAs (miRNAs) were also found to be mobile into host tissues, where they trigger generation of small interfering RNAs that appear to target specific host mRNAs for destruction. These miRNAs may thus function as virulence factors that facilitate parasitism. Taken together, we speculate that RNA trafficking is an important component of parasitism, and reflects a broad role for mobile RNAs in intact plants as well as grafted plants.

W554: Graft Genetics and Genomics
Graft Transmission of RNA Silencing for Crop Improvements: Virus Resistance and High-Temperature Tolerance
Masamichi Nishiguchi, Shinya Nakamura, Emran Md. Ali and Kappei Kobayashi, Ehime University, Matsuyama, Japan
RNA silencing is a nucleotide sequence specific gene regulation, which caused RNA degradation, translation inhibition or DNA methylation. One of the characteristics of this phenomenon in plants is graft transmission of RNA silencing between rootstocks and scions in both up and down directions. We have focused on grafting to confer useful functions such as biotic and abiotic stresses in non transgenic scions from transgenically silenced rootstocks. Here we present two cases: virus resistance and high-temperature tolerance in non-transgenic scions. *NtTOM1* and *NtTOM3* are the tobacco homologs of *Arabidopsis TOM1* involved in tobamovirus replication. Double RNA silenced transgenic tobacco lines of both *NtTOM1* and *NtTOM3* were produced and used for grafting as rootstocks. Non transgenic tobacco or tomato plants grafted onto these silenced rootstocks showed induction of RNA silencing of both genes and virus resistance. siRNAs of both genes were detected. In the next experiments we produced RNA silenced tomato plants of fatty acid desaturase gene *LeFAD7* and used as rootstocks.

Finally siRNA of the gene was detected in non transgenic tomato scions after grafting. These scions showed normal growth under high temperature conditions (40 ˚C) while control grafted tomato plants died. These results showed successful use of grafting for conferring virus resistance and high-temperature tolerance. This approach may pave the way for other traits and crops as far as grafting is available. This work was supported by the Program for Promotion of Basic and Applied Researches in Bio-oriented Industry of Japanese Government and other funds.

W555: Graft Genetics and Genomics

From Small RNA and Gene Expression Studies to the Discovery of a Novel Class of Antimicrobial Peptides for Controlling Citrus Honglongbing

Hailing Jin, University of California, Riverside, CA

W556: Graft Genetics and Genomics

Opportunities for Crop Improvement by Inter-Cellular Movement of Organelles through Graft Junctions

Csanad Gurdon, Rutgers University, Department of Plant Biology, New Brunswick, NJ, Zora Svab, Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ and Pal Maliga, Rutgers University, Piscataway, NJ

We report on cell-to-cell movement of plastids and mitochondria through graft junctions of Nicotiana species. Selection for organelle movement is achieved by grafting a plant that carries a selectable nuclear marker to another that has a selectable plastid marker. Graft junctions are then sliced, and the tissue slices selected in culture for both the nuclear and the plastid markers. Plastid movement in the absence of any input of mitochondria or chromosomal DNA is commonly recovered after selection. Such one-step substitution of chloroplasts avoids the need for repeated back-crosses for cytoplasmic substitution, and enables separation of plastid and mitochondrial traits. Thus, graft transmission enables utilization of engineered chloroplasts in sexually incompatible, but graft compatible species such as tobacco, tomato and potato. In contrast to plastids, when mitochondria move, they always fuse, and their mitochondrial genomes recombine, allowing the transmission of mitochondrial traits such as cytoplasmic male sterility. Because grafting is feasible between sexually incompatible species, mitochondrial movement, fusion and recombination may be exploited to create cytoplasmic male sterile forms of crops such as tomato, where the trait currently does not exist.

W557: Graft Genetics and Genomics

Genome-Wide Analysis of Small RNA and mRNA Expression Profiles during Fruit Development in Grafted Citrus

Rachel J. Rattner, University of California, Riverside, Riverside, CA

Citrus, one of the most economically important fruit crops in the world, is commercially propagated through grafting. Varying scion-rootstock combinations causes substantial effects on trees that often influence yield and fruit quality traits. Presently, the explanation for these differences has not been extensively studied at the molecular level. One potential reason for rootstock effects on fruit quality is the presence of small RNAs, molecules known to affect gene expression and plant development. Species-specific small RNAs have been discovered in a wide variety of plants. Many are differentially expressed and can cross a graft union to influence scion traits.

We hypothesize that grafting diverse rootstocks influences small RNAs populations in citrus, which can greatly impact the quality of fruit produced. In this study, fruit and root samples were collected from sweet orange scions grafted onto four genetically differing rootstocks (trifoliate orange, Carrizo citrange, rough lemon, and sweet orange). Three replicate samples of each genotype were collected from a Washington navel orange rootstock trial planted in Riverside, CA at four time points throughout the growing season. Using mRNA-seq and small RNA-seq, an integrative analysis of microRNA and mRNA expression profiles was performed. Differentially expressed mRNAs and microRNAs were identified between the different rootstocks throughout fruit development. Furthermore, ~15% of small RNA reads sequenced matched known, conserved microRNAs, leaving many reads to be assessed as potentially novel, species-specific small RNAs. The sequencing results will be correlated to phenotypic data to identify small RNAs likely influencing changes in fruit quality.

W558: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities

Structural Annotation of Genes and Transposable Elements in Maize using Computational and Manual Validation

Michelle C. Stitzer, University of California, Davis, Davis, CA

Transposable elements (TEs) contribute the majority of DNA to most plant genomes, and their relentless activity moving to new positions in a genome generates genetic variation. TEs have incontrovertibly succeeded in generating this variation in the maize (*Zea mays* subsp. *mays*) genome, where more than 85% of genomic sequence can be ascribed to past transposition. TEs have been implicated in dramatic phenotypic changes, but in genome-wide analyses, they are often represented as fragmentary relics of past transposition, which obscures their evolutionary history and interaction with their host genome. Using an updated TE annotation of the maize inbred line B73 that identifies intact TE copies, we investigate the family-level ecological and evolutionary dynamics of TEs in maize, approximating every TE family as a species of TE. Integrating a variety of data, from descriptors of each TE like coding capacity, expression, and methylation, as well as features of the region of the genome in which it is found, we use a machine learning framework to model the relationship between these features of the genomic environment and TE survival. These analyses reveal that individual TE families persist in the genome using dramatically different strategies, and a multitude of combinations of these features can determine and characterize the ecological niche of a TE family. We conclude that while the impact of transposition is highly family- and context-dependent, a family- level understanding of the ecology of TEs in the genome will refine prediction of the role TEs play in generating genetic and phenotypic diversity.

W559: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities

Transcriptome Analysis of Rice Leaf Sheath Blight Response

Noor Al-Bader[1], Austin Meier[1], Matthew Geniza[1], Yamid Sanabria[2], James Oard[2] and Pankaj Jaiswal[3], (1)Oregon State University, Corvallis, OR, (2)Louisiana State University Agricultural Center, Baton Rouge, LA, (3)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR

In rice, *Rhizoctonia solanii* causes leaf sheath blight (SB) disease, which leads to widespread decline in yield. Almost all of the US rice varieties are susceptible. Therefore, we carried out an RNA-seq based transcriptome study on two rice varieties, Cocodrie (CCDR; rice sheath blight susceptible) and MCR10277 (MCR; rice sheath blight resistant), to profile the time-series, genome-scale wide transcriptional differences in response to the *R. solanii* infection.
In-house bioinformatics workflows, was used to identify variety-specific SNPs, including indels, and pairwise comparison between time points per variety for differential gene expression (DGE). During this process, tools and data resources hosted on the Gramene Database, from achieving the raw data on ArrayExpress to inferring biological relevance of DEG and SNP analyses using Gramene's genome browser, tools such as the Variant Effect Predictor (VEP), and Plant Reactome, were utilized. Our time-course analysis, by cross-referencing DGE with significant variants, yielded 4 genes of interest within the major QTL region on chromosome 9 that plat a role in the resistant rice's physiological response to SB.

W560: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities
Looking for Insights across Genomes: Searching and Visualizing Data in Gramene
Andrew Olson, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
The Ensembl Compara gene tree pipeline reconstructs the evolutionary history of protein coding gene families, documenting speciation and duplication events in ancestral genomes. This foundational resource has been used by the Gramene project (www.gramene.org) to identify gene annotation errors (split gene models), project curated pathways from rice to other plant species, and suggest functional annotation for less well annotated orthologs.
Gramene's search engine (data.gramene.org) indexes gene models together with their associated ontology terms, InterPro domains, pathways and homology relationships. The search interface result list displays the closest model species homolog for gene models that lack a meaningful gene name or description. When a search only matches genes in well annotated model organisms, the homology pane can be used to modify the search to show homologs.
The homology pane contains an integrated visualization component which displays the gene tree expanded to show the gene of interest and the closest well annotated homolog. Additional information is displayed as a track next to each leaf node or collapsed subtree. The default display mode shows the regions of proteins participating in the multiple sequence alignment color-coded by InterPro domain. Users can navigate the view or switch to a scrollable multiple sequence alignment mode. A new neighborhood conservation view has been added which shows local syntenic relationships across the gene family. A gene expression display mode is under development which would show expression profiles for comparable tissues/conditions.
Navigating a gene tree with hundreds of genes from dozens of species is overwhelming and unnecessary for most users. To cope with large gene families, especially as Gramene plans to more than double the number of hosted reference genomes, we have implemented a filter to limit search results and prune species trees and gene trees on the fly to only display the genes from a user defined subset of genomes. The gene tree pruning code also removes excess gaps from the multiple sequence alignment which result from removing less closely related species.
These developments add value beyond the standard Ensembl views by integrating diverse data into a fast gene tree interface while giving scientists the flexibility to focus on the species most relevant to their research.
All code developed for the web service and the search interface are maintained on GitHub (https://github.com/warelab)
Gramene is supported by NSF grant IOS-1127112, and partially from USDA-ARS (1907-21000-030-00D).

W561: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities
Integrating and Displaying Plant Gene Expression in Expression Atlas
Laura Huerta, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom
Expression Atlas (https://www.ebi.ac.uk/gxa) is a database and web-service at EMBL-EBI that selects, curates, re-analyses and displays gene expression data in a baseline context, e.g. to find genes expressed in different tissues in potato, and in a differential context, e.g. to find up-regulated genes in response to stripe rust and powdery mildew in wheat. Plant experiments from ArrayExpress, GEO and SRA/ENA/DDBJ are selected for curation and analysis. Data curation involves enriching sample annotation with additional metadata, annotating metadata with Experimental Factor Ontology (EFO) terms and deciding comparisons for differential expression analysis based on associated publications and correspondence with the original researchers. Data analysis is performed using open source tools for microarray data and our standardized pipeline iRAP (https://github.com/nunofonseca/irap) for RNA-seq data. Currently, we provide gene expression analysis results for more than 700 plant experiments across 20 different plant species. Expression Atlas can be searched by gene, gene set and biological condition queries. The use of EFO annotations allows efficient search via ontology-driven query expansion and facilitates data integration across multiple experiments. We offer downstream analysis and visualization such as gene co-expression, biological variation among replicates, transcript quantification, visualization of gene expression in Gramene genome browser and enrichment of Gene Ontology terms and Reactome pathways. Finally, we have developed an automatic pipeline that discovers new plant RNA-seq data at ENA for 45 different species, performs quality control, alignment to the genome reference in Ensembl plants and quantification of gene and exon expression. The analysis results are available via our RNASeq-er API (https://www.ebi.ac.uk/fg/rnaseq/api/).

W562: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities
Scaling Comparative Analysis across the Taxonomic Space
Paul J. Kersey, EMBL - The European Bioinformatics Institute, Cambridge, United Kingdom
The Ensembl Compara suite of comparative analysis tools is used across the taxonomy, including for plant species, where it generates data made available through the collaborating web portals of Ensembl Plants and Gramene. Among the outputs are gene trees, implied evolutionary histories for every gene family. As the number of known genomes and gene models increases, the computational cost of running these pipelines grows, and it becomes harder to perform broad-range analyses within reasonable timeframes. Therefore, we are developing a new pipeline to improve the speed and range of our collaborative analysis, using Hidden Markov Models (HMMs) to identify members of gene families before attempting to constructing the tree. The benefits are twofold: firstly, the same set of models can be built outside of the release schedule, and applied to multiple releases before periodic update; and secondly, that one can identify higher-level families of well-conserved proteins that are

found across the taxonomy (i.e. match plant families to their corresponding families in fungi or vertebrates). Rather than develop a completely novel set of HMMs, our approach involves using the HMMs that identify proteins as members of PANTHER families, supplementing these by additional HMMs to capture families present in Ensembl gene sets but not yet matched by PANTHER.
We hope to put the new pipeline into production in 2018. In this talk, I will present preliminary data and compare the results with the current version.

W563: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities
Functional Curation - How Do We Prioritize and Scale up?
Pankaj Jaiswal, Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR
We will showcase an ongoing project, domain informational vocabulary extraction (DIVE). This project developed in collaboration with the Gramene database, American Society of Plant Biologists (ASPB), and CyVerse, aims to facilitate data-mining and enrichment of digital publications through entity and key informational words detection including gene and phenotype annotations. The DIVE implements strategies for biological entity detection via natural language processing including using regular expression rules, ontologies, keywords dictionary and supervised learning approach. The results of the biological entities are then stored in a database and accessible through an interactive web application for curation and evaluation by experts and publication authors. As part of the submitted peer-reviewed scientific manuscript submission and approval process, integrated with DIVE, in a web interface, a user can make additional annotations and corrections to the current results. The updates are then shared with the publishers to improve the entity detection in subsequent processed articles. The system can be beneficial to publishers and various domain curators, and researchers at large on making first pass natural language processing by DIVE to build curation workflows and set priorities. The project is supported by the Gramene database award (NSF IOS-1127112) and the CyVerse.

W564: Gramene: Unifying Comparative Genomics and Pathway Resources for Plant Communities
Gramene: Progress and Future Plans
Marcela Karey Tello-Ruiz, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
The Gramene database (http://www.gramene.org) is an integrated resource for comparative genomics, phylogenetic and pathway analysis in plants. The database provides agricultural researchers and plant breeders with valuable biological information on genomes and plant pathways of numerous crops and model species, thus enabling powerful comparisons across species. During this talk, I will provide a brief overview of the data, as well as analysis and visualization tools at Gramene, as well as future perspectives. The current data release of Gramene includes the reference genomes and corresponding gene annotations for 53 plant species including rice, maize, wheat, barley, soybean, Arabidopsis, Brassicas, medicago, cassava, Guinea yam, white jute, tomato, potato, beet, banana, cocoa, peach, grapevine, poplar, red clover, sunflower, Amborella, spikemoss and algae, as well as over 265 curated Japonica rice pathways, and orthology-based pathway projections for 75 crops and plant models, including the above 53, wild peanut ancestors, and the common bean.

W565: Grape Genome Initiative
Insights into the Evolution of *Vitis vinifera*
Brandon S. Gaut[1], Yongfeng Zhou[1], Melanie Massonnet[2], Andrea Minio[2] and Dario Cantu[2], (1)University of California, Irvine, Irvine, CA, (2)University of California Davis, Davis, CA
We have been studying the evolutionary genetics of grapes (*Vitis vinifera* ssp. *vinifera*) to answer fundamental questions about the history of grapes and, more generally, about the processes that shape genetic diversity in domesticated plants. With SNPs from genome resequencing data, we have used population genetic approaches to address questions about the duration of domestication, the genes that contribute to domestication phenotypes, and apparent costs of domestication. We have found, for example, that grape genomes contain 5.2% more predicted deleterious variants than wild individuals, and that these deleterious variant have largely accumulated in a heterozygous state, probably due to clonal propagation and the accumulation of recessive deleterious mutations. We are also studying structural variants (SVs). In most crops, SVs remain an understudied component of genetic diversity, even though they contribute substantively to crop adaptation. To study the evolutionary dynamics of SVs, we are comparing two Pacbio genome assemblies, and using this baseline to extrapolate our results to the analysis of SV in population genomic data. Preliminary analyses suggest that SVs are deleterious globally but may have been causal for important domestication phenotypes such as the shift to hermaphroditic mating system.

W566: Grape Genome Initiative
Root Transcriptomic Responses of Grafted Grapevines to Heterogeneous Nitrogen Availability Depend on Rootstock Genotype
Noe D. Cochetel[1], Frederic Escudie[2], Sarah Jane Cookson[3], Zhanwu Dai[3], Philippe Vivin[3], Pierre-François Bert[3], Mindy Stephania Munoz[4], Serge Delrot[3], Christophe Klopp[2], Nathalie Ollat[3] and Virginie Lauvergeat[3], (1)Dept of Biochemistry & Molecular Biology, University of Nevada, Reno, Reno, NV, (2)Genotoul Bioinformatics Platform, UR875, INRA, France, (3)EGFV, Bordeaux Sciences Agro, INRA, University of Bordeaux, France, (4)Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Chile
In many fruit species, including grapevine, grafting is used to improve scion productivity and quality and to adapt the plant to environmental conditions. However, the mechanisms underlying the rootstock control of scion development are still poorly understood. The ability of rootstocks to regulate nitrogen uptake and assimilation may contribute to this control. A split-root system was used to grow heterografted grapevines and to investigate the molecular responses to changes in nitrate availability of two rootstocks known to affect scion growth differently. Transcriptome profiling by RNA sequencing was performed on root samples collected 3 and 24 h after nitrogen supply. The results demonstrated a common response involving nitrogen-related genes, as well as a more pronounced transcriptomic reprogramming in the genotype conferring the lower scion growth. A weighted gene co-expression network analysis allowed the identification of co-regulated gene modules, suggesting a role for nitrate transporter 2 family genes and some transcription factors as main actors controlling this genotype-dependent response to heterogeneous nitrogen supply. The relationship between nitrate, ethylene, and strigolactone hormonal pathways was

found to differ between the two genotypes. These findings indicated that the genotypes responded differently to heterogeneous nitrogen availability, and this may contribute to their contrasting effect on scion growth.

W567: Grape Genome Initiative

Developing Grapevine FAIR Data

Cyril Pommier[1], Eric Duchêne[2], Thierry Lacombe[3], Timothee Flutre[4], Michael Alaux[1], Guillaume Cornut[5], Florian Philippe[5], Sophie Durand[1], Raphael Flores[1], Erik Kimmel[1], Thomas Letellier[1], Nacer Mohellibi[1], Célia Michotey[1], Hadi Quesneville[1] and **Anne-Francoise Adam-Blondon**[1], (1)URGI, INRA, Université Paris-Saclay, Versailles, France, (2)Université de Strasbourg, INRA, SVQV UMR-A 1131, Colmar, France, (3)UMR 1334 AGAP, INRA, Montpellier, France, (4)INRA UMR 1334 AGAP Amélioration génétique et adaptation des plantes méditerranéennes, Montpellier, France, (5)Unit of Research in Genomic-Info (URGI), INRA, Université Paris-Saclay, VERSAILLES, France

URGI-INRA has been involved in the last 6 years in several international projects (e.g. EU FP7 TransPLANT ; EU H2020 ELIXIR-Excelerate) and initiatives (Wheat Initiative, Research Data Alliance, Breeding API) contributing to the development of (i) community recommendations for data standardisation (e.g. wheat guidelines; http://www.wheatis.org/DataStandards), (ii) standard Application Programming Interfaces for plant genetics and breeding (Breeding API ; www.brapi.org), (iii) data standards for phenotyping data (MIAPPE ; www.miappe.org), (iv) crop specific ontologies in the frame of the Crop Ontology (www.cropontology.org). All these activities are aiming at creating a framework contributing to the development of Findable, Accessible, Interoperable and Reusable (FAIR) data for plant biology, genetics and breeding. A vision paper has recently been produced in the frame of the International Grapevine Genome Initiative (Adam-Blondon et al, 2016, doi:10.1038/hortres.2016.56) that discusses how to best move towards FAIR data for grapevine research. The french grapevine community is now actively organizing a sustainable implementation of the resulting guidelines using as local community levers three important resources, the INRA information system for crop genetics and genomics data (GnpIS), the Crop Ontology portal and the French Grapevine National Biological Resource Center that manages the french grapevine germplasm collection.

W568: Grape Genome Initiative

A Phased, Chromosome-Scale, Diploid Genome Reference for the Wine Grape Cultivar Cabernet Sauvignon

Dario Cantu[1], Andrea Minio[1], Rosa Figueroa-Balderas[1], Grant R. Cramer[2], Massimo Delledonne[3], Melanie Massonnet[1] and Amanda Vondras[1], (1)University of California Davis, Davis, CA, (2)Dept of Biochemistry & Molecular Biology, University of Nevada, Reno, Reno, NV, (3)University of Verona, Verona, Italy

This presentation will describe an update on the assembly and annotation of a diploid genome reference for the wine grape cultivar Cabernet Sauvignon. The initial assembly of single-molecule real time sequencing reads (147X coverage; Pacific Biosciences) was carried out using FALCON-unzip. Contigs (N50=2.17Mb; N90=0.42Mb) were scaffolded using HiC libraries and the HiRise algorithm developed by Dovetail. This resulted in ~4 fold-improvement in contiguity (N50=9.4Mbp; N90=0.42Mbp). The assemblies were further scaffolded using SSPACE and gaps were filled with PBJelly. Scaffolds were then manually curated to remove primary assembly redundancy due to the retention of divergent homologous loci. These steps greatly reduced assembly fragmentation (N50=11.8Mbp; N90=2.8Mbp). Analysis of the assembly revealed numerous haplotype-switches within contigs and between scaffolded contigs. To improve phasing and increase contiguity, we built a 300X coverage optical map (Bionano Genomics) and incorporated it into the assembly bringing the final scaffolding to a primary phased assembly of 443Mb in 56 sequences (N50=16.5Mb; N90=6.9Mb) and a secondary phased assembly representing all alternative haplotypes of 330Mb in 33 sequences (N50=6.9Mb; N90=6.1Mb). Both primary and secondary assemblies were sorted and grouped in 19 chromosome-scale pseudomolecules based on collinearity with PN40024. Gene models were annotated using *ab initio* and evidence-based prediction methods combined with full length cDNA sequencing. A total of 29,294 and 16,806 protein coding-genes were identified in the primary and secondary assemblies, respectively.

W569: Grape Genome Initiative

Genomic and Transcriptomic Tools to Develop New Insight on the Molecular Bases of Inter-Varietal Phenotypic Differences and New Tools for Breeding

Maria Francesca Cardone[1], Carlo Bergamini[1], Rocco Perniola[1], Fiammetta Alagna[1], Lucia Rosaria Forleo[1], Donato Antonacci[1], **Riccardo Velasco**[1] and CREA-VE , (1)CREA Research Centre for Viticulture and Enology, Turi, Italy

Grapevine is one of the most important crop plants in the world because of its economically valuable role in fruit and wine production. Recently there was great expansion of genomics resources about grapevine genome, thus providing increasing efforts for molecular breeding. Current cultivars display a great level of inter-specific differentiation that needs to be investigated at molecular level to find responsible genes selected by cross breeding programs.

In the present work, we present the recent advancement of our research team in develop genomic tools for the molecular breeding in *Vitis vinifera* L.

Recently we combined high throughput technologies, to reveal the first inter-varietal atlas of structural variation and single nucleotide variants for the grapevine genome. We detected roughly 8% of the grapevine genome affected by genomic variations, thus demonstrating their importance in shaping the grapevine genome. Taken into account phenotypic differences existing among the studied varieties, we were able to identify genes showing differences in copy number correlated to different phenotypes and thus putative functional candidates for important traits in grapevine cultivation.

The identification of candidate genes represent the first step forward the definition of molecular markers to be applied in the marker assisted selection (MAS). In this context, additional association studies, transcriptomic and functional analysis are in progress to demonstrate the role of candidate genes in important traits. These studies represent a landmark for future studies aiming and to develop new tools to be used in association with new biotechnologies to support the molecular breeding in *Vitis vinifera*.

W570: Grape Genome Initiative

Genomic Basis of Clonal Variation in Cabernet Sauvignon

Patricia Agudelo-Romero and **Michael J Considine**, University of Western Australia, Crawley, Australia

The major winegrape varieties have been established for several centuries, with clonal propagation used for replication and expansion of vineyard areas. It is widely known that somatic mutations can arise through these practices, and the extensive history of propagation results in the accumulation of mutations through recurrent selection. Indeed, clonal selection programs have existed in several countries for decades. However, there is little knowledge of the underlying genomic mechanism of clonal variation. More importantly, the grapevine industries lack the tools to identify or verify clones, notwithstanding their value. Here, we sequenced the genomes of 16 individuals (PE150), representing 10 clones plus replication, with a varying confidence in their identity. This provided a design that reflects the current state of the industry, while also capturing commercially important clones. Although at the time of writing, data are still being analysed, we have identified and begun to validate a number of SNPs, and have established pipelines to explore larger scale transpositions, as well as viral integration into the genome. We are also seeking to identify regions that may uniquely mark the Cabernet Sauvignon genome. The outcome of these studies will be a set of valuable tools for the genetic fidelity of vineyards, and knowledge to underpin future research to understand relationships from genes to wine.

W571: Grasslands (Lolium Genome Initiative)
***Lolium perenne* Genome Sequencing**

Torben Asp, Aarhus University, Slagelse, Denmark

W572: Grasslands (Lolium Genome Initiative)
***Lolium perenne* Genome Sequencing Bioinformatics**

Narcis Fernandez-Fuentes, Aberystwyth University, Aberystwyth, United Kingdom

W573: Grasslands (Lolium Genome Initiative)
Pooling Resources: Allele Frequency Fingerprinting in *Lolium perenne*

Tom Ruttink, ILVO Plant Sciences Unit, Melle, Belgium

Allele frequency fingerprinting of heterogeneous plant populations of outbreeding species can be used for variety identification, association mapping, genomic selection or characterization of genetic resources. In the FACCE-JPI GrassLandscape project, we empirically validated a pool-GBS method for Genome-Wide Allele Frequency Fingerprinting (GWAFF). As pool-GBS cannot be targeted to predefined loci such as candidate genes, we integrated it with targeted resequencing using a highly multiplexed amplicon sequencing strategy to measure allele frequencies (pool-HiPlex). A pool-HiPlex assay was designed that amplifies 185 amplicons in 41 *L. perenne* genes in just two parallel PCR-reactions. We validated pool-GBS and pool-HiPlex using pools of 48 individuals, chosen to represent a wide range of genetic diversity in *L. perenne*, and addressed completeness, reproducibility and accuracy of allele frequencies with >1000 HiPlex SNPs and >150.000 GBS SNPs. We consistently found high correlations between allele frequencies obtained by genotyping individual plants and pool genotyping on leaf tissue pools and DNA extract pools. We also analyzed the error introduced at various steps of the protocol such as weighing, DNA-quantification, pooling, ligation and/or PCR-amplification. Applying a minor allele frequency threshold of 5% or 3% effectively removed non-reproducible SNPs in pool-GBS and pool-HiPlex, respectively. Allele frequency spectra could be obtained for single SNPs as well as for haplotypes spanning neighboring SNPs using read-backed phasing. Application of this methodology to a set of 470 natural populations of *L. perenne* sampled across Europe and the fertile Crescent revealed a geographical pattern of genetic differentiation in this species.

W574: Grasslands (Lolium Genome Initiative)
Quaternary Climate Changes Explain the Current Genetic Variation in a Major European Grassland Species, *Lolium perenne*

José L. Blanco-Pastor[1], Stèphanie Manel[2], Philippe Barre[1], Anna M. Roschanski[3], Evelin Willner[3], Klaus J. Dehmer[3], Matthew Hegarty[4], Hilde Muylle[5], Tom Ruttink[5], Isabel Roldán-Ruiz[5], Thomas Ledauphin[1] and Jean-Paul Sampoux[1], (1)INRA, Lusignan, France, (2)EPHE - CNRS, Montpellier, France, (3)IPK, Gatersleben, Germany, (4)IBERS, Ceredigion, United Kingdom, (5)ILVO, Melle, Belgium

Grasslands have been pivotal in the development of domestic herbivore breeding since the Neolithic and are nowadays the most widespread agricultural land-use over Europe. We have demonstrated that the genealogy of natural *Lolium perenne* (perennial ryegrass), from permanent grasslands across Europe and the Fertile Crescent, can be traced back to the pre-Holocene epoch (> 12 kya). Demographic events in this species, as reconstructed from genetic data, can be explained by the changing climatic conditions during Quaternary glaciations in Europe. Specific population splits and admixtures fit historical sea level changes across the Mediterranean basin. According to our estimations, population expansion across continental Europe took place during the Würm glaciation (12-110 kya), a cooling period that decreased the dominance of trees in favor of herbs, and notably grasses. The development of agriculture and herbivore breeding that started in the late Holocene (3.5 kya) and caused an increase in the abundance of grasses across Europe did not impact the genetic structure of *Lolium perenne* at continental scale. Furthermore, our results reveal little differentiation between cultivars released by contemporary plant breeding and natural strains of *L. perenne*, pointing to the very recent start of intensive human directional selection in herbage grasses. We expose that natural grasslands harbor wide genetic variability at continental scale that has enabled grasses to survive extreme climatic changes across millennia. This genetic variability has been underused by recent breeding activities and constitute valuable standing genetic variation for future adaptation of grasslands to Climate Change, safeguarding the agricultural services they provide.

W575: Grasslands (Lolium Genome Initiative)
Systems Biology of Grass-Endophyte Symbiota

Tim Sawbridge[1,2], Ian Tannenbaum[2], Tongda Li[2], Jatinder Kaur[1] and Ross Mann[1], (1)Department of Economic Development, Jobs, Transport and Resources, Bundoora, Australia, (2)La Trobe University, Bundoora, Australia

Perennial ryegrass has a well-known, and commercial exploited symbiotic relationship with the fungus *Epichloe festucae* var *lolii*. This interaction has been studied in our lab for decades. More recently research has begun on bacterial members of the symbiome. Over 300 bacteria have been isolated from surface sterilised seed from the commercial cultivar Alto, grown 7 days on sterile filter paper. Sixty-five of these bacteria have been genome sequenced, identifying 40 different strains. Some of these strains have been used for plate based bacterial phenotyping assays. Results from these assays will be presented. In parallel, pooled seedling samples were used for 16S v4 amplicon sequencing. These amplicons have been compared to genome sequences from the isolated bacteria to assess the culturability of seedling microbiome. A subset of the surviving seedlings were planted into two different growth matrixes, sand and potting mix, and allowed to grow for 3 months. A further collection of ~ 300 bacteria were isolated from these plants, and 16S v4 amplicon assays were run again. Results comparing the initial seedling microbiome, and that seen in the mature plants will be presented. In addition, individual seedlings containing different endophyte combinations were assessed by 16S v4 amplicon sequencing. This has identified a core set of OTUs seen in these seedlings. A subset of the initial plants have now flowered and seed has been collected to assess the seed transmissibility of members of the bacterial community.

W576: Host-Microbe Interactions
Insights into Virulence Evolution from 31 Phytophthora Plant Pathogen Genomes produced by Illumina and PacBio Sequencing
Brent Kronmiller, Oregon State University, Corvallis, OR

W577: Host-Microbe Interactions
Host Genetic Background Drives Developmental Trajectories in Hosts with Disrupted Microbiota
Kathryn CA Milligan-Myhre, University of Alaska Anchorage, Anchorage, AK
Resident microbiota play an important role in host health, and have complex relationships with their host. Host microbe interactions can be influenced by several factors, including the immune response to microbes and host genetic background. To determine what role the host genetic background plays in host-microbe interactions, we adapted well developed evolution animal model, threespine stickleback (Gasterosteus aculeatus) as a model host for host-microbe interaction studies. Threespine stickleback are well characterized, have an annotate sequenced genome, and, important for our studies, have undergone parallel evolution as ancestral oceanic populations have adapted to freshwater environments around the northern hemisphere. In Alaska we have ready access, and sample, populations of stickleback that have evolved in hundreds of different microbial, abiotic, and macrobiotic environments. We hypothesize that these populations have evolved different relationships with their microbiota. To test this hypothesis, we have examined the immune response and somatic development of several populations raised in the presence of complex microbial communities, absence of microbes, or with a microbiota challenged with antibiotics. We have found that stickleback that have evolved in different environments indeed have different immune responses and developmental trajectories when their microbiota is disrupted. We are now expanding our studies to new populations, developing new assays to measure developmental changes, and determining the role of stress in this delicate balance between the host and its microbial community.

W578: Hybridization, heterosis and balancing selection
Unlocking Genetic Basis of Complex Traits and Heterosis in Rice
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
Most of agronomic traits, which are called complex traits, are usually controlled by multiple genes and affected by various environmental conditions. Although a lot of quantitative trait locus (QTL) and genes related to rice complex traits have been cloned and functionally characterized, genetic basis and regulatory mechanisms underlying these complex traits are still unclear. We have implemented an integrated approach of genome-wide association study (GWAS) and phenomics with functional analysis to catch up agronomic trait genes or QTLs in a diverse cultivated rice population. This approach informs that the associated loci with the agronomic traits such as panicle length, grain sizes, grain weight and grain filling rate can be further characterized through expressional profiling, in-depth genome analysis, transgenic study, genome editing, and population genetic analysis. We believe that allelic genetic variations responsible for complex traits can be effectively explored.
Exploitation of heterosis is one of the most important applications of genetics in agriculture. However, the genetic mechanisms of heterosis are only partly understood, and a global view of heterosis from a representative number of hybrid combinations is lacking. We have developed an integrated genomic and forward genetic approach to construct a genome map for elite hybrid rice varieties and their inbred parental lines. We identified that the accumulation of numerous rare superior alleles with positive dominance is an important contributor to the heterotic phenomena. We have further done large-scale genomic mapping for yield related traits and heterotic effects by analyzing over 10,000 rice lines produced from 17 elite rice lines. The large data of genomics and phenomics from the well-designed populations enabled us, for the first time, to identify the genetic contributors and find out the exact causes of heterosis using “a composite interval-mapping method”. For the individual yield components, the heterozygous state of the heterosis-related genes generally acted through the way of dominance complementation. Taking all the components into account, the hybrids with yield heterosis resulted from an optimal combination of multiple yield-related components, meaning better performance of overall yield in crop productions. These results inform on the genomic architecture of heterosis for yield traits in rice, which will be useful information for crop improvement program.

W579: Hybridization, heterosis and balancing selection
A Tailored Quantitative Genetic Framework Reveals the Important Role of Epistatic Effects for Grain Yield Heterosis
Yong Jiang, Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, Germany
Increasing wheat yield is a key global challenge to produce sufficient food for a growing human population. Wheat grain yield can be boosted by exploiting heterosis, the superior performance of hybrids over the midparent values. Here we present a tailored quantitative genetic framework to study the genetic basis of midparent heterosis in hybrid populations based on crosses among diverse parents and applied it to a

particularly extensive dataset assembled for winter wheat. Grain yield was assessed for 1,604 hybrids and their 135 parental elite breeding lines in 11 environments. The hybrids outperformed on average the midparent values by 10%. This equals approximately fifteen years of breeding progress in wheat, thus further substantiating the remarkable potential of hybrid wheat breeding. Genome-wide prediction and association mapping implemented based on the developed quantitative genetic framework revealed that dominance effects played a less prominent role than epistatic effects for grain yield heterosis in wheat.

W580: Hybridization, heterosis and balancing selection

Phenotypic Effects of Deleterious Alleles and their Contributions to Heterosis in Maize

Jinliang Yang, Department of Plant Sciences, Davis, CA, Edward S. Buckler, USDA-ARS-Cornell University, Ithaca, NY, Michael D. McMullen, USDA-ARS, Columbia, MO, Rita H. Mumm, University of Illinois, Urbana, IL and Jeffrey Ross-Ibarra, University of California, Davis, CA

W581: Hybridization, heterosis and balancing selection

Hybrid Wheat from a Practical Breeder's Perspective

P. Stephen Baenziger[1], Amanda Easterly[1], Nicholas Garst[1], Anil Adhikari[2], Geraldine Opena[2], **Vikas Belamkar**[1], Amir Ibrahim[2], Jackie C Rudd[3], Bhoja R Basnet[4], Friedrich Longin[5], Jochen Christoph Reif[6], Jesse Poland[7] and Jean-Benoit Sarazin[8], (1)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, (2)Texas A&M University, College Station, TX, (3)Texas A&M AgriLife Research, Amarillo, TX, (4)CIMMYT, Col. El Batan, Texcoco, Mexico, (5)University of Hohenheim, Stuttgart, Germany, (6)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, (7)Department of Plant Pathology, Kansas State University, Manhattan, KS, (8)SAATEN-UNION RECHERCHE, Estrées-Saint-Denis, France

Hybrid wheat has proven to be elusive. From its previous highpoint in the 1980s, currently, only a few companies are producing hybrids in Europe, India, and South Africa. However, interest in hybrids has recently increased due to the need for greater and more efficient production to meet the projected future needs coupled with the availability of advanced breeding tools and insights. While most hybrid wheat discussions concentrate on the theoretical or genetic aspects, this talk will present how an applied wheat breeder is trying to make hybrid wheat a reality. Hybrid wheat requires: converting a self-pollinated crop into a cross-pollinated crop, the ability to make experimental hybrids, the use of genome-wide molecular markers and theory to develop genome-based high-yielding heterotic groups and patterns augmented by improved crossing block designs such as balanced incomplete factorial, and a path to commercial hybrid production. Preliminary results indicate that pollinator lines with good pollen shed can be readily found in existing breeding programs and anther extrusion genes can be mapped in doubled haploid populations from crosses that greatly differ in anther extrusion; identifying pollen receptive lines need more research perhaps through using genetic male sterility and random-mating populations; chemical hybridizing agents can be used to make experimental hybrids in sufficient seed quantity for multi-location trials; the genome-wide markers and algorithms are being developed to build heterotic groups which like maize will need to be bred, not discovered; and chemical hybridizing agents and cytoplasmic male sterility systems appear to be commercially viable.

W583: Increasing Genetic Gains for Food Security in the Developing World

A Novel Approach of Rapid and Targeted Gene Transfer from Wild Relatives into Crop Plants

Kanwardeep Singh, Ramanjot Bhullar, Muhammad Ahsan Khan and **Kulvinder S. Gill**, Washington State University, Pullman, WA

The focus of the USAID funded innovation lab is to develop heat tolerant wheat varieties while understanding the heat tolerance trait at molecular, genetic, physiological and biochemical level. Since wild relatives of crop plants are known for their biotic and abiotic stress tolerance, one aspect of the project is to develop a fast, accurate, targeted and efficient method of transferring value added genes such as those controlling heat stress tolerance from wild relatives into cultivated wheat. But so far targeted transfer of such genes has been difficult because of *Ph1* gene imposed restriction on chromosome pairing and recombination between wheat and wild relative chromosomes. With more than 300 useful genes transferred from the wild relatives into wheat, most have not been used in breeding because the transfers were either complete chromosome/arm or large segments which often carried undesirable traits along with the useful genes. We cloned a candidate for the *Ph1* gene, silencing of which resulted in a phenotype characteristic of *Ph1* gene mutants. Complementation of a *Ph1* gene mutant (*ph1b*) with the candidate gene under its native promoter restored the chromosome pairing function. In this study, we transiently silenced the gene via VIGS to induce chromosome pairing and recombination between chromosome 1BS of wheat with 1RS of rye. Out of 250 plants that were analyzed, 66 plants showed recombination between wheat and rye chromosome arm. With an average of 5, the number of rye segments in each recombinant plant ranged from 1 to 6. The size of the rye segment transferred to wheat background ranged from 2 to 100 Mb. Although recombination hot-spots were obvious, recombination events were distributed on the entire chromosome arm.

W584: Increasing Genetic Gains for Food Security in the Developing World

Using Chickpea's Wild Relative Plants and Microbes to Expand Agricultural Diversity and Address Agricultural Challenges in the Developing World.

Douglas R Cook, University of California-Davis, Davis, CA

Chickpea is a pulse legume of critical importance in low-income food insecure countries, in advanced developing economies, and in developed countries. Paradoxically, countries with the highest nutritional demand for chickpea are also those with the lowest yields, often ½ to ¼ of yields found in the developed world. Whole genome sequencing reveals that ~95% of genomic variation was lost from modern elite cultivars during domestication. This has profound implications, because corresponding reductions to trait variation limit the ability to adapt the crop to changing environments and to meet emerging needs, raising an urgent need for new sources of diversity. We have collected and are characterizing a large and systematic set of wild *Cicer* species and co-occurring microbes from a representative range of natural environments.

Genomic technologies have been used to characterize diversity, phenotyping has been used to identify high value traits, and genetic crosses combined with sequence-based genotyping are revealing marker-trait associations to facilitate introgression breeding. Current efforts focus on domestication and yield related traits (e.g., pod shattering, biomass conversion efficiency, seed size, total seed production), while a second generation of traits is under analysis and yielding a range of interesting trait values (e.g., for drought and heat tolerance, pod borer pest resistance, *Ascochyta* blight and *Fusarium* wilt resistance, nitrogen fixation, flowering time and plant architecture). In parallel, we are harnessing chickpea's microbes, with systematic sampling and genomic analysis of both nitrogen-fixing *Mesorhizobium* symbionts and of plant-enriched members of a broader microbial community. Outcomes of this project will be high-yielding, climate-resilient chickpea varieties within the context of user-preferred traits: climate resilience, seed quality and nutrient density, reduced inputs through nitrogen fixation, and biotic stress resistance among them.

W585: Increasing Genetic Gains for Food Security in the Developing World
Genetic Improvement of Banana for Diseases and Pest Resistance
Leena Tripathi, International Institute of Tropical Agriculture (IITA), Nairobi, Kenya
Banana including plantain (*Musa* spp.) is among the world's most important staple food crops cultivated with an annual production of about 145 million tonnes. Africa contributes one-third of the global production with east Africa being the largest banana-growing region accounting for about 40% of the total production in Africa. This crop contributes about 25-30% of calorie intake and mainly grown by smallholder farmers, who rely on banana for food and income. Banana production is constrained by a range of diseases such as fusarium wilt, banana Xanthomonas wilt, black Sigatoka and banana bunchy top disease and pests including weevils and nematodes. Development of improved varieties of bananas is fundamental in order to tackle these challenges. However, conventional breeding of banana is limited by narrow genetic diversity and the sterility of the polyploid genomes of cultivated banana. Biotechnology has opened unprecedented avenues for exploring biological systems. Genetic engineering is one of the key techniques particularly useful for the genetic improvement of crops not amenable to conventional breeding. Progress has been made for improving banana using genetic engineering such as transgenic banana resistance to Xanthomonas wilt disease and nematodes**.** Genome editing is another strong rapidly advancing tool, which has potential to improve agriculture and food security in Africa. The progress in transgenic and genome editing research at IITA for improvement of banana for diseases and pests resistance will be presented and discussed.

W586: Increasing Genetic Gains for Food Security in the Developing World
Role of Animal Genetic Improvement in Enhancing Food Security?
Susan J. Lamont, Department of Animal Science, Iowa State University, Ames, IA, Carl J. Schmidt, Dept. of Animal & Food Sciences, University of Delaware, Newark, DE and Huaijun Zhou, Animal Science, University of California, Davis, CA
Animals are an essential component of a comprehensive program to meet the needs to feed the world, alleviate poverty, and empower women. Chickens are especially suitable because of their portability, wide cultural acceptance, minimal space requirements and nutrient-dense products of eggs and meat. Furthermore, in many cultures, poultry are managed and owned by women. Thermal stress and infectious disease present major limitations to poultry production in much of the developing world. Our collaborative research teams are applying cutting-edge genomic approaches to elucidate the genetic bases for adaptability to heat stress and resistance to disease in chickens. The goals of these studies are to identify the beneficial genetic variation that naturally occurs within the species, and to use that information to breed populations of chickens that are naturally more resistant to the ill effects of heat stress and pathogen challenge. Supported by USAID Feed the Future Innovation Lab for Genomics to Improve Poultry and USDA-NIFA-AFRI Grant 2011-67003-30228.

W587: Increasing Genetic Gains for Food Security in the Developing World
Building International Collaborations
Jagger Harvey, Kansas State University, Manhattan, KS

W588: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
Finding the Missing Links: A Journey Towards Meeting African Farmers' Needs
Jean-Marcel Ribaut, Integrated Breeding Platform, Texcoco, Mexico
Africa imperatively needs to increase food and nutritional security to serve a growing population and reduce food importation costs (currently estimated at US$ 35 billion/year). There is considerable potential to raise agricultural productivity through the development of improved cultivars that lift yields, and respond to both local and global market demands. However, and despite decades of major investment in R4D, the impact in farmers' field remains limited, especially for subsistence crops. Farmers still have difficulty accessing water, fertilizers and phytosanitary products, amongst others, and seed quality and distribution are a major bottleneck in most places. Even if improved germplasm with large genetic potential is available, it often lacks critical or specific local characteristics, or only performs well under optimal conditions. In the African context, some links of the crop value chain are either broken or missing, and only an integrated approach – from crop diversity to production in the field – can have a sustainable impact on agricultural productivity. Improvement toward sustainable change will include the implementation of a demand-led breeding practice, that is based on modern technologies aligned with local reality, and supported by a strong capacity development component (human and infrastructure). Stimulating entrepreneurial spirit to implement local/regional businesses at strategic points down the chain is also a must to succeed. The case for this vision builds on examples and lessons learnt from the Generation Challenge Programme and the Integrated Breeding Platform, after working in R4D, with and for African partners, for more than 15 years.

W589: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
The Breeding Management System: A Robust Informatics Platform for Plant Breeding Workflows
Graham McLaren, Integrated Breeding Platform, Texcoco, Mexico

W590: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
Modernizing ICRISAT Breeding Programs through BMS of IBP and Linking Various Informatics Solutions
Abhishek Rathore, ICRISAT, Patancheru, Hyderabad, India

W591: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
From Breeding Data to Decision by BrAPI and Open Rap
Fred A. van Eeuwijk, Maikel Verouden, Bart-Jan van Rossum, Martin P. Boer and Marcos Malosetti, Wageningen University & Research - Biometris, Wageningen, Netherlands
We present an R based analytical pipeline and an Application Programming Interface (API) that can extract data from the BMS and other data bases and can perform single and multi-site phenotypic analyses as well as single trait QTL analyses.
The Breeding Application Programming Interface (BrAPI) specifies a standard interface for plant phenotype/genotype databases to serve their data to crop breeding applications. It is a shared, open API, to be used by all data providers and data consumers who wish to participate. Initiated in May 2014, it is currently in an actively developing state. The specifications of BrAPI can be found at https://github.com/plantbreeding/API
Based on these BrAPI specifications a package is simultaneously being developped, which uses BrAPI to import data from phenotyp/genotype databases into R, the freely available language and environment for statistical computing and graphics. This R package is available at https://github.com/c5sire/brapi. Currently all version 1 API calls have been implemented in the R brapi package.
In BMS the following BrAPI calls have been succesfully implemented, which also means the matching R brapi functions can be used:
- Authentication.
- Listing of the available crops.
- Listing of the programs available for a specific crop.
- Listing of the trials for a specific crop.
- Retrieving study observation details as a table.
- Listing of locations used.

The R Analytical Pipeline contains

- Single site analysis: default SpATS analysis, if no spatial information, then mixed models based on lme4 (including forming predictions using the lmer object with standard predict function). If no spatial information is available both asreml and lme4 models are included with lme4 as a default for CRAN purposes.
- GxE analysis: simple summary statistics, diagnostic plots, Finlay-Wilkinson model, AMMI model, GGE model, and variance component models for GxE. The variance component models for GxE are computed with asreml. With lme4 only a compound symmetry model will be fitted.
- QTL mapping: single trait QTL mapping
-

W592: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
Revisiting the Modern Synthesis: Integrating Marker-Assisted Selection into a Population Improvement Based Rice Breeding Program
Josh Cobb[1], Holden Verdeprado[1], Juan D. Arbelaez[1], Eng Hwa Ng[2] and Tobias Kretzschmar[1], (1)International Rice Research Institute, Los Baños, Philippines, (2)ICRISAT, Hyderabad, India
Rice breeding is both challenged and benefited by the fact that a successful varietal improvement program must embrace both the integration single genes that segregate in a simple Mendelian fashion as well as complex traits that are inherited in more quantitative ways. For decades the rice genetics community has produced a wealth of knowledge about these single genes and has developed markers that allow a breeder to track them in a population. However, marker assisted selection (MAS) alone is insufficient to drive the rates of genetic gain for more complex traits that are equally necessary. This presentation will describe the attempts made in the Favorable Environments Breeding program at IRRI to integrate the selection for single genes appropriate for MAS into a more complex population improvement strategy designed to improve quantitatively inherited traits.

W593: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
A Journey from Paper and Pencil Based Breeding to a Fully Digitalized Modern Breeding Program
Anupama Hingane, ICRISAT, HYDERABAD, India

W594: International Cotton Genome Initiative (ICGI)
Introduction and Update
John Z. Yu, USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX and Tianzhen Zhang, Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China
Introduction to the ICGI workshop and update of the ICGI status

W595: International Cotton Genome Initiative (ICGI)
Domestication and Improvement of Cotton: From Tree Cotton to American Upland Cotton and to the World's Largest Fiber Crop
Lei Fang[1], Yan Hu[1], Jiedan Chen[1], Xueying Guan[1,2], Baoliang Zhou[2], Wangzhen Guo[2], Xuehui Huang[3], Xiongming Du[4] and Tianzhen Zhang[1], (1)Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China, (2)Nanjing Agricultural University, Nanjing, China, (3)National Center for Gene Research,SIBS,CAS, Shanghai, China, (4)Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Ministry of Agriculture, Anyang, China

Allotetraploid upland cotton (*Gossypium hirsutum* L.) provides the most important natural fiber in the world. It was initially domesticated from tree cotton in the Mesoamerican and Caribbean regions, and subsequently further domesticated and improved in the southern United States, from the perennial to the annual crop. Because of its high-yield property, American upland cotton was extensively introduced and grown in Asian countries such as India and China. To unfold this evolution and domestication history, we resequenced the genomes of 147 cotton accessions, including diverse wild relatives, landraces, and modern cultivars. By comparing the genetic diversity among wild *G. hirsutum* cultivars and races, we identified 109 domestication-related selective sweeps, including 723 fiber related and 115 seed germination-related genes, indicating the contribution of selective sweeps in the domestication for fiber yield and fiber qualities. To further unveil the improvement history, we report a comprehensive genomic assessment of modern improved upland cotton based on the genome-wide resequencing of 318 landraces and modern improved cultivars. More associated GWAS loci for lint yield (71) are detected than those for fiber quality (45), which suggests that lint yield has stronger selection signatures than other traits. Moreover, we evaluated the contributions of various genetics pools to the current cultivars and found that 54.8% of the elite GWAS alleles detected were transferred from three founder landraces. Our findings uncover the domestication and improvement history of allotetraploid upland cottons and provide genomic bases for improving cotton product and for further evolution analysis of polyploid crops.

W596: International Cotton Genome Initiative (ICGI)
The Origin, Diversity, and Domestication of Tetraploid Cotton (*G. hirsutum* L.)
Joshua A. Udall, NCGR, Santa Fe, NM
Genome resequencing of cotton germplasm can uncover the natural history of cotton from its origination to its domestication. Numerous accessions were selected from the USDA collection for resequencing with high coverage, including wild, federal and cultivar tetraploid and diploid cotton. A comprehensive variation map for more than 420 *G. hirsutum* accessions has been constructed, including SNPs and InDels. SNPs were used to create a phylogenetic tree and identify population structure. Four main populations were identified (domesticated, Landrace 1, Landrace 2, and wild). The cultivars appear to be derived from a single landrace suggesting a single domestication event in *G. hirsutum*. The genome was scanned for regions of domestication (pi and Fst) and putative selective sweeps were identified. Our new understanding of the genetic variation in the *G. hirsutum* genome will assist cotton breeders efforts to improve the fiber quality, disease resistance, and yield of modern cotton varieties.

W597: International Cotton Genome Initiative (ICGI)
CRISPR-Cas System for Cotton Genome Editing: High Efficiency, High Throughput and High Precision
Shuangxia Jin, Huazhong Agricultural University, Wuhan, China
As one of gene editing technologies, the CRISPR/Cas9 system has exerted its broad applications from prokaryotes to eukaryotes. Its robustness and high efficiency give it obvious advantage and potential compared to over ZFN and TALENs systems. Up to now, it had conspicuous effects in many plants, including major crops like rice, wheat and zea maize. As an important economic crop, the widely cultivated upland cotton (*Gossypium hirsutum*) is an allotetraploid with a complex genome structure. Most genes have at least two copies originated from A and D subgenome. This feature results in no obvious phenotype in gene functional analysis, because of gene redundancy when using of RNA interference strategy. So, CRISPR-Cas 9 is highly desirable for cotton genome editing. Recently we successfully knock out several cotton genes by CRISPR-Cas 9 system with an average of 65-85% efficiency!
Then, we further developed a high-throughput genome editing system in cotton. A sgRNAs library (containing 1100 sgRNAs targeted to 600 independent genes) was constructed and cloned into the CRISPR-Cas 9 vector. By this way, we can edit several hundred target genes in one transformation. This system needs a very high efficient cotton genetic transformation system to generate thousands of regenerated plantlets by somatic embryogenesis. The data we obtained recently suggested that this system works pretty well in cotton.
Recently, we start to work a new genome editing system in cotton by using a *Francisella novicida* (Fn) CRISPR-Cpf1-based genome-editing method. Cpf1, a single-strand RNA-guided endonuclease of the class 2 CRISPR-Cas system, cleaves targeted DNA with features distinct from those of Cas9. For example, preferring a T-rich protospacer-adjencent motif (PAM) and cutting in staggered ends. This system has several advantages over the CRISPR-Cas 9 system including small Cas protein size, generating sticky end after cleaving the DNA, cutting the RNA target and lower off-target risk.

W598: International Cotton Genome Initiative (ICGI)
Elucidating the Roles of MicroRNAs in Cotton Fiber Initiation and Early Development
Baohong Zhang, East Carolina University, Greenville, NC
Cotton fiber development is a fundamental biological process; investigating cotton fiber initiation and development provides a unique window into the regulation of cell differentiation, cellulose biosynthesis, and further increasing cotton fiber quality and yield. For several decades, a great deal of research has been aimed at elucidating the underlying molecular pathways. Nevertheless, the mechanisms by which cotton fiber differentiates and develops remains unclear. In this study, we employed high throughput deep sequence, degradome sequence as well as quantitative real time PCR to identify and functionally analyze microRNAs (miRNAs), an important gene regulator, in both fiberless mutants as well as its wildtypes. We found that both conserved and novel miRNAs have unique expression pattern in cotton fiber development. During the cotton fiber development, particularly at the 10 DPA, lots of miRNAs show different modification that suggests miRNA-regulating cotton fiber through miENA modification. The miRNA genes controlling cotton fiber development are majorly from subgenome A. Using transgenic, RNAi and genome editing technologies show that overexpression and knockout/knockdown of an individual miRNAs affected cotton fiber development and further affect cotton fiber length and quality.

W599: International Cotton Genome Initiative (ICGI)
Role of Cuticular Wax in Cotton Leaf Curl Virus Disease Resistance in Upland Cotton (*Gossypium hirsutum* L.)

Muhammad Saeed, Government College University, Faisalabad, Pakistan and Sana A., Raza M. H., Anam T., Areba A., Pervaiz M., Hafiz N. A., Shahzad M., Ayesha M., Aisha S., Humaira F., Muniba I., Riffat A. S., Mahtab H., Maria S., Iqra S., Song X., Sun X, Riaz M.

Cotton is important for its diverse economic uses such as fiber, vegetable oil, and bioenergy source. Cotton production is severely affected by Cotton Leaf Curl Virus Disease (CLCuVD) in major cotton growing countries of the world including India, Pakistan, and Sudan. Various strategies are suggested to cope with this disastrous disease of cotton. Cuticular wax is reported to impart resistance against various stresses in plants such as drought, mechanical injury and pathogens. In this study, role of cuticular wax in resistance to CLCuVD was investigated. Cuticular wax load of 120 varieties/lines of *Gossypium hirsutum* was evaluated during 2015 and 2016. Data for CLCuVD infestation was also recorded. Marked genotypic differences were observed for cuticular wax content and CLCuVD infestation. There was significant negative correlation between cuticular wax content and CLCuVD infestation indicating that cuticular wax content is involved in resistance to CLCuVD infestation. Some cotton varieties/lines with high cuticular wax content were found highly tolerant to CLCuVD, while, some other varieties/lines with reduced cuticular wax content showed highest susceptibility to CLCuVD. Results of this research will accelerate efforts to characterize CLCuVD resistance mechanism at genetic level.

W600: International Cotton Genome Initiative (ICGI)

Single-Molecule Sequencing of *G. arboreum* Genome and GWAS Analysis of Chinese *G. arboreum* Accessions

Xiongming Du, Institute of Cotton Research, Chinese Academy of Agricultural Science, Anyang,Henan, China and Xiongming Du1,4,7, Shoupu He4,7, Gai Huang2,3,7, Zhaoen Yang4,7, Gaofei Sun1,7, Xiongfeng Ma4,7, Nan Li5,7， Hongkun Zhen6, Sanwen Huang5, Tao Lin5*, Yuxian Zhu2,3* & Fuguang Li1,4* 1 Research Base， Anyang Institute of Technology, State Key Laboratory of Cot

The allotetraploids of cotton plants are designated as AADD genomes. The diploid ancestral A of the allotetraploid are thought to have diverged from *G. arboreum* or *G. herbaceum*. Here, a higher quality *G. arboreum* genome assembly was got by using PacBio single molecule real-time sequencing technology sequencing and High-throughput chromosome conformation capture (Hi-C) technology. We generated 142.54 Gb of raw PacBio reads (approximately 77.6-fold genome coverage) assembled 8,223 contigs with a N50 length of 1.1 megabases. 40,960 genes were annotated assistant by the PacBio transcriptome analysis. We further sequenced 230 *G. arboreum* accessions (average 6× depth) to generate a map of genome variation including~18 million SNPs and ~2 million indels. A phylogenetic tree based on SNP genotypes, together with an analysis of population structure suggested that Chinese *G. arboreum* originated in South China and was subsequently introduced to the Yangtze River and Yellow River regions. A total of 98 significant peak associations (-logP>6) for 19 agronomically-important traits in *G. arboretum* were identified in our genome-wide association study (GWAS). Most of the population divergence related traits, including yield and disease resistance, have experienced geographic isolation. Gain of *Fusarium wilt* disease resistance in YZ and YR accessions is associated with *GaGSTF9*, whose expression is highly inducible upon fungi inoculation. A non-synonymous SNP mutant substitution of *GaKASIII* seems to have conferred significant alteration of the fatty acid composition (C16:0 and C16:1) in cotton seed. We further identified a new QTL related to seed fuzz development by combining our GWAS analysis with bulk population sequencing. Our work deepens our knowledge about A-genome evolution, especially Chinese *G. arboreum* population.

W601: International Goat Genome Consortium

Introduction

Alessandra Stella, Parco Tecnologico Padano Foundation, Lodi, Italy

W602: International Goat Genome Consortium

Genome-Wide Linkage Disequilibrium (LD) and Signatures of Selection in African Goat Populations

Getinet Mekuriaw Tarekegn, SLU, Uppsala, Sweden

W603: International Goat Genome Consortium

Genome-Wide Association Study of Goat Milk Traits using Low-Depth Genotyping-by-Sequencing

Andrew S Hess[1], Ken Dodds[1], Rudiger Brauning[1], Judy and Barry Foote[2], John C. McEwan[1] and Shannon Clarke[1], (1)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)Foote, RD4, Hikurangi, New Zealand

Genotyping-by-sequencing (GBS) is a sequencing method that uses restriction enzymes to target portions of the genome to obtain genomic information at a higher density and reduced cost compared to standard SNP genotype panels. However, given that the average depth of coverage is typically targeted at 2-4x, there is some level of uncertainty in the SNP genotype scores obtained. This uncertainty is particularly evident when distinguishing between homozygous and heterozygous genotypes at loci where only one allele is observed for an individual. Therefore, methods for genome-wide association studies (GWAS) that account for uncertainty in GBS genotype calling are desirable when searching for loci associated with a trait. Such methodologies have already been developed for conducting GWAS using imputed genotypes, in which the probability of the individual carrying each genotype (e.g. AA, AB, or BB) is used instead of the genotype score. This allows individuals whose genotypes have been imputed with low levels of certainty to be included in the analysis where they may have otherwise been left out. Similarly, we use this approach to account for uncertainty in genotyping due to low-depth sequencing. We demonstrate the utility of this method for GBS data by performing a GWAS on milk traits in dairy goats, whereby we identified strong functional candidates for genes associated with milk traits. Thus, GBS can be used as a suitable substitute for SNP chips when performing GWAS.

W604: International Goat Genome Consortium

Most of the Variation of the Extinct *Capra pyrenaica pyrenaica* is still present in Two Related *Capra pyrenaica* Subspecies

Marcel Amills[1], Taina Figueiredo-Cardoso[2], Raul Tonda[3], Rayner Gonzalez[4], Ricardo Garcia-Gonzalez[5], Alberto Fernandez Arias[6], Jose Folch[7], Antonia Noce[2], Armand Sanchez[8], Alex Clop[9] and Sergi Beltran Beltran[3], (1)Center for Research in

Agricultural Genomics, BARCELONA, Spain, (2)Center for Research in Agricultural Genomics, Bellaterra, Spain, (3)Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain, (4)Center for Research in Agricultural Genomics - University Autonomus of Barcelona, Bellaterra, Spain, (5)Instituto Pirenaico de Ecología, Jaca, Spain, (6)Service of Hunting, Fishing and Wetland of Aragon, Zaragoza, Spain, (7)Agrifood Research and Technology Centre of Aragon, Zaragoza, Spain, (8)Universitat Autonoma de Barcelona, Bellaterra, Catalunya, Spain, (9)Center for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Bellaterra (Barcelona), Spain
The Spanish ibex (*Capra pyrenaica*), a wild goat species, suffered a strong population reduction during the 19-20th centuries as a consequence of habitat destruction, hunting and epidemics. Two subspecies, *Capra pyrenaica hispanica* (CPH) and *C.p.victoriae* (CPV), survived this genetic bottleneck, while *Capra pyrenaica pyrenaica* (CPP) disappeared 17 years ago and *Capra pyrenaica lusitanica* became extinct in 1892. One of the consequences of population decline and inbreeding is the increase in the proportion of deleterious mutations. By sequencing the genome of one of the last CPP representatives, we aimed to investigate the genetic consequences of such processes on the variation of this subspecies. Moreover, we sequenced genomic DNA pools of CPH and CPV individuals with the goal of quantifying the variation that is shared between CPP and CPH and CPV. Genome sequencing of CPP, CPH and CPV demonstrated the existence of 2,074,770 SNPs. The majority of these polymorphic sites (83.84%) were shared amongst the three subspecies, and also CPH and CPV had a substantial amount of SNPs (≈92%) in common. The most abundant SNPs in *Capra pyrenaica* were those mapping to intergenic (43%) and intronic (44%) regions. Deleterious variants had a low frequency, and the most abundant ones were those altering protein structure. Our results suggest that most of the variation of the extinct CPP is still present in the two remaining subspecies. A next step will be to provide a quantitative estimate of the deleterious mutation load in CPP.

W605: International Goat Genome Consortium
Impact of Haplotype Phasing on Genome Assembly and Inference
Ramadimetja Prescilla Mohlatlole, University KwaZulu-Natal, Pietermaritzburg, South Africa

W606: International Goat Genome Consortium
Transcriptomics of Innate Immune Response in the Domestic Goat and Comparative Analysis with Sheep
Charity Muriuki[1], Mike Coffey[2], Morris Agaba[3], David A. Hume[1,4] and Emily L. Clark[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)Scotland Rural College (SRUC), Edinburgh, United Kingdom, (3)Sarissa Ltd, Private box, Kampala, Uganda, (4)Mater Research, University of Queensland, Translational Research Institute, Woolloongabba, Australia
Despite great genetic similarity, ruminants vary in their susceptibility to similar pathogens but the underlying molecular mechanisms remain largely unknown. To elucidate the molecular basis of variation in disease response, we generated a gene expression atlas of the domestic goat from a subset of tissue and cell types and compared it to sheep and other ruminants. We produced 54 TruSeq 75bp paired-end libraries on the Illumina platform at 25M reads per sample including six libraries from bone-marrow derived macrophages stimulated with lipopolysaccharide to mimic pathogen challenge by gram-negative bacteria. Transcripts were quantified using Kallisto to generate gene expression estimates as transcripts per million (TPM), capturing nearly 90% of the annotated protein-coding genes in the goat reference transcriptome. We visualize the data as gene-gene network graphs in Miru (www.kajeka.com) to investigate tissue-specific gene expression profiles and assign function to unannotated genes through principle of guilt-by-association. Additionally, using percentage similarity of protein-identity between goat and sheep, we are running a comparative transcriptomic analysis with the recently released high-resolution gene expression atlas of the domestic sheep. We identify numerous, previously unannotated genes involved in innate immune response in goat and provide useful data to improve annotation of the goat reference genome (ARS1). The sample metadata have been loaded into BioSamples and sequence data submitted to the ENA available for use by the wider research community. This project aids the understanding of molecular mechanisms controlling variation in disease susceptibility across ruminants and provides a valuable resource for the study of ruminant functional genomics.

W607: International Goat Genome Consortium
Can We Apply Lessons Learned from Manual Gene Annotation in Human and Mouse to Goat?
Adam Frankish, EMBL-EBI, Hinxton, United Kingdom

W608: International Goat Genome Consortium
Conclusion
Alessandra Stella, Parco Tecnologico Padano Foundation, Lodi, Italy

W609: International Phytomedomics and Nutriomics Consortium (ICPN)
Genomics Resources Towards Breeding for Medicinal Value in *Humulus lupulus* and Related Species
Paul D. Matthews, Hopsteiner, S.S. Steiner, Inc., New York, NY

W610: International Phytomedomics and Nutriomics Consortium (ICPN)
Nicotine Biosynthesis Pathway: Beyond Tobacco
Nikolai V. Ivanov, PMI R&D, PMP S.A., Neuchatel, Switzerland

W611: International Phytomedomics and Nutriomics Consortium (ICPN)
Genomics of the Healthful and Multi-Use Crop Globe Artichoke

Alberto Acquadro, Ezio Portis, Lorenzo Barchi, Cinzia Comino, Andrea Moglia, Danila Valentino, Anna Milani and Sergio Lanteri, DISAFA, Plant Genetics and Breeding, University of Torino, Grugliasco, Italy
Globe artichoke (*Cynara cardunculus* var. *scolymus*, *Asteraceae*, 2n=2x=34, genome = 1,06 Gb) is a multi-use crop. Other than producing immature inflorescences used as a vegetable, it is exploited as a source of inulin and bio-active compounds such as phenolics (mono- and di-caffeoylquinic acids, flavonoids) and sesquiterpene lactones (cynaropicrin, grosheimin), the latter being responsible for its peculiar bitter taste. Both globe artichoke and its related taxa cultivated cardoon (*C. cardunculus* var. *altilis)* are also grown as ornamentals and for the production of lignocellulosic biomass and seeds rich in oil of good alimentary quality and exploitable for biodiesel production. We recently released the first globe artichoke reference genome sequence (www.artichokegenome.unito.it), which includes about 28,000 genes. Following the re-sequencing of an F1 segregating progeny (globe artichoke x cultivated cardoon), 73% of the assembled genome was anchored to 17 chromosomal pseudomolecules. We also performed the re-sequencing of four globe artichoke genotypes, which are representative of the core varietal types in cultivation, and a genotype of cultivated cardoon. Their genomes were reconstructed at chromosomal scale and analogous numbers of genes were predicted, while distinctive variations in miRNAs and resistance genes (RGAs) were identified. Among the 23,5M spotted SNPs, some were predicted to influence the biological functions of genes involved in phenolics and sesquiterpene lactones biosynthesis. The publicly available genomic resources represent key tools to dissect the path from sequence variation to phenotype and key information for assisting molecular breeding for quality traits, as well as for future genome editing projects.

W612: International Phytomedomics and Nutriomics Consortium (ICPN)
Analysis of Variations in Genes Controlling Fruit Colors of Capsicum using SMRT Sequencing

Min-Young Kang and **Byoung-Cheorl Kang**, Seoul National University, Seoul, South Korea
In plants, carotenoid plays important roles in photosynthesis, photo-protection, and phytohormones synthesis such as ABA and strigolactone. In addition, carotenoid provides color to flowers and fruits for attracting insects or animals for pollination and seed dispersal. In *Capsicum*, ripe pepper fruits display various colors ranging from white to deep red. In the carotenoids synthesis pathway, capsanthin-capsorubin synthase (*CCS*), phytoene synthase (*PSY1*), β-Carotene hydroxylase (*CrtZ-2*) and lycopene β-cyclase (*LcyB*) genes are known for regulators of fruit color. In this study, a total 101 of accessions were used for sequence analysis of these genes using single molecule real time (SMRT) sequencing technology. We detected various variants in the coding regions of the each gene. The nucleotide sequences were aligned to those of red color fruit accessions. Sequencing analysis showed that 9 point mutations and 2 frame shift mutations in the coding region of the *PSY1* gene. Among these mutations, 2 nonsense mutation and 2 frame shift mutations lead to premature stop codon. In the coding region of *CCS*, 6 missense mutations, 5 nonsense mutations, and 6 frame shift mutations were identified. In addition, we found structural mutation in the *CCS* promoter region. In the coding region of *LcyB*, there were 7 missense mutations and 1 frame shift mutation. In coding region of the *CrtZ-2* gene, 8 missense mutations and 1 frame shift mutation were detected. Specially, sequence analysis of the *CrtZ-2* gene revealed a big structural mutation in the non-red color fruits that may result in no expression of this gene. Taken together, we revealed various genetic variations of *PSY1, CCS, LcyB*, and *CrtZ-2* genes which are responsible for non-red color fruit in *Capsicum* spp.

W613: International Phytomedomics and Nutriomics Consortium (ICPN)
Functional Genomics of Sesame for Developing Varieties with Healthy Oil Composition

Hongmei Miao, Henan Sesame Reserach Center, Henan Academy of Agricultural Sciences, Zhengzhou, China and Haiyang Zhang, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, China
Sesame (Sesamum indicum L.) belongs to the Pedaliaceae family and is a high- quality oilseed crop with the honor of 'Queen of oil seeds' for the lignan content up to 1.5%. Sesame seeds contain 50-55% oil, and the principal fatty acids are oleic acid (18:1) (39.6%) and linoleic acid (18:2) (46.0%). The genome structure and the comparative genomics analysis reflected that more than 900 gene families were found contributed to oil biosynthesis and oil metabolism process in sesame. About 53 gene families, such as ketoacyl- ACP Synthase III, stearoyl-ACP desaturase and acyl-CoA thioesterase genes significantly expanded. Transcriptomes of seed formation of several sesame varieties presented the various expression profile of the gene families which contributed to the high oil content and the equal oleic and linoleic acid ratio. Moreover, the genome analysis of eight hundreds of sesame germplasm indicated that amounts of SNP positions in some genes related with the oil biosynthesis process experienced the positive selection during artificial cultivation and breeding. The transition of the genes controlling the oil content and fatty acid composition in 20 descent lines of the elite Chinese variety Yuzhi 4 were presented accordingly. In addition, we performed the sesamin function analysis using 3T3-L1 cell model and the C57BL/6 mice model. The obvious nutrition effects of sesamin on adipocyte differentiation and morphology was proved, and the molecular mechanism of sesamin effect was being analyzed. The findings supply the basis for future studies of molecular breeding, nutrition and metabolism in sesame.

W614: International Phytomedomics and Nutriomics Consortium (ICPN)
Improvement of Human Nutritional Value in *B. rapa* (Chinese Cabbage) through Molecular Breeding

Jana Jeevan Rameneni[1], Vignesh Dhandapani[1], Xiaonan Li[2], SuRyun Choi[1] and **Yong Pyo Lim**[1], (1)Chungnam National University, Daejeon, South Korea, (2)Shenyang Agricultural University, Shenyang, China
The *Brassica rapa* (2n = 20, AA) is one among the six economically important cultivated *Brassica* species of U's triangle. *The B. rapa* subspecies has wide genetic and morphological diversity which grown as leafy vegetables, vegetable oils, turnip greens, turnip roots, turnip tops and as a fodder crop. Among *Brassica* species,*B. rapa* is the smallest genome (529Mbp) and an ideal candidate genome as a reference, compared to *B. nigra* (BB, 632Mbp) and *B. oleracea* (CC, 696Mbp). In general plant secondary metabolites plays vital roles during different stages of growth and development. These functional compounds add high nutritional value to humans so the enrichment of these nutritional supplements in Chinese cabbage cultivars is one of the important objectives. We have generated doubled haploid (DH) lines through microspore culture from the collected germplasm accessionswith high functional compounds like glucosinolates, vitamin C, total sugars and calcium. For genetic analysis on glucosinolates metabolism through molecular markers approach, we have performed a conventional QTL analysis using F2/3 mapping populationof *B. rapa* combined with candidate gene association approach by using natural population, in order to identify the genomic region and genes regulating glucosinolates biosynthesis in *B. rapa* crops. Recently, we have re-sequenced 145 Chinese

cabbage accessions and identified ~1.7 million high quality SNPs for the association studies. Results suggest several alleleswith very high association for important compounds like gluconapin, neoglucobrassicin, sinigrin and glucobrassicanapin. Additionally, the comparative analyses of several association results were completely matching with previous analyzed QTL maps.The further analysis will be done to study the identified candidate genes related to glucosinolates enhancement.

W615: International Phytomedomics and Nutriomics Consortium (ICPN)
Toward an Integration of Phytomedomics and Omics Approaches to Developing Functional Vegetable Variety in *Allium*
Masayoshi Shigyo, Yamaguchi University, Yamaguchi, Japan, Shusei Sato, Tohoku University, Sendai, Japan, Hideki Hirakawa, Tohoku University, Kisarazu, Japan, Yuji Sawada, RIKEN Center for Sustainable Resource Science, Yokohama, Japan and Satoshi Fujito, National Agriculture and Food Research Organization (NARO), Tsu, Japan
Allium cepa (bulb onion, shallot) is widely cultivated species of the genus *Allium*, containing chemical compounds with potential anti-inflammatory, anticholesterol, anticancer and antioxidant properties. A complete set of *A. fistulosum – A. cepa* monosomic addition lines has been used for obtaining genomic information in *A. cepa*. Toward the advancement of whole genome sequencing project an F_2 mapping population was produced from a single F_1 plant between shallot and bulb onion doubled haploid (DH) lines, and could be used for constructing *A. cepa* ultra-high density genetic linkage map via the use of numerous SNP markers generated by RNA-sequencing along with a role in Transcriptome-based genotyping. Furthermore, an advanced metabolomics technique was utilized as a tool for characterizing phytochemical variations in the genetic stocks mentioned above. This will develop capability and plant materials to support metabolomics-informed biomedical studies of whole plants or extracts, as well as enabling detection of associations between phytochemical content, gene expression and specific genome regions. Together with the method of simultaneous measurements for anti-inflammatory, antioxidative and innate immunopotentiative activities, the knowledge and technologies will be applicable for targeting breeding, production and processing towards enhanced functionality.

W616: International Phytomedomics and Nutriomics Consortium (ICPN)
A Novel Carotenoid Accumulation Mechanism Revealed through Analysis of Genetic Structure and Domestication in Carrot (*Daucus carota* L.)
Shelby Ellison[1], Keo Corak[2], Claire Luby[2], Douglas Senalik[3], Massimo Iorizzo[4], Irwin Goldman[2], Philipp W. Simon[5] and Julie Dawson[6], (1)USDA-ARS, Madison, WI, (2)University of Wisconsin, Madison, Madison, WI, (3)USDA, ARS, Madison, WI, (4)Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC, (5)USDA-Agricultural Research Service, Vegetable Crops Unit, University of Wisconsin-Madison, Madison, WI, (6)University of Wisconsin, Madison, WI
Carrots accumulate more carotenoids than any other crop. Genetic variation in the carotenoid biosynthetic pathway does not fully explain the accumulation of such high levels of carotenoids in carrot roots. Using a diverse collection of modern and historic cultivated varieties and wild carrot accessions, an association analysis revealed a significant association with a region of the genome that contains the *Or* gene, advancing the *Or* gene as a candidate for carotenoid accumulation in carrots. This gene family has been demonstrated to control carotenoid accumulation in other crop families but has not previously been described in carrot. Our analysis also allowed us to more completely characterize the genetic structure of carrot, showing that the western domesticated carrot largely forms one genetic group, despite dramatic phenotypic differences among market classes within this group. Eastern domesticated and wild accessions form a second group, which reflects the more recent cultivation history of carrots in the specific eastern region of central Asia; and other wild accessions form distinct geographic groups, with a very well-defined group on the Iberian peninsula. Using genome-wide Fst, nucleotide diversity, and Tajima's D statistics, we analyzed the genome for regions putatively under selection during domestication, and identified five regions that were significant for all three methods of detection, one of which includes the *Or* gene. This provides further support that *Or* was important in the early stages of carrot domestication and improvement and may explain why it has not been found with less genetically diverse mapping populations.

W617: International Sheep Genomics Consortium
ISGC Welcome and Updates
Shannon Clarke, AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

W618: International Sheep Genomics Consortium
The Ovine FAANG Project
Brenda M. Murdoch[1], Stephen N. White[2], Michelle R. Mousel[2], Alisha T. Massa[3], Kim C. Worley[4], Alan L. Archibald[5], Emily L. Clark[5], Brian Dalrymple[6], James W. Kijas[7], Shannon Clarke[8], Rudiger Brauning[8], Timothy P.L. Smith[9], Tracy Hadfield[10] and Noelle Cockett[10], (1)University of Idaho, Moscow, ID, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (3)Washington State University, Pullman, WA, (4)Human Genome Sequencing Center, Department of Molecular and Human Genetics, Houstin, TX, (5)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (6)Institute of Agriculture, The University of Western Australia, CRAWLEY, Australia, (7)CSIRO Animal, Health and Food Science, St Lucia, Australia, (8)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (9)USDA, ARS, USMARC, Clay Center, NE, (10)Utah State University, Logan, UT
The functional annotation of the sheep genome will facilitate a better understanding of the complex nature of gene regulation within this globally important food and fiber species. Through the generation and compilation of transcription and sequence data, we will characterize and define the multifaceted biological mechanisms that contribute to gene regulation. This project examines regulatory signals for coding and non-coding transcript isoforms and alternative splicing, promoters and cis-acting regulatory elements, open chromatin, histone modifications, and DNA methylation across a wide range of sheep tissues. Members of International Sheep Genomics Consortium (ISGC) and other researchers used protocols compliant with the FAANG Consortium assays to collect approximately 100 tissues from the same Rambouillet female used in

a *de novo* genome assembly. In order to characterize the transcriptome, we performed three types of RNA sequencing including long read PacBio IsoSeq, and short read Illumina mRNA and miRNA sequencing. Furthermore, to complement gene expression data and to identify active promoters and confirm transcription start sites, cap analysis of gene expression (CAGE), is being performed on all tissues. Further, to assess chromatin accessibility, ATAC-seq is being performed for all tissues using fresh nuclear DNA collected from 10 tissues at slaughter for library preparation and sequencing. Individual cells were collected from 18 additional tissues, and intact samples were taken from remaining tissues and slow-frozen for ATAC-seq analyses. Overall, this project will provide tissue-specific detailed understanding of gene regulatory signals and gene products that contribute to evolution, development and functional phenotype of sheep.

W619: International Sheep Genomics Consortium
Rambouillet Sheep Reference Genome and Initial Analysis of FAANG Sample RNA Sequencing
Kim C. Worley, Baylor College of Medicine, Houston, TX
We report the high quality Rambouillet sheep reference genome and initial analysis of FAANG sample RNA sequencing from the reference ewe, Benz2616.
We *de novo* assembled 200 Gb of Pacific Biosciences (PacBio) sequence with 12.6 kb N50 sub-read length with Celera Assembler and polished with Arrow. Scaffolding the contigs using Hi-C data and Phase PGA incorporated 98.1% of the assembly into 32 large scaffolds. Scaffold gaps were filled using PBJelly, misassemblies identified with misFinder and additional gap-filling completed. Error correction using Pilon and Illumina data produced the final 2.87 Gb genome. More contiguous, complete and correct than most, the contig N50 is 2.6 Mb, with half the genome in 309 contigs (longest 16.3 Mb). Most ESTs (98% of 338,551) align to the genome, 90% with nearly complete alignments, aligning over >90% of their length. Base quality is high, (error rate <1%).
FAANG assays from over 100 collected reference animal tissues including PacBio IsoSeq, Illumina RNAseq and miRNAseq, ATAC-Seq and other assays are underway, to complement the genomic PacBio, Illumina and Hi-C sequence. RNA sequence analysis of PacBio IsoSeq, Illumina RNAseq and microRNAseq (5, 9 and 20 tissues respectively) is ongoing. IsoSeq matched ~11,000 per tissue (15,888 total) of 20,921 annotated proteins. Illumina RNAseq with ~100x more reads per tissue identified ~15,500 transcripts per tissue. MicroRNA sequences analyzed using miRDeep2 identified a total of 6,523 novel miRNAs and 659 known miRNAs, with most (471) of the known and many of the novel miRNAs similar to annotated bovine miRNAs.

W620: International Sheep Genomics Consortium
SheepGenomeDB
Hans D. Daetwyler, Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Australia

W621: International Sheep Genomics Consortium
Can We Apply Lessons Learned from Manual Gene Annotation in Human and Mouse to Sheep?
Jane Loveland, EMBL-EBI, Cambridge, United Kingdom

W622: International Sheep Genomics Consortium
Signatures of Selection in Ethiopian Sheep Populations Adapted to Diverse Environments
Kwan-Suk Kim, College of Agriculture, Life & Environmental Sciences, Chungbuk National University, Cheongju, South Korea

W623: International Sheep Genomics Consortium
Ancient Sheep across Europe and Western Asia
Andrew John Hare, Trinity College Dublin, Dublin 2, Ireland

W624: International Sheep Genomics Consortium
Effects of Maternal Diets on Fetal Genomic Imprinting in Sheep
Jingyue (Ellie) Duan, Department of Animal Science, University of Connecticut, Storrs, CT
Genomic imprinting is an epigenetics phenomenon that causes differential allelic gene expression based on parental origin. To date, 255 imprinted genes have been identified or predicted in all mammals combined. However, imprinting study in sheep lags behind, as only 21 imprinted genes have been described. The current study used DNA/RNA throughput sequencing to identify monoallelically expressed and imprinted genes in day 135 sheep fetal organs, and to access the influence of maternal nutrition on imprinting. We solved several technical challenges in NGS data analysis pipeline including alignment bias of RNA sequencing reads and filtering potential false positives. We identified 80 monoallelically expressed gene and 18 imprinted genes, five of which were previous known imprinted in sheep, and thirteen were known imprinted in other species. Sanger sequencing confirmed four new sheep imprinted genes *INPP5F, PLAGL1, CASD1* and *PPP1R9A*. Among the thirteen new imprinted genes, five located in the sheep known imprinting clusters of *MEST* domain on chromosome 4, *DLK1/GTL2* domain on chromosome 18 and *KCNQ1* domain on chromosome 21, three were in a novel sheep imprinted cluster on chromosome 4 known in other species as *PEG10/SGCE*. Additionally, we found *PHLDA2, SLC22A18, DIRAS3*, and *IGF2* differentially expressed, but no allele expression reversal or loss of imprinting among three maternal nutritional groups. Our results expand the imprinted gene list to 34 in sheep and demonstrate the influence of maternal diet on fetal imprinting in sheep under the conditions studied.

W625: International Sheep Genomics Consortium
Characterizing Allelic Variation in the Recombination Hotspot Mediator Gene PRDM9 in U.S. Sheep
Kimberly M. Davenport, University of Idaho, Moscow, ID
Meiotic recombination is an important process during gametogenesis that ensures proper chromosome segregation and contributes to genetic variation. It is clear from previous studies that recombination events are not entirely random, and exhibit distinct location preferences with high

activity, termed “hotspots.” The *PRDM9* gene encodes a zinc finger protein that mediates hotspot usage and locations in mice, humans, and presumably sheep. In addition, incompatibility between *PRDM9* alleles is known to cause sterility in male mice. In this study, we characterized *PRDM9* alleles and identified two (C and D) not previously reported in U.S. sheep. These alleles were identified by the size of the zinc finger region, and were similar to those in populations abroad. The D allele (9 zinc fingers in length) was the most common, and the C allele (8 zinc fingers in length) was only observed in a heterozygous CD genotype. To further characterize *PRDM9* in U.S. sheep, we identified haplotypes in the zinc finger region using the USMARC Sheep Diversity Panel v2.4 which differed from those previously reported. Characterizing *PRDM9* will allow us to better discern the use and location preferences of recombination hotspots in the genomes of U.S. sheep populations. Further, defining *PRDM9* alleles in U.S. sheep will enhance our ability to understand whether they significantly affect reproduction in males or the accuracy of genetic predictions.

W626: International Sheep Genomics Consortium
Transcriptomic Analyses Reveals Peroxisome Proliferator Activated-Receptor Signaling Pathway as a Potential Regulator Mechanism in Gastrointestinal Parasite Infections in Sheep
Pablo Peraza, Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay

W627: International Wheat Genome Sequencing Consortium (IWGSC)
The Reference Sequence for the Bread Wheat Genome
Frederic Choulet, INRA GDEC, CLERMONT-FERRAND, France and International Wheat Genome Sequencing Consortium
The generation of a high-quality reference genome sequence for bread wheat, linked to genetic and genomic resources, has been the goal of the International Wheat Genome Sequencing Consortium (IWGSC) since its foundation in 2005. Here, we report on the assembly and deep analysis of the 21 chromosomes of the allohexaploid bread wheat *cv*. Chinese Spring: IWGSC RefSeq v1.0. We used an Illumina-based whole genome shotgun approach integrated with a wealth of community resources and were able to assemble 21 high-quality pseudomolecules representing 94% of the predicted wheat genome size, with a scaffold N50 of 23 Mb. We predicted 107,886 high confidence gene models and ~4 million transposable elements accounting for 85% of the genome.
Comparative analyses of the A-B-D sub-genomes revealed no subgenome dominance, and a highly conserved gene set although only 55% of the homeologous groups correspond to 1:1:1 triplets, meaning that A-B-D have been strongly impacted by lineage-specific gene duplications. Insights into gene expression have been described through a transcriptome atlas developed from 850 RNASeq datasets representing all stages of wheat phenological development. With a sequence assembly that now supports the resolution of complex gene families associated with important traits, the community now has a key resource in place for future research and breeding.

W628: International Wheat Genome Sequencing Consortium (IWGSC)
Development of Genetic and Genomic Resources to Evaluate Wheat Organellar Genome Variants and their Functional Implications
Katie L. Liberatore[1,2], Marisa E. Miller[1,2], Kevin A. T. Silverstein[3], Castle Raley[4], Bao Tran[4] and Shahryar F. Kianian[1,2], (1)Cereal Disease Laboratory, United States Department of Agriculture-Agricultural Research Service, St. Paul, MN, (2)Department of Plant Pathology, University of Minnesota, St. Paul, MN, (3)Supercomputing Institute, University of Minnesota, Minneapolis, MN, (4)Leidos Biomedical Research, Inc; Frederick National Laboratory for Cancer Research, Frederick, MD
Organellar genome diversity represents a potential untapped source for improving agronomic traits. Due to the nature of hybridization and domestication events that led to modern cultivated wheat, organellar (mitochondria and chloroplast) genome diversity was dramatically reduced. However, wheat has the largest known collection of alloplasmic (alien cytoplasm) lines. In these lines, the cytoplasms, and therefore the organelles, of wild relatives have been mismatched with the nuclear genome of domesticated wheat through extensive backcrossing. Disruption of native nuclear-cytoplasmic interactions (NCI) impacts a number of agronomic traits including fertility, biomass, grain yield, and stress response. To understand the functional implications of organellar genome variants and genetic mechanisms underlying NCI, we generated organellar genomic resources. We developed a method to couple organellar DNA enrichment from total gDNA utilizing a pull-down approach and ultra-low input library preparations for long-read sequencing. We sequenced 20 organellar-enriched samples with PacBio, including 13 diverse wild species, *T. durum*, *T. aestivum* cv. Chinese Spring, and three alloplasmic lines. We also generated Illumina short-read sequences for >75 cultivars, wild species, and alloplasmics. Comparative analyses are investigating organellar diversity across the Triticum-Aegilops complex and how sequences change in the alloplasmic condition. In parallel, we are generating additional genetic and genomic resources in wheat and the model system *Brachypodium distachyon* for functional genomics studies. These resources will be useful to the broader community for furthering our understanding of the molecular mechanisms involved in NCI and their affects on plant phenotype as well as for breeding efforts to improve agronomic traits.

W629: International Wheat Genome Sequencing Consortium (IWGSC)
Characterisation of the Pentatricopeptide Repeat Protein Family in the Wheat IWGSC Refseq v1.0 Reference Genome
Joanna Melonek, The University of Western Australia, Crawley, Australia
The family of pentatricopeptide repeat (PPR) proteins is one of the largest gene families in flowering plants and has agronomical importance as a source of restorer of fertility (Rf) genes used to suppress cytoplasmic male sterility during the development of F1 hybrids. Typically, flowering plant genomes contain 550-700 PPR genes, in the wheat IWGSC RefSeq v1.0 reference genome we found 1686 PPRs. The large number of PPR genes is primarily due to polyploidy and it’s actually lower than expected from simply adding genes present in the progenitor diploid genomes. This implies PPR gene inactivation and loss during polyploidization, for which we found evidence in the form of truncated or frame-shifted gene fragments. 207 PPRs were identified as restorer of fertility-like (RFL) genes in the wheat reference genome, far more than in any other plant genome analysed to date. We show that locations of some of the previously mapped restorer genes overlap with the genomic locations of RFL clusters identified in our study. This is the first comprehensive analysis of the PPR and RFL families in wheat. The sequence knowledge gained from this project has the potential to accelerate hybrid wheat breeding programs by facilitating the identification of active

restorer genes in potential restorer lines. Hybrid wheat varieties are expected to have higher and more consistent yields by better adaptation to increasingly unpredictable weather conditions in the era of global climate change.

W630: International Wheat Genome Sequencing Consortium (IWGSC)
Map-Based Cloning of Powdery Mildew Resistance QTL Introgressed to Bread Wheat from the Timopheevi Group Reveals a Highly Divergent Region with Suppressed Recombination containing a Cluster of NLR Gene Homologues

Miroslav Valárik, Institute of Experimental Botany, Olomouc, Czech Republic

Introgression of *QPm.tut-4A* locus from *Triticum militinae* into the distal end of bread wheat chromosome 4AL confers improved resistance against powdery mildew. The locus was high-density mapped and delimited to 0.024 cM using 8327 individuals and 75 markers. Using additional 2052 *ph[1]* lines seven new recombinations were identified. After chromosome walking, final flanking markers *owm169* and *owm228* were mapped and the region was found 640.8 kbp and 480.2 kbp long in cv. Chinese Spring (CS) and *T. militinae* (TM), respectively. The cM/Mb ratio is much smaller compared to these commonly found at the end of wheat chromosomes. The sequenced region was annotated and 16 and 12 protein coding genes were identified in CS and TM, respectively. Out of them, seven CS and six TM genes were not syntenic. Furthermore, intergenic regions do not show a significant similarity between CS and TM. The TM region containing the remaining six genes has a syntenic counterpart in CS, but that region was duplicated and one of the duplications was inverted. The duplication and inversion were accompanied by gene loss and four of the TM genes have their counterparts in both duplicated regions in CS. Finally, three genes from the CS region do not have their homologs in the TM region. These structural and sequence differences are major reasons for the discrepancy between the expected and observed cM/Mb ratio. This work was supported by award LO1204 from the National Program of Sustainability I and by the Estonian Ministry of Agriculture.

W631: International Wheat Genome Sequencing Consortium (IWGSC)
Hunting Yellow Rust Resistance Genes in Bread Wheat

Clemence Marchal, John Innes Centre, Norwich, United Kingdom

Achieving wheat yields to meet current and future demands is crucial. This, however, remains challenging in part due to the numerous pathogens threatening wheat production, including yellow (stripe) rust (*Puccinia striiformis fsp tritici; Pst*). Despite over 70 designated yellow rust resistance genes *(Yr)* in wheat, few have been cloned. This lack of knowledge hinders efficient marker assisted breeding and exploitation of novel allelic variation. We recently exome sequenced a mutant population of UK cultivar Cadenza which carries the major gene *Yr7*. Screening 1,000 mutagenized individuals with Pst identified seven susceptible lines presumed to carry mutations in *Yr7*. To test this, mutational resistance gene enrichment sequencing (MutRenSeq) was conducted on the susceptible lines and a candidate for *Yr7* was identified. Taking advantage of the IWGSC RefSeqv1.0 assembly, we quickly determined the physical position of the closest Chinese Spring homolog within the Yr7 region and confirmed this linkage in an F2 population. Previously, *Yr5* was proposed to be allelic to *Yr7*. Therefore, a similar approach was carried out on *Yr5* susceptible mutants and a single candidate gene, different from *Yr7*, was identified. The *Yr5* homolog is located in close physical proximity to *Yr7* on IWGSC RefSeqv1.0. This suggests that *Yr5* and *Yr7* are very closely linked genes rather than true alleles. The closest homologs to both *Yr7* and *Yr5* reside in a complex disease resistance cluster in RefSeqv1.0. We will present a phylogenetic analysis of this resistance gene cluster in Chinese Spring and additional commercial varieties and discuss the implications for breeding.

W632: International Wheat Genome Sequencing Consortium (IWGSC)
Worldwide Phylogeography and the History of Wheat Genetic Diversity

Etienne Paux, GDEC, INRA, UCA, Clermont-Ferrand, France

The history of bread wheat (*Triticum aestivum* L.) is intertwined with the history of humankind. Since its formation in the Fertile Crescent during the Neolithic, ~8,000 to 10,000 years ago, wheat has undergone a complex history of spread, domestication and selection. To get better insights into the wheat phylogeography and genetic diversity, a set of 4,600 accessions originating from 105 different countries and comprising both landraces and cultivars was genotyped with the TaBW280K SNP array. Eventually, high-quality genotyping data were obtained for 4,506 wheat accessions with 113,457 SNPs (99,333 Polymorphic High Resolution and 14,124 Off-Target Variant markers) whose genomic position was determined using the IWGSC RefSeq v1.0. Based on 632 landraces, 8,741 haplotypes were identified and used to analyze the population structure, as well as the genetic diversity in the whole set of 4,506 accessions. In parallel, the 14,124 Off-Target Variants were employed to investigate further structural variations of the wheat genome. Our results shed light on the complex history of bread wheat and show how man has influenced the worldwide genetic diversity of this species.

W633: Interoperability and Federation Across Bioinformatic Platforms and Resources
Introduction

Eric Lyons, University of Arizona; BIO5 Institute; CyVerse, Tucson, AZ

W634: Interoperability and Federation Across Bioinformatic Platforms and Resources
MiCloud and BioDocklets: A Plug and Play, on-Premises Bioinformatics Cloud, Providing Seamless, Single-Step Execution of NGS Pipelines.

Konstantinos Krampis, City University of New York, Weill Cornell Medical College, New York, NY

Processing of Next-Generation Sequencing (NGS) data requires significant technical skills, involving installation, configuration, and execution of bioinformatics data pipelines, in addition to specialized post-analysis visualization and data mining software. In order to address some of these challenges, developers have leveraged virtualization containers, towards seamless deployment of preconfigured bioinformatics software and pipelines on any computational platform.

We will present an approach for abstracting the complex data operations of multi-step, bioinformatics pipelines for NGS data analysis. As examples, we have deployed two pipelines for RNAseq and CHIPseq, pre-configured within Docker virtualization containers we call Bio-Docklets. Each Bio-Docklet exposes a single data input and output endpoint and from a user perspective, running the pipelines is as simple as

running a single bioinformatics tool. This is achieved through a 'meta-script' that automatically starts the Bio-Docklets, and controls the pipeline execution through the BioBlend software library and the Galaxy Application Programming Interface (API). The pipelne output is postprocessed using the Visual Omics Explorer (VOE) framework, providing interactive data visualizations that users can access through a web browser.
The goal of our approach is to enable easy access to NGS data analysis pipelines for nonbioinformatics experts, on any computing environment whether a laboratory workstation, university computer cluster, or a cloud service provider. Besides end-users, the Bio-Docklets also enables developers to programmatically deploy and run a large number of pipeline instances for concurrent analysis of multiple datasets.
Web links and citations:
https://goo.gl/iL5bH6
https://goo.gl/AdKQF8

W635: Interoperability and Federation Across Bioinformatic Platforms and Resources
GAMER: GO Annotation Method, Evaluation, and Review
Kokulapalan Wimalanathan[1], Iddo Friedberg[1], Carson M Andorf[2] and **Carolyn J. Lawrence-Dill**[1], (1)Iowa State University, Ames, IA, (2)USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA
We created a new high-coverage, robust, and reproducible functional annotation of maize protein coding genes based on Gene Ontology (GO) term assignments. Whereas the existing Phytozome and Gramene maize GO annotation sets only cover 41% and 56% of maize protein coding genes, respectively, this study provides annotations for 100% of the genes. We also compared the quality of our newly-derived annotations with the existing Gramene and Phytozome functional annotation sets by comparing all three to a manually annotated gold standard set of 1,619 genes where annotations were primarily inferred from direct assay or mutant phenotype. Evaluations based on the gold standard indicate that our new annotation set is measurably more accurate than those from Phytozome and Gramene. To derive this new high-coverage, high-confidence annotation set we used sequence-similarity and protein-domain-presence methods as well as mixed-method pipelines that developed for the Critical Assessment of Function Annotation (CAFA) challenge. Our project to improve maize annotations is called maize-GAMER (GO Annotation Method, Evaluation, and Review) and the newly-derived annotations are accessible via MaizeGDB (http://download.maizegdb.org/maize-GAMER) and CyVerse (B73 RefGen_v3 5b+ at doi.org/10.7946/P2S62P and B73 RefGen_v4 Zm00001d.2 at doi.org/10.7946/P2M925). The GAMER pipeline can be used to functionally annotate genes from any plant species, enabling comparable GO annotations across many species and resources.

W636: Interoperability and Federation Across Bioinformatic Platforms and Resources
Benefits of Track Hubs and Byte-Range Requests
Brian Lee, UCSC, Santa Cruz, CA
Track Hubs are remotely hosted collections of text files describing features about underlying data files that can be visualized on the UCSC Genome Browser and other genome browsers. Assembly Track Hubs are an extension of Track Hubs that enable viewing a new custom assembly and related created data files. The data files for hubs must be hosted on a server, such as CyVerse, that accepts byte-range requests. The benefits of these data formats is that only portions of the file needed to display in a particular genomic region are transferred over the Internet and cached as a sparse file, allowing the display of vast collections of experimental data while transferring only specifically requested views.

W637: Interoperability and Federation Across Bioinformatic Platforms and Resources
An E-Infrastructure Model for the Plant Research Communities
Hadi Quesneville, URGI, INRA, Université Paris-Saclay, Versailles, France
With continuous advances in high-throughput technologies which generate an ever-growing amount of data, biological research is gradually entering a new data centric era. In order to face this data deluge, researchers, institutions and governments build computer infrastructures able to store, manage, share, and compute the data they produce. These computer infrastructure, called e-infrastructure or cyber-infrastructure, seek to meet the needs of large national or international research communities. First conceived as centralized infrastructure, they evolve toward federations of distributed resources.
We are currently building such e-infrastructures for (i) the wheat research community, and (ii) the plant community in the frame of the ELIXIR European infrastructure for life science. They evolve both towards a common scheme where these e-infrastructures are conceived as distributed information system relying on networks of bioinformatics platforms. These platforms, each considered as a node, share their resources and propose several dedicated integrative databases, e.g. for genomic, genetic, and phenotypic information, comparative genomics, and functional genomics. A portal acting as a hub provides centralized access to (i) the nodes and their resources, (ii) description of recommended data standards, (iii) a file repository to deposit and share data among the scientific community, and (iv) a data discovery search able to query distributed and heterogeneous databases. Data interoperability among the component of these e-infrastructures is a key issue that we address through web semantics, ontologies and data standards. We present how data in these infrastructures can achieve the FAIR principles to be Findable, Accessible, Interoperable, and Re-usable.

W638: Interoperability and Federation Across Bioinformatic Platforms and Resources
Challenges and Recommendations for Genomics, Genetics and Breeding Databases from the AgBioData Consortium
Monica Poelchau[1], Jacqueline D. Campbell[2], Ethalinda Cannon[2], Sook Jung[3], Dorrie Main[3], Ramona Walls[4] and Lisa C Harper[5], (1)USDA/Agricultural Resarch Service/National Agricultural Library, Beltsville, MD, (2)Iowa State University, Ames, IA, (3)Washington State University, Pullman, WA, (4)University of Arizona, Tucson, AZ, (5)USDA-ARS, Albany, CA
Genomics, Genetics and Breeding (GGB) databases are critical for the effective transmission of agricultural knowledge. However, to date there has been only limited movement toward interoperability standards among GGB databases. The AgBioData consortium (https://www.agbiodata.org/) was founded with the objective to work together to ensure standards and best practices for acquisition, display

and retrieval of genomic, genetic and breeding data. In April 2017, 45 AgBioData members convened to identify challenges faced by GGB databases and develop recommendations for future progress. Here, we present results from this meeting for six different domains relevant to GGB databases – Biocuration, Ontologies, Metadata, Database platforms, Data sharing, and Communication.

W639: Interoperability and Federation Across Bioinformatic Platforms and Resources
TBD
Sam Hokin, National Center for Genome Resources (NCGR), Santa Fe, NM

W640: Interoperability and Federation Across Bioinformatic Platforms and Resources
Building a Better Transcriptomics Pipeline using the CoGe API, CCtools WorkQueue and the JetStream Cloud
Asher K Haug-Baltzell[1], Eric Lyons[2] and Sean Davey[1], (1)University of Arizona, Tucson, AZ, (2)University of Arizona; BIO5 Institute; CyVerse, Tucson, AZ
High-throughput sequencing has pushed the idea of “big data” to the forefront of the discussion in many biological research programs. While this explosive growth of data will enable the answering of some of the most complex questions in biology, it brings with it additional challenges in designing experiments and workflows that can handle this new volume of information. Fortunately, rising to meet this challenge is a wealth of bioinformatics tools, pipelines, and platforms. However, while each of these resources offers unique strengths, researchers often want to take advantage of different components from various resources to answer their specific questions. As such, linking tools together via APIs has become of increasing interest.
One area of the sequencing boom that can benefit greatly from such linkages is transcriptomics. Not only are many new projects generating transcriptomic data to answer their questions, but an enormous amount of already generated and publically available data exists which can be used to complement newly generated data or mine for further discovery. Transcriptome-centered analysis pipelines are naturally parallelizable on multiple fronts: the variety of independent QC, assembly, and analysis steps, and the application across hundreds to thousands of samples, can all be separated relatively simply. Cloud computing resources, such as those offered by the NSF-funded JetStream cloud, can be tied into workflows using job distribution frameworks such as CCTools WorkQueue. Furthermore, the large volume of data storage required by these workflows can be eased by leveraging data services, such as those offered through the CoGe API. Visualization and results management (e.g. sharing) are also built into the CoGe API, and can be used to create a complete raw-data to discovery pipeline.
This tutorial-style talk will walk through the process of designing and implementing a distributed transcriptomics pipeline that takes advantage of powerful open-source projects for data management, job management, and results collection, visualization, and sharing. Many of the analysis tools will already be well-known to researchers familiar with transcriptome sequencing and analysis, and most discussion will center around leveraging offerings from diverse bioinformatics and cloud-computing platforms to improve the efficiency and usability of existing workflows.

W641: IRIC: Rice Informatics for the Global Community
2017 Update on IRIC Portal and Consortium
Nickolai Alexandrov, International Rice Research Institute, Los Baños, Philippines

W642: IRIC: Rice Informatics for the Global Community
DeepVariant -- Creating a Universal SNP and Indel Variant Caller with Deep Neural Networks
Ryan Poplin, Google, Inc., Mountain View, CA and Allen Day, Google, Seattle, WA
Next-generation sequencing (NGS) is a rapidly evolving set of technologies that can be used to determine the sequence of an individual's genome. The most common application of NGS technologies is to call genetic variants present in an individual using billions of short, errorful sequence reads. Despite more than a decade of effort and thousands of dedicated researchers, the hand-crafted and parameterized statistical models used for variant calling still produce thousands of false positive and false negative variants in each genome. Here we show that a deep convolutional neural network can call genetic variation in aligned next-generation sequencing read data by learning statistical relationships (likelihoods) between images of read pileups around putative variant sites and ground-truth genotype calls. This open-source tool, called DeepVariant, outperforms existing methods, even winning the "highest performance" award for SNPs in a FDA-administered variant calling challenge. The learned model generalizes across genome builds and even to other species, allowing non-human sequencing projects to benefit from the wealth of human ground truth data. We present validation results from applying DeepVariant to the Rice genome samples from the 3000 Rice Genome Project together with the International Rice Research Institute. DeepVariant is available as open-source on GitHub.

W643: IRIC: Rice Informatics for the Global Community
Improving Rice Genome Annotation Based on Public RNA-Seq Data
Lili Hao, BIG Data Center, Beijing, China
Rice is one of the most important staple food for a large portion of the world’s population and also a key model organism for cereal crops due to its great agricultural importance. In order to take maximum advantage of reference genome in extensive post-genomics studies, it is advisable to keep annotation information updating all the time by integrating the latest massive multiple-omics data. Here, we present a revision of the *Oryza Sativa Japonica* genome annotation: *BIGD_IC4R-1.0,* based on more than 500 recent available high-quality RNA-seq data sets along with annotation contributions from NCBI, EBI and UniProt, thereby providing substantive improvements over the previous version *MSU Rice Genome Annotation Project Release 7.0*. Our near-final release of the *BIGD_IC4R-1.0* is consisted of 57,905 protein coding genes, including 2,259 novel loci which do not overlap with the previous annotations. A total of 67 percent of these gene models are corroborated by evidence of expression and the structural annotation of over 20,682 loci in the *MSU7.0* has been updated. Moreover, the number of genes in *BIGD_IC4R-1.0* with splice variants is significantly increased compared with *MSU7.0*. In addition, 11,841 long ncRNAs were predicted from 658,655 StringTie and PASA assembled transcripts and then included in *BIGD_IC4R-1.0*. This updated genome re-

annotation has revised hundreds of incorrect predicted gene models in rice and provided a number of alternatively spliced isoforms as well as long ncRNAs, thus would hopefully promote the future functional genomics and transcriptomics studies in *Oryza* plants.

W644: IRIC: Rice Informatics for the Global Community
Speed Matters: Rapid Browsing across Multiple Genomes using Persephone
Max Troukhan, Persephone Software, LLC, Agoura Hills, CA
A multi-genome browser, called Persephone, is presented. The main advantages of the viewer are its speed, interactivity and the ability to handle large amounts of data in real time.
Persephone provides a framework for genomic data storing and visualization. It is a platform that includes a relational database designed for fast data retrieval, an interactive viewer with smooth animation, a loader application, and other services.
The main characteristic feature of Persephone is its ability to show multiple chromosomes on one screen. This allows interactive exploring of syntenic relations between genomic maps. A map can be based on a sequence or on a genetic linkage group. The maps can be aligned by connecting orthologous genes or common markers, thus allowing comparison of genetic and physical maps.
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. Applying proprietary algorithms to data compression has allowed 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 level of details shown on each map changes smoothly with varying zoom factor.
The live demo of Persephone will show several reference genomes of rice with data that allow comparison of accessions from the 3000 Rice Genomes project. A recently-published free web version of Persephone will also be introduced.

W645: IRIC: Rice Informatics for the Global Community
KnetMiner – an Integrated Data Platform for Gene Mining and Biological Knowledge Discovery
Keywan Hassani-Pak[1], Ajit Singh[1], Ramil Mauleon[2], Marco Brandizi[1] and Christopher Rawlings[1], (1)Rothamsted Research, Harpenden, United Kingdom, (2)International Rice Research Institute, Los Baños, Philippines
KnetMiner, with a silent "K" and standing for Knowledge Network Miner, is an integrated data and analytics platform for candidate gene mining and biological knowledge discovery. KnetMiner mines the myriad databases that describe an organism's biology to present links between relevant pieces of information, such as genes, biological pathways, phenotypes and publications with the aim to provide leads for scientists who are investigating the molecular basis for a particular trait. The approach is based on 1) integration of heterogeneous, complex and interconnected biological information into a knowledge graph; 2) text-mining to enrich the knowledge graph with novel relations extracted from literature; 3) graph queries of varying depths to find paths between genes and evidence nodes; 4) evidence-based gene rank algorithm that combines graph and information theory; 5) fast search and interactive knowledge visualisation techniques. Overall, KnetMiner is an open-source and public resource (http://knetminer.rothamsted.ac.uk) that helps scientists connect and visualise diverse biological databases for clues to design better crop varieties. The key strength of KnetMiner is to include the end user into the "interactive" knowledge discovery process with the goal of supporting human intelligence with machine intelligence. Here, we will present our recent work, in collaboration with IRRI, to build a rich KnetMiner resource the rice research community.

W646: IRIC: Rice Informatics for the Global Community
Rice Comparative and Functional Genomics at Syngenta
Matthieu Conte, Syngenta Seeds, Toulouse, France
To meet the growing global demand for agricultural crops like Rice it is crucial to understand the molecular components controlling crop productivity. In recent years, decreasing costs for whole genome sequencing and advances in high throughput data generation have resulted in a deluge of large, high-dimensional datasets for important agricultural crops. As a result, it has become increasingly difficult to maintain, integrate, and efficiently leverage this escalating amount of data across crops.
With fruitful collaborations with Academic institutes Syngenta is developing innovative processes and platforms utilizing a gene-centric data model, multiple reference genomes and comparative genomics to significantly improved data organization, knowledge capture, and data mining capabilities.
In this talk, I will focus on two systems that we have implemented to integrate data into a gene-centric information system (called GIN) and to share information across crops using a comparative genomics system (called GreenPhyl). I will describe how these systems have extended our knowledge of the Rice genome to support breeding projects at identifying gene candidates that potentially control important agronomic traits.

W647: IWGSC - Standards and Protocols
Can We Apply Lessons Learned from Manual Gene Annotation in Human and Mouse to Wheat?
Jane Loveland, EMBL-EBI, Cambridge, United Kingdom
The Ensembl-HAVANA team have significant expertise in manual genome annotation and over the last 15 years have been providing reference gene annotation for whole genomes (human, mouse and zebrafish), individual chromosomes (Pig chr X and Y), genes (Rat, Pig) and regions (MHC of Gorilla, Pig, Dog, Wallaby, Tasmanian devil) of community interest. Comprehensive manual annotation of high quality genomes is labour intensive and as such is not practical for very many genomes, however, automated gene annotation methods such as the Ensembl genebuild pipeline, can do a good job of a capturing the geneset, particularly protein coding genes. It is clear that experts in individual communities will want to improve the baseline automated annotation, for example to adequately capture their knowledge of functionally important genes or resolve annotation errors in complex regions such as gene clusters that present particular challenges for automated pipelines. We have a history of successful annotation workshops that have been co-ordinated by our team, namely for cow, pig and rat, where we provided training and annotation expertise to particular communities. As individual groups and communities create their own gene annotation, there is a danger that any divergence in their approach could hinder accurate downstream analysis both within and between species. For

example, the CCDS collaboration between ourselves and RefSeq was established to agree common annotation for at least one CDS in every protein-coding gene in the already well annotated human and mouse genomes. Despite the technical expertise in both groups and the wealth of available experimental data in these species, small differences in starting annotation guidelines led to significant differences in the annotated genes, requiring the resolution of many hundreds of annotation differences. We will present our guidelines and practices for annotation, based on our accumulated knowledge from producing reference gene annotation as framework that could be used to inform the approach of a community towards manual annotation, for example, by providing guidelines that can be used in a platform agnostic way to help inform decisions on annotating structural and functional information for genes and transcripts.

W648: IWGSC - Standards and Protocols

Getting to the Finishing Line: Integrating Optical and Physical Mapping for Megabase-Scale Resolution and Correct Annotation of Wheat Chromosome 7A

Rudi Appels, Murdoch University, Perth, Australia

Recent advances in whole-genome sequence assembly algorithms and chromosome conformation capture (Hi-C) have enabled the production of the first full-scale pseudomolecules in hexaploid wheat (IWGSC RefSeq v1.0). Resources developed by the IWGSC over the last decade, including BAC libraries and physical maps were critical in validating the sequence and extending scaffolds into super-scaffolds, tripling the assembly N50 from 7Mb to 21Mb in IWGSC RefSeq v1.0.

An independent assembly of chromosome 7A based on integrating a variety of datasets, including a BAC-based physical map, mate-pair sequencing, and optical mapping, resulted in the chromosome being assembled into 129 scaffold islands covering 735.1Mb. This assembly combined with the IWGSC RefSeq v1.0 enabled extensive validation as well as the elevation of regions of the chromosome from its existing high-quality draft status to finished status (less than one error per 100,000 base pairs). Integrating all available assembly resources, provided a complete classification of the chromosome into 17 contiguous regions with an N50 of 120Mb, the linear order of which could be independently validated using an 8-way MAGIC molecular genetic map. The value of fully validated sequence at long- and short-range is demonstrated using a number of regions of agronomic importance, including manual curation of the gene space.

W649: IWGSC - Standards and Protocols

Deep Learning Methods for Understanding Wheat Genomics Data

Ken Heyndrickx[1], Xi Wang[2], Thomas Janssens[3], Steve Robbens[3] and **Fred Van Ex**[4], (1)Bayer Cropscience, Diegem, Belgium, (2)Bayer Cropscience NV, Diegem, Belgium, (3)Bayer CropScience, Ghent, Belgium, (4)Bayer CropScience NV, Diegem, Belgium

Over the past few years both the volume and complexity of genomics and phenotyping data has exponentially increased. Today, the bottleneck is often no longer the ability to generate high throughput data sets required to solve specific biological problems, but rather the inherent limitations of conventional analysis methods to find structure and make accurate predictions based on large and complex data sets.

Deep Learning (DL) is a breakthrough technology in data science which shows great promise to help us overcome these limitations and increase the pace of research in biological sciences.

Bayer is actively researching applications of Deep Learning methods on diverse sets of complex data such as high throughput images and genomics data in order to accelerate our ability to improve crops. Rapid progress has been made on the development of deep learning approaches for annotation of high throughput phenotype data of a variety of wheat tissues. As a next step, Bayer is preparing to apply Deep Learning approaches on a diverse set of wheat genomics data in order to increase both the speed and accuracy of the functional annotation of the wheat genome.

W650: IWGSC - Standards and Protocols

Deciphering Structural Variations in the Wheat Genome using Resequencing Data

Romain De Oliveira[1], Emeric Dynomant[1], Hélène Rimbert[1], Hana Simkova[2], Jaroslav Dolezel[2], Federica Cattonaro[3], François Balfourier[1], Etienne Paux[1] and Frédéric Choulet[1], (1)GDEC, INRA, UCA, Clermont-Ferrand, France, (2)Institute of Experimental Botany, Olomouc, Czech Republic, (3)Istituto di Genomica Applicata, Udine, Italy

Structural variations (SVs) such as copy number and presence-absence variations (CNVs, PAVs) are polymorphisms that are known to be involved in the expression of phenotypes. In the absence of a reference genome sequence, their study has long been hampered in wheat. The recent advent of new wheat genomic resources has led to a paradigm shift, making it possible to investigate the extent of SVs among cultivated and wild populations. Our project aims at characterizing SVs in a *Triticeae* diversity panel of 44 accessions from seven tetraploid and hexaploid *Triticeae* species. To cope with the wheat genome complexity, we developed strategies combining shotgun sequencing of sorted chromosomes 3B with bioinformatics tools and we studied SVs affecting not only genes but also transposable elements (TEs). Our results show that 14% of the genes are variable within this panel. In addition, they reveal a very high level of intra- and interspecific variability affecting TEs, contrasting with the weak polymorphism rate usually reported with SNPs. Chromosomal extremities are the regions where we see most of the variability, confirming previous hypotheses made when comparing wheat with the other grasses.

W651: IWGSC - Standards and Protocols

The Intracellular Immune Receptor Repertoire of Wheat

Burkhard Steuernagel and Brande B.H. Wulff, John Innes Centre, Norwich, United Kingdom

With the release of the IWGSC reference sequence of bread wheat we are now in a good position to study gene families. The nucleotide binding and leucine-rich repeat (NLR)-encoding genes encompass one of the largest gene families in plants. These intracellular immune receptors perceive the presence of pathogens by detecting the presence of pathogen effectors. [perhaps a sentence here explaining why NLRs are difficult to annotate with conventional methods?] We developed NLR-Annotator, a tool for *de novo* annotation of loci associated with NLRs. NLR-Annotator can be applied directly on genomic sequence and is therefore independent from gene annotations. In the genome of the wheat accession Chinese Spring, we found 3,400 loci that may either be functional NLR genes or pseudo genes. Using our annotation, we can

now also evaluate the performance of gene annotation pipelines on the complex NLR family and devise new strategies to characterise this gene family with greater precision.

W652: IWGSC - Standards and Protocols

The IWGSC Data Repository and Wheat Data Resources Hosted at URGI: Overview and Perspectives

Michael Alaux[1], Jane Rogers[2], Thomas Letellier[1], Raphael Flores[1], Cyril Pommier[1], Nacer Mohellibi[1], Sophie Durand[1], Erik Kimmel[1], Célia Michotey[1], Mikael Loaec[1], Véronique Jamilloux[1], Mathide Lainé[1], Frédéric Choulet[3], Etienne Paux[3], Jérôme Salse[4], Kellye Eversole[5], Anne-Francoise Adam-Blondon[1] and Hadi Quesneville[1], (1)URGI, INRA, Université Paris-Saclay, Versailles, France, (2)International Wheat Genome Sequencing Consortium, Cambridge, United Kingdom, (3)GDEC, INRA, UCA, Clermont-Ferrand, France, (4)GDEC, INRA, UCA, Clermont-Ferrand, France, Clermont-Ferrand, France, (5)Eversole Associates, Bethesda, MD

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. Under the umbrella of the IWGSC (International Wheat Genome Sequencing Consortium), we have set up a Data Repository in the Wheat@URGI portal:
http://wheat-urgi.versailles.inra.fr/Seq-Repository
to store, browse and query the data being generated by the consortium and especially the wheat reference sequence (IWGSC RefSeq v1.0) and all related annotation data.
We will give an overview of the wheat genomic, genetic and phenomic data available and suggest ways to develop the database in the future.

W653: IWGSC - Standards and Protocols

Updating the IWGSC Refseq Genome Sequence and Annotation - a Discussion

Frederic Choulet, INRA GDEC, CLERMONT-FERRAND, France

W654: IWGSC - Standards and Protocols

IWGSC Phase II: What's Next for the IWGSC

Kellye Eversole, IWGSC, Bethesda, MD

In 2017, the International Wheat Genome Sequencing Consortium (IWGSC) achieved the first high quality, annotated reference sequence of bread wheat, IWGSC RefSeq v1.0. In phase II, the consortium will continue to deliver valuable tools and resources for wheat scientists and breeders through the functional annotation of the reference sequence and the generation of a pan genome that represents the breadth of diversity for bread wheat. An overview of the future activities of the IWGSC will be presented.

W655: JBrowse, a Next Generation Genome Browser

JBrowse Hands-On Workshop

Scott Cain, Ontario Institute for Cancer Research, Medina, OH

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.

W656: JBrowse, a Next Generation Genome Browser

JBrowse Workshop

Scott Cain, Ontario Institute for Cancer Research, Medina, OH

W657: Legumes

Exploiting *Medicago* Structural Variation to Discover Novel Genes for Nodulation.

Nevin D. Young[1], Shaun J Curtin[2], Joseph Guhlin[1], Diana Trujillo[3], Peng Zhou[4], Robert M. Stupar[5] and Peter L. Tiffin[6], (1)Department of Plant Pathology, University of Minnesota, St. Paul, MN, (2)Cereal Disease Laboratory, St. Paul, MN, (3)University of Minnesota, Saint Paul, MN, (4)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (5)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (6)Department of Plant Biology, University of Minnesota, Saint Paul, MN

Medicago truncatula is a widely-studied model for legume biology and genomics with multiple reference quality genomes, dozens of deeply sequenced accessions, and hundreds of survey-sequenced genotypes suitable for SNP discovery and GWAS. *M. truncatula* is especially valuable for studies of symbiosis, nodulation, and mycorrhization, leading to many of the seminal discoveries in plant-microbe communication. Nodulation in *Medicago / Ensifer (Sinorhizobium)* symbiosis is indeterminant, with nodulation relying on a large gene family – the nodule-specific cysteine-rich peptides (NCRs) – primarily found in *Medicago* and close taxonomic relatives. *De novo* assembly of 16 *Medicago* accessions enabled genome-scale exploration of structural variation (SV), especially the variant architecture of gene families important in

plant-microbe interactions. This study demonstrated that 22% of the genome is involved in large structural changes, 42% of reference genome is missing from one or more accessions, and gene families vary notably in their SV architecture. *Medicago* re-sequencing also enabled discovery of novel proteins important in nodulation. SNP-based GWAS of nodulation uncovered several promising candidates. Ten candidates were tested in carefully replicated, statistically robust validation experiments involving *Tnt1*, RNAi, TALEN, or CRISPR knockout / knockdown lines. Among the newly validated nodulation genes was *Pho2* (a modulator of phosphate). A parallel GWAS based on read depth variants revealed that NCRs in a cluster on chromosome 4 play an especially important role in nodulation. Finally, detailed analysis of gene family expansions specific to different lineages led to the discovery of *Medicago*-specific PLAT proteins. CRISPR-based knockouts of *Medicago* PLAT proteins confirmed their functional role in nodulation.

W658: Legumes
Can We Improve Alfalfa Yield? Lessons from Breeding and Genomics.
E. Charles Brummer, University of California, Davis, Davis, CA

W659: Legumes
Development of Sequencing for Genotyping within Soybean Breeding Programs.
David L. Hyten, University of Nebraska-Lincoln, Lincoln, NE

W660: Legumes
Gene Networks that Govern Soybean Seed Development
John Harada[1], Leonardo Jo[1], Julie Pelletier[1], Russell Baden[1], Jungim Hur[2], Min Chen[2] and Robert B Goldberg[2], (1)University of California, Davis, CA, (2)University of California, Los Angeles, CA
A unique aspect of seed development is that it is temporally biphasic. The earliest phase is morphogenesis during which the embryo and endosperm undergo regional specification and the establishment of subregions, tissues, and cell types, and the basic body plan of the plant is established. The maturation phase occurs late in seed development. During this phase, storage macromolecules including lipids and proteins, accumulate to massive amounts to serve as a food source for the seedling, and the embryo acquires the ability to withstand desiccation. Although the morphological and physiological events that occur during the morphogenesis and maturation phases have been characterized extensively, little is known of the gene regulatory networks that operate during seed development.
To dissect the processes that govern the maturation phase of soybean seed development, we queried the mRNA transcriptomes of every subregion of soybean seeds at several stages of development to identify transcription factors (TFs) that accumulate primarily in the embryo during the maturation phase. We focused on four of these TFs that are known to play critical roles in the maturation phase, ABI3, AREB3, bZIP67, and LEC1. We identified target genes that are directly regulated by these TFs in chromatin immunoprecipitation-RNA sequencing experiments. We showed that there is significant overlap in the target genes regulated by these TFs and that specific combinations of these TFs regulate genes involved in distinct biological processes, including maturation, photosynthesis, and morphogenesis. We discussed the implications of these results for understanding the processes that occur during seed development.

W661: Legumes
Could Small Seeded Wild Relatives of Cultivated Peanut be used to Increase the Size of Peanut Seeds?
Daniel Fonceka, Centre d'Etudes Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), Thies, Senegal

W662: Legumes
Advances in Genotyping Peanut, a Challenging Polyploid.
Josh Clevenger[1], Walid Korani[2], Ye Chu[3], Scott A. Jackson[1] and **Peggy Ozias-Akins**[2], (1)University of Georgia, Athens, GA, (2)University of Georgia Tifton Campus, Tifton, GA, (3)University of Georgia, Tifton, GA
Peanut (*Arachis hypogaea* L.), also commonly known as groundnut, is a polyploid crop whose diploid ancestors only recently (3 Mya) diverged from one another, and therefore retain high sequence similarity, particularly in coding regions of the genome. Sequence-based genotyping of peanut using standard bioinformatics tools for diploids, or polyploids with more highly divergent subgenomes, has resulted in a high frequency of false single nucleotide polymorphism (SNP) calls. A novel pipeline, SWEEP (Clevenger et al. 2015), was designed to distinguish between homeologous SNPs and true allelic SNPs using a sliding window approach. SWEEP increased the rate of true SNP calls from 10% to 40% based on large-scale validation with the Affymetrix Axiom_Arachis array. Further improvement to 75% positive calls was obtained using a machine-learning approach, validated with a Version 2 Axiom_Arachis array, and around 90% has been achieved with the newly developed HAPLOSWEEP. These improvements will now enable more reliable sequence-based SNP calling when sequencing is preferred over array genotyping to reduce ascertainment bias.

W663: Linkage and Deletion Mapping
Method for Computing Map Likelihood Given Marker Observations and an Arbitrary Population Structure
Damien Leroux, MIAT, Institut National de la Recherche Agronomique, Castanet-Tolosan CEDEX, France
Classically, genetic cartography has been achieved using simple population models which feature few independent meioses, such as F2, Backcross, or Outbred. More complex population structures have recently emerged, such as MAGIC, that render classic population models obsolete. We propose a method to account for the actual population structure under study and quickly compute the likelihood of a genetic map by first inferring once and for all the recombinant state of all gametes in a pedigree on each marker independently using a cunning variant of Bayesian methods, and then computing the global likelihood for a given set of marker order and 2-point distances using these one-point state probabilities.
We follow the pedigree to build our Bayesian Network. Such a network becomes very loopy (and hard to estimate by message-passing) as soon

as an individual has multiple descendants. Despite this loopy nature, the pedigree defines a strong causality. We overcome this loop issue by creating a recursive network called a Hierarchical Bayesian Network. This network can be solved exactly and reasonably fast. Once we extract the beliefs on each gamete (modeled as a selection variable) on each observed marker, it is straightforward to compute the likelihood of the set of gametes given a marker order and the corresponding inter-marker distances.
This method allows for massively parallel computations to increase speed, and also supports defining observational noise at various levels, per marker, per individual or generation, and so on…
We implement a proof of concept as part of our QTL analysis software, Spell-QTL.

W664: Linkage and Deletion Mapping
Rapid Marker Ordering Using Multidimensional Scaling
Christine A. Hackett, Biomathematics and Statistics Scotland, Dundee, United Kingdom
Modern genotyping techniques are producing increasingly high numbers of genetic markers that can be scored in experimental populations of plants and animals. Ordering these markers to form a reliable linkage map is computationally challenging and most effort has focussed on populations derived from diploid, homozygous parents. Here we explore the use of weighted multidimensional scaling (MDS) to order markers from a cross between autotetraploid parents, using simulated data and experimental data from a potato population. We compare different functions of the recombination fraction and LOD score to form the MDS stress function, and find that a LOD^2 weighting generally performs well, including when missing values and genotyping errors are present. We recommend an initial analysis using unconstrained MDS to give a rapid way to detect and remove problematic markers, followed by fitting a principal curve to give a reliable marker orders. This MDS approach can also be applied to experimental populations of diploids or higher polyploids.

W665: Linkage and Deletion Mapping
Linkage Analysis and Haplotype Phasing in Experimental High Autopolyploid Populations, with Applications in Hexaploid Sweetpotato
Marcelo Mollinari[1], Guilherme Da Silva Pereira[1], Bode Olukolu[1], Dorcus C. Gemenet[2], Awais Khan[2], Mercy N. Kitavi[3], David Maria[2], Marc Ghislain[3], Craig Yencho[1], Antonio Augusto Franco Garcia[4] and Zhao-Bang Zeng[1], (1)North Carolina State University, Raleigh, NC, (2)International Potato Center (CIP), Lima, Peru, (3)International Potato Center (CIP), Nairobi, Kenya, (4)University of São Paulo (ESALQ/USP), Piracicaba - SP, Brazil
Autopolyploid species are very important in agriculture and play fundamental role in evolutionary process. Despite all advances in genetic linkage analysis in autotetraploids, there is a lack of statistical models to perform linkage analysis in organisms with higher ploidy levels. We present a method to estimate recombination fractions and infer linkage phases in full-sib populations of autopolyploid species with high even ploidy levels in a sequence of SNP markers using hidden Markov models (HMM). Our method is based on the sequential addition of SNPs using efficient two-point procedures. The unsolved linkage phase configurations are evaluated using the multipoint HMM likelihood. To evaluate our method, and to show its properties, we used simulations of autotetraploid, autohexaploid and autooctaploid populations. We also construct a map of a hexaploid sweetpotato biparental population comprising 7,907 simplex SNPs, 1,665 double simplex and 6,535 in several multiplex configurations (16107 SNPs, in total) distributed in 15 homology groups with a total length of 1303.1 cM. The results demonstrate the reliability of the method, even for complex linkage phase scenarios. The method is implemented in the R package "polymap", available at https://github.com/mmollina/polymap.

W666: Linkage and Deletion Mapping
Lep-MAP3: Robust Linkage Mapping Even for Low-Coverage Whole Genome Sequencing Data
Pasi M Rastas, University of Helsinki, Institute of Biotechnology, Helsinki, Finland
Accurate and dense linkage maps are required for family-based linkage and association studies, quantitative trait locus (QTL) mapping, analysis of genome synteny and other analyses. Moreover, linkage mapping is one of the best tools to detect errors in *de novo* genome assemblies, as well as to anchor assembled contigs within chromosomes. Even a mapping cross of ten individuals will detect many assembly errors. With more individuals and markers, even more local errors can be detected and more contigs can be anchored. Linkage maps with more markers than recombinations have multiple markers at most map positions. This will anchor contigs more reliably by locally pinpointing each recombination. However, the tools that are currently available for linkage mapping are not well suited for very large number of markers nor individuals.
Here we present linkage mapping software Lep-MAP3, capable of analysing large datasets. It is fast and has small memory footprint, it can simultaneously analyse multiple families and requires little manual work and data curation. It can analyse low-coverage whole genome sequencing datasets on millions of markers and thousands of individuals. Such cost-efficient data enables comprehensive validation and refinement of genome assemblies. We demonstrate that Lep-MAP3 obtains very good performance already on 5x sequencing coverage and outperforms the fastest available software on accuracy and often on speed. We also construct *de novo* linkage maps with millions of markers on real 5-12x whole-genome sequencing data. Lep-MAP3 is freely available with the source code under GNU general public license from http://sourceforge.net/projects/lep-map3.

W667: Linkage and Deletion Mapping
Efficient Filtering of Markers for Building Ultra-High-Density Linkage Maps
Andrii Fatiukha, Institute of Evolution and the Department of Evolutionary and Environmental Biology, Faculty of Science and Science Education, University of Haifa, Haifa, Israel and Yefim I. Ronin, Institute of Evolution, University of Haifa, Haifa, Israel
The increase in the number of markers by orders of magnitude owing to new genomic technologies was perceived as a breakthrough that enables building quality ultradense genetic maps. This expectation may be far from reality due to genotyping errors and missing marker calls that may lead to wrong orders of markers. However, big datasets make it possible to apply filtering for building high-quality maps. In addition to the standard filters based on segregation ratios and missing level, we propose a pre-mapping selection of informative markers based on a

principle that less erroneous data should generate more reproducible local patterns than more erroneous data. In particular, for big datasets, a considerable part of markers should appear as groups of co-segregating markers (twin groups, TGs) or slightly dispersed "clouds" of markers resulted from twins due to single errors. We developed two procedures in situations of low (0.01–0.02) and high (0.02–0.04) rates of genotyping errors. The process of construction a skeletal map includes choosing delegate markers of TGs; testing map quality using jackknife resampling and deletion of markers violating local map stability; and insertion in the resulting map additional markers. In situations of insufficient number of TGs (at higher error levels), we employ marker clustering to extend the twin-based filtration idea. We select a representative marker from each cluster and use these representatives as additional candidate markers. Our system was implemented in MultiPoint-Ultradense software (Ronin et al. 2017, Genetics 206:1285) and tested on a variety of plant and animal organisms.

W668: Linkage and Deletion Mapping
Mapping 1000000 Markers with Flipper
Charles F. Crane, USDA-ARS, West Lafayette, IN
Flipper is a fast linkage and deletion mapper that implements Kruskal's algorithm to produce a minimum spanning tree of genetic markers. Additional heuristics allow Flipper to recognize and repair many instances of misplaced markers and misjoined subsets of markers. Flipper requires memory in proportion to the square of the number of markers. With 16 Gb of system memory, it cannot run more than about 43000 markers at once without paging to disk. Flipper allows efficient deduplication of marker sets to mutually recombining exemplar markers, and this can be used to reduce 1000000 perfectly genotyped markers to a sufficiently small number of exemplars. Since Flipper keeps track of the nonrecombining markers that accompany each exemplar, those can be inserted into the map based on exemplars. However, this method fails with even 0.1% genotyping error, since the errors introduce spurious recombination and greatly increase the number of apparent exemplars. A better, more general method is to map a subset of enough markers to place ca. 100 per linkage group, and then use this map as a set of mathematical attractors to cluster the entire marker set into linkage groups that might fit within the 43000-marker limit. Since each recombination fraction is calculated only once, this grouping of markers requires very little memory. The same idea can be applied to abutting subsets of markers within a single linkage group. The markers that fall into each non-overlapping cluster can then be mapped and the clusters can be sutured together with local permutation of 8-10 markers at each junction. An example of each of these three methods is given for a simulated genotyping-by-sequencing experiment with 1000000 markers.

W669: Maize
Origins of Temperate Adapted Maize
Kelly Swarts[1], Rafal Gutaker[1], Bruce Benz[2], Michael Blake[3], Robert Bukowski[4], James B. Holland[5], Melissa Kruse-Peeples[6], Nicholas K. Lepak[7], Lynda Prim[6], M. Cinta Romay[4], Jesus Sanchez-Gonzalez[8], Chris Schmidt[6], Verena J. Schuenemann[9], Johannes Krause[10], R.G. Matson[3], Detlef Weigel[11], Edward Buckler[12] and Hernan Burbano[1], (1)Research Group for Ancient Genomics and Evolution, Dept. of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany, (2)Dept. of Biology, Texas Wesleyan University, Fort Worth, TX, (3)Dept. of Anthropology, University of British Columbia, Vancouver, BC, Canada, (4)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (5)North Carolina State University, Raleigh, NC, (6)Native Seeds/SEARCH, Tucson, AZ, (7)USDA-ARS, Ithaca, NY, (8)Centro Universitario de Ciencias Biológicas y Agropecuarias, Zapopan, Mexico, (9)Institute of Archaeological Sciences, Tuebingen, Germany, (10)Max Planck Institute for the Science of Human History, Jena, Germany, (11)Max Planck Institute for Developmental Biology, Tübingen, Germany, (12)USDA-ARS, Cornell University, Ithaca, NY
People introduced maize to the Southwest US approximately 4,000 years ago, but full maize agriculture was not successful in the temperate uplands of the Colorado Plateau for another 2,000 years. Because early flowering characterizes modern temperate maize, we used a large inbred panel to predict days to flowering in an archaeological maize population from the dawn of temperate maize agriculture on the Colorado Plateau, inferring marginal adaptation. Cross-population predictions were validated on a population of descendant temperate and tropical Southwest US landraces with high predictive ability. Demographic modelling with modern outbred *Zea* and the archaeological population supported *in situ* adaptation to temperate environments based primarily on ancient standing variation. The impacts of this historical process continue to resonate in commercially important modern germplasm.

W670: Maize
Mining Big Data in Maize: From Genetics of Deleterious Alleles to Epigenetics of Methylation Variation
Jinliang Yang, University of Nebraska-Lincoln, Lincoln, NE
Genomic data have been generated in an unprecedented rate in the maize community. In order to repurpose the large amount of genomic data, we founded ZeaBigData.org to curate the existing datasets in maize. We characterized the deleterious alleles with the curated data and incoporated the evolutionary information into the genomic selection model to increase the prediction accuracy for phenotypic traits. We also work on epigenetics using whole genome bisulfite sequencing (WGBS). In maize, more than 40% of cytosines in the genome are methylated. Major progress has been made recently in describing the variation and functional consequences of DNA methylation across the genome, and also in understanding the molecular mechanisms driving methylation, yet the evolutionary forces shaping the variations and landscapes of DNA methylation are largely unknown. Here we conducted WGBS on 20 individuals sampled from a natural population of teosinte (Zea mays ssp. parviglumis) to investigate the evolutionary forces acting on methylation variation in nature. We inferred the methylation site frequency spectrum (mSFS) from the methylation data, and implemented a Markov Chain Monte Carlo (MCMC) approach to fit the data to an explicit population genetic model incorporating forward and backward mutation as well as natural selection. Preliminary analyses suggest evidence of selection to maintain gene body methylation, but that much of the intergenic variation in methylation may be largely neutral. Our results will enable a better understanding of the evolutionary forces acting on patterns of methylation in different contexts and functional regions of the genome.

W671: Maize

Impact of Microorganisms on Gene Expression and Traits
Slavica Djonovic, IndigoAg, Boston, MA

W672: Mango genomics
Sequencing, Assembly and Annotation of the Mango Cultivar 'Tommy Atkins'
David N. Kuhn[1], Michael Campbell[2], Ian Bally[3], Natalie Dillon[3], David Innes[4], Jordon Rahaman[5], Amy Groh[6], Barbara Freeman[7], Maria A. Islas-Osuna[8], Ron Ophir[9], Yuval Cohen[10] and Amir Sherman[11], (1)USDA ARS SHRS, Miami, FL, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)Queensland Department of Agriculture and Fisheries, Mareeba, Australia, (4)Queensland Department of Agriculture and Fisheries, Brisbane, Queensland, Australia, (5)Florida International University, Miami, FL, (6)USDA-ARS, Miami, FL, (7)USDA-ARS-SHRS, Coral Gables, FL, (8)Lab. de Biologia y Genetica Molecular (CIAD, A.C.), Hermosillo, Sonora, Mexico, (9)Agriculture Research Organization, Volcani Center, Bet Dagan, Israel, (10)Volcani Research Center, Rishon Lezion, Israel, (11)Agricultural Research Organization, The Volcani Center, Rishon Lezion, Israel
The mango (*Mangifera indica* L.) cultivar 'Tommy Atkins' is the most commonly available market mango in the United States, the world's largest importer of mango. Although 'Tommy Atkins' is a highly heterozygous cultivar, most available mango cultivars are also highly heterozygous, so 'Tommy Atkins' was chosen due to its commercial importance to the US market. Two different technologies were used for sequencing the mango genome, ~100x Illumina paired end coupled with 10X Genomics. Assembly of the data by NRGene resulted in 571 super-scaffolds and ~80% coverage of the haploid mango genome (~490Mb). The assembly was improved using two previously published genetic maps encompassing ~5,000 SNP markers and a recombination block map (NRGene) to allow correct ordering of scaffolds within a pseudomolecule and merging of unincorporated scaffolds into pseudomolecules. The final assembly was 20 pseudomolecules corresponding to the 20 chromosomes in the haploid genome. Pseudomolecules ranged in size from 11-22 Mb. A mango specific repeat library was created for masking and annotation was done with MakerP resulting in the identification of ~27,000 genes.

W673: Mango genomics
A Reference Genome Assembly of the Mango Variety Amrapali (*Mangifera indica* L.)
Nagendra K. Singh[1], Ajay Kumar Mahato[1], Pawan K. Jayaswal[1], Sangeeta Singh[1], Nisha Singh[1], Neera Yadav[1], Vandna Rai[1], Amitha Mithra S. V.[1], Kishor Gaikwad[1], Neha Sharma[2], Nimisha Sharma[2], Anand K. Singh[2], Manish Srivastava[2], Jai Prakash[2], Usha Kalidindi[2], S. K. Singh[2], Kasim Khan[3], Rupesh K. Mishra[3], Shailendra Rajan[3], Anju Bajpai[3], B.S. Sandhya[4], Puttaraju Nischita[4], Kundapura V Ravishankar[4], M.R. Dinesh[4], Neeraj Kumar[5], Sarika Jaiswal[5], Mir A. Iquebal[5], Dinesh Kumar[5], Anil Rai[5], Tilak R. Sharma[1] and Trilochan Mohapatra[6], (1)ICAR-National Research Centre on Plant Biotechnology, New Delhi, India, (2)ICAR-Indian Agricultural Research Institute, New Delhi, India, (3)ICAR-Central Institute for Subtropical Horticulture, Lucknow, India, (4)ICAR-Indian Institute of Horticultural Research, Bengaluru, India, (5)ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India, (6)Indian Council of Agricultural Research, New Delhi, India
Mango (*Mangifera indica* L.) is an important tropical fruit with high production volume; huge variability in look and taste but its long juvenile phase, high heterozygosity and lack of genomic resources has hampered the varietal improvement. We estimated a genome size of 402±10 Mbp by flow cytometry and assembled 403 Mbp of high-quality reference genome of cultivar 'Amrapali' arranged in 4,312 scaffolds (NCBI Acc. No. LMWC01000000). Eighty percent of the scaffolds were anchored to 20 mango chromosomes with 6,242 SLAF markers of a recently published high-density linkage map. Mapping of 18 different published transcriptome sequences on the assembly showed >96% coverage of the gene space, while BUSCO analysis showed 91.4% genome coverage. We annotated 46,395 protein-coding genes of which the highest 11,349 showed strong homology to *Citrus sinensis*. The mango genome comprised 181.75 Mbp (45.02 %) of repeat elements. A comprehensive analysis of non-coding RNA was also performed. Pairwise alignments of 20 mango chromosomes with each other revealed eighteen large segmental duplications, indicating at least a recent (15.87-31.74 Mya) and an ancient (253.96-269.84 Mya) whole genome duplication event. We identified by re-sequencing of 24 diverse mango cultivars 10.5 million high-quality SFPs and developed a 80K genic SNP genotyping chip, which was used for delineating population structure and genome wide association studies of fruit quality traits including fruit size and acidity in 368 mango genotypes. An accurate reference genome of mango will accelerate the development of stress tolerant desired quality mango varieties.

W674: Mango genomics
Genomics of Fruit Ripening Genes in Mango
Maria Islas-Osuna[1], David N. Kuhn[2], Adrian Ochoa-Leyva[3], Rogerio Sotelo-Mundo[4], Mitzuko Dautt-Castro[5], Carmen Contreras-Vergara[6] and Sergio Casas-Flores[5], (1)Lab. de Biologia y Genetica Molecular (CIAD, A.C.), Hermosillo Sonora, Mexico, (2)USDA ARS SHRS, Miami, FL, (3)IBT/UNAM, Cuernavaca Morelos, Mexico, (4)Lab. de Estructura Biomolecular (CIAD, A.C.), Hermosillo Sonora, Mexico, (5)IPICYT, División de Biología Molecular, Lab. Genómica Funcional y Comparativa, San Luis Potosi SLP, Mexico, (6)Lab. Biologia y Genetica Molecular (CIAD, A.C.), Hermosillo Sonora, Mexico
Based on the transcriptomes of mango cv. 'Kent' and 'Ataulfo', several transcripts were mapped against the 'Tommy Atkins' genome in order to detect polymorphisms among selected ripening genes. Interestingly, important genes that encode for the carbohydrate metabolic process such as polygalacturonases, xyloglucan endotransglucosylase hydrolases, α and β-galactosidases, pectate lyases, rhamnogalacturonate lyases, pectinmethyl esterases, β-glucanases, endoglucanases, expansins, α-amylases and α-L-arabinofuranosidases presented SNVs (single nucleotide variations). Also, genes from the carotenoid biosynthetic pathways, plant hormone signal transduction pathway and cytochrome P450 had SNVs. Hormone-related transcripts such as ethylene receptors and transcription factors also presented SNVs. The implications of the gene variation among domesticated mangoes will be discussed in this talk to further correlate known phenotypes in commercial cultivars with the genotype. Molecular breeding program of this important fruit is becoming reachable in the near future.

W675: Mango genomics

RNA-Seq Analysis of the Mango (*Mangifera indica L*) Fruit Epidermis: Elucidating the Molecular Mechanism of Cuticle Biosynthesis

Martín-Ernesto Tiznado-Hernández, Coordinación de Tecnología de Alimentos de Origen Vegetal, Hermosillo, Mexico and Tafolla-Arellano, Julio César, Zheng, Yi; Sun,Honghe; Jiao, Chen; Ruiz-May, Eliel; Hernández-Oñate, Miguel A.; González-León, Alberto; Báez-Sañudo, Reginaldo; Fei, Zhangjun; Domozych, David; K.C. Rose, Jocelyn

W676: Mango genomics

A Fruitful Phylogeny: Revealing the Evolutionary Relationships of *Mangifera* Species

Emily Warschefsky, Florida International University, Miami, FL and Eric Bishop-von Wettberg, University of Vermont, Burlington, VT

Crop wild relatives have recently garnered much attention for their potential as reservoirs of genetic diversity that can be used for crop breeding and improvement. In tree crops, such as mango (*Mangifera indica* L.), these wild relatives are of additional value as potential rootstocks onto which cultivars can be grafted. Because graft compatibility between two species depends on a complex suite of characteristics, including genetic similarity, rootstock selection can be informed by a systematic understanding of the crop and its wild relatives.

The mango genus, *Mangifera,* is one of the largest in the family Anacardiaceae, with the latest revision including 69 species spanning from Eastern India to the Solomon Islands. In addition to the economically important *M. indica*, some 26 other species are regionally cultivated for their edible fruits, and many others present potentially beneficial traits such as salinity tolerance and disease resistance. We have used restriction-site associated DNA sequencing to produce the first comprehensive and multilocus phylogeny of *Mangifera*. Using samples from 201 individuals representing approximately 36 species, we recover five main clades of *Mangifera* and its sister genus *Bouea*, and find *Mangifera,* as traditionally circumscribed, to be polyphyletic. We identify a clade of species that are closely related to *M. indica* and could serve as important resources for breeding programs, but additionally note that each of the five major clades contain both wild and cultivated species. Therefore, *Mangifera* represents a promising system in which to study, in parallel, the wild to domesticated transition in woody perennial species.

W677: Mango genomics

Organ-Specific and Temporal Expression Profiles of Auxin-Related Genes during Mango Fruitlet Drop

Youlia Denisov[1], Shani Glick[1], Tali Zviran[1], Mazal Ish-Shalom[1], Adi Doron-Faigenboim[2], Yuval Cohen[1] and **Vered Irihimovitch**[1], (1)Volcani Research Center, Rishon Lezion, Israel, (2)Department of Vegetable and Field Crops, ARO, Rishon LeZion, Israel

In mango, natural fruitlet abscission leads to yield loss. Fruitlet abscission initiates with a decrease in polar auxin transport through the abscission zone (AZ), triggered by ethylene. To explore the molecular components affecting this process, a Fluidigm™ array was used to analyze the expression patterns of distinct indole-3-acetic acid (IAA)-related genes, comparing control vs. ethephon-treated pericarp and AZ profiles. Over 48 h, the accumulation of *MiPIN1* and *MiLAX2* IAA-carrier genes decreased in both control and treated tissues. Nevertheless, ethephon-treated tissues displayed significantly lower levels of these transcripts within 4–18 h. Ethephon treatment also induced early and pronounced downregulation of distinct auxin-response factors and of 10 out of 16 IAA-responsive genes, contrasting with significant upregulation of *Gretchen Hagen3* transcript (*MiGH3.1*) encoding an auxin–amino synthetase. For both control and treated AZ, the decrease in IAA-carrier transcripts was associated with a decrease in IAA content and an increase in IAA–Asp:IAA ratio, suggesting that fruitlet drop is accompanied by formation of this non-hydrolyzed IAA–amino acid conjugate. Despite these similarities, ethephon-treated AZ displayed a sharper decrease in IAA content and higher IAA–Asp:IAA ratio within 18h. Lastly, the response of IAA-related genes to an exogenous IAA treatment was also examined. Our results will be discussed, highlighting the roles that distinct IAA-related genes might assume during mango fruitlet drop.

W678: Mango genomics

The Kensington Pride Mango Genome

Natalie Dillon, Queensland Department of Agriculture and Fisheries, Mareeba, Australia

W679: Meiotic Recombination

The Meiotic DSB Landscape in Arabidopsis

Ian R Henderson, University of Cambridge, Cambridge, United Kingdom

Plant genomes can be broadly divided in euchromatin and heterochromatin, which are cytologically defined, and generally correspond to gene versus transposon rich regions. We now appreciate that epigenetic modifications of the genome, for example DNA methylation, underlie differentiation of the genome into these different chromatin states. In addition to associating with different patterns of transcription, it is known that plant heterochromatin is typically also silenced for recombination during meiosis. In many of the large grass genomes, including wheat, the majority of the chromosomes consist of non-recombining expanses of heterochromatin, which can cause significant limitations for breeding. Our research investigates the genetic and epigenetic factors that shape recombination in plant genomes. I will present new data where we have profiled Arabidopsis recombination factors genome-wide, which has revealed hotspots associated with genes and also, surprisingly, DNA transposons. I will discuss the implications of these finds for plant genome evolution and the relationship between genes and transposons. I will also present new work profiling chromatin states in the hexaploid wheat genome and show how this correlates with recombination. In summary I will explore the relationships between chromatin, transcription and recombination, with implications for the stability of plant chromosomes and how we improve crops via breeding.

W680: Meiotic Recombination

The Genomic and Epigenomic Features of Meiotic Crossover Hotspots in Plants
Alexandre Marand, University of Wisconsin-Madison, Madison, WI

W682: Meiotic Recombination
Characterization of Meiosis and the *Pairing Homoeologous 1* (*Ph1*) Locus in Wheat
Sateesh Kagale, National Research Council Canada, Saskatoon, SK, Canada
Meiotic recombination between related but diverged sequences (homoeologous recombination) influences genome stability in polyploid crop species and the ability to introgress desirable traits through inter-specific crosses. Disruption of genetic barriers, such as sequence divergence and strict regulation of chromosome pairing, is the key to enable 'genetic accessibility' of natural variation and introgression of favorable traits from related or wild species into polyploid crops. In wheat, a single locus on Chromosome 5B known as *Ph1* (*pairing homoeologous 1*) controls orderly pairing of homologous chromosomes during meiosis. A *Ph1* deficient stock in Chinese Spring (CS-*Ph1b*) produced through radiation treatment has been used for inducing homoeologous recombination between wheat chromosomes and their alien homoeologues. Using the reference genome sequence of wheat and RNA sequencing of meiocytes from CS and Cs-*Ph1b*, we have identified the complete repertoire of genes in the *Ph1* locus region. Comprehensive structural and expression analysis of these genes suggest *Ph1* is a complex locus carrying multiple candidate genes with redundant functions. These findings along with our efforts towards (1) characterizing gene regulatory networks that are specifically involved in complex chromosome pairing behavior and subsequent recombination initiation during meiosis in wheat, (2) suppression of *Ph1* through mutagenesis in an elite wheat cultivar, and (3) developing pollen based single cell genomic sequencing approach for monitoring recombination frequency in F_1 wheat plants will be discussed.

W683: Meiotic Recombination
Live Imaging of Arabidopsis Meiosis
Arp Schnittger, University of Hamburg, Hamburg, Germany

W684: Meiotic Recombination
Variation in the Genomic Recombination Landscape in Adaptively Diverging Sticklebacks
Felicity Jones, Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany

W685: NCBI Genome Resources
What's New in the NCBI Eukaryotic Genome Annotation Pipeline
Francoise Thibaud-Nissen, National Center for Biotechnology Information (NCBI), Bethesda, MD and the Eukaryotic Genome Annotation Team, NCBI/NLM/NIH, Bethesda, MD
The NCBI Eukaryotic Genome Annotation Pipeline (www.ncbi.nlm.nih.gov/genome/annotation_euk/) has been used to annotate over 400 organisms, including plants and animals of agricultural importance. The pipeline provides content for various NCBI resources including RefSeq sequence databases, Gene, BLAST databases and the Genome Data Viewer.
The pipeline uses a modular framework for the automated execution of all annotation tasks from the fetching of genome assemblies to the loading of RefSeq-accessioned annotation products to public databases. Protein-coding and long non-coding RNA are predicted by Gnomon, an alignment- and HMM-based gene prediction program developed at NCBI, using the alignments of proteins, ESTs, full-length cDNAs and RNA-Seq data. Since November 2017, rRNAs, snRNAs and snoRNAs are also annotated by applying RFAM HMM models to the genome. A final set of models is selected for each genomic location by applying rules of precedence to conflicting or overlapping curated RefSeq alignments, and Gnomon and RFAM models. Finally, protein models are named based on orthology to model organisms or Blast hits to SwissProt/UniProtKb. The annotated genomic sequences products are available for download and distributed through various NCBI resources including sequence databases, Gene, BLAST and genome browsers. A web page summarizing the annotation results is also provided with each annotated genome.
This talk will provide a high-level view of the annotation pipeline. We will show how we assess the quality of the annotation products that we generate, and how contiguity and accuracy of genome assemblies affect the results. See all eukaryotes annotated by the NCBI Eukaryotic Annotation Pipeline at: http://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/ .

W686: NCBI Genome Resources
NCBI's Gene and Expression Profiles from SRA
Lukas Wagner and Terence D. Murphy, National Center for Biotechnology Information (NCBI), Bethesda, MD
Analyzing gene expression across a range of tissues or developmental stages can be highly informative for understanding gene function. RNA-seq on the Illumina platform has become a favorite method for assessing gene expression, with an abundance of data available in the public SRA archives. NCBI's eukaryotic genome annotation pipeline makes extensive use of public RNA-seq data for computing gene annotations, typically using alignments of one to 20 billion RNA-seq reads from dozens to hundreds of samples to refine the gene models. In addition to dramatically improving annotation quality, these data are highly informative for understanding gene expression.
NCBI is now computing gene expression profiles from RNA-seq alignments for select organisms. The initial data release in February 2017 included human, mouse, and rat, and the profiles are being expanded to key organisms including pig, cow, corn, rice, and others of interest to the plant and animal genomic communities. NCBI's Gene resource has been enhanced to display these expression profiles and also allow basic searches for genes with particular expression patterns. RNA-seq coverage graphs for individual samples are also available in the browser display.
Many organisms have RNA-seq data from hundreds to thousands of additional samples available in the public SRA archives. Querying SRA effectively and submitting your data to SRA are both tasks that can be straightforward. Having an understanding of how typical sequencing efforts are represented in SRA and in related databases is helpful for useful queries. Similarly, submitting an RNA-Seq study with as complete a description as possible in the metadata allows other researchers to identify, explore and perhaps cite the work.

W687: NCBI Genome Resources

NCBI Assembly: All You Need to Know about Submitting and Retrieving Genome Assemblies

Francoise Thibaud-Nissen[1], Avraham Kimchi[2], Karen Clark[3], Terence D. Murphy[1] and Paul Kitts[2], (1)National Center for Biotechnology Information (NCBI), Bethesda, MD, (2)NIH/NLM/NCBI, Bethesda, MD, (3)National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD

The NCBI Assembly resource (https://www.ncbi.nlm.nih.gov/assembly) is a collection of genomic assemblies for eukaryotes and prokaryotes, that have been submitted to GenBank, EBI or DDBJ. It encompasses near-finished as well as draft assemblies, each defining an unambiguous set of sequences. The Assembly data model allows the update and tracking of successive versions of a sequenced genome, and supports complex assemblies such as diploid or polyploid assemblies. The model also includes numerous attributes that allow searching and filtering, and closely integrates Assembly with other NCBI resources such as BioProject and BioSample.

Assemblies are submitted to GenBank through the submission portal (https://submit.ncbi.nlm.nih.gov/subs/genome/). We will discuss efforts to streamline the submission process, including the option to submit sequences with gaps, and the option to submit annotation in GFF format. We will also present how to download assembly data, one by one or in bulk, and the variety of formats in which the data is available.

W688: NCBI Genome Resources

NCBI's Genome Data Viewer, Remap, and Other Tools for Navigating an Abundance of Genomic Information

Terence D. Murphy, Anatoliy Kuznetsov, Peter Meric and Valerie A. Schneider, National Center for Biotechnology Information (NCBI), Bethesda, MD

NCBI provides a diverse set of powerful tools to aid genomics researchers in exploring both public and personal datasets.

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 genomic, GEO tracks, aligned data in SRA, and more. Users can also upload their own tracks in a variety of formats (BED, GFF3, GTF, VCF, GVF, WIG, BAM), with the data privately stored on NCBI's servers, or work with remote BAM files from other servers.

NCBI's Genome Workbench (https://www.ncbi.nlm.nih.gov/tools/gbench/) provides many of the visualization tools of Genome Data Viewer and more in a standalone application available for Mac, Windows, and Linux. Users can work with local data or data from NCBI's servers, and use a suite of visualization and analysis tools including graphical and text views, multiple alignment and tree building, and a variety of BLAST and other alignment options.

NCBI's Remap service (https://www.ncbi.nlm.nih.gov/genome/tools/remap) is a tool that allows users to project annotation data from one coordinate system to another, such as when an upgraded assembly version or an assembly for another breed or cultivar becomes available. The service works with a variety of formats (BED, GFF3, GTF, VCF), and can run from either a web interface or through an API. Assembly-assembly alignments and an alpha version of a standalone Remap application are also available. The assembly-assembly alignments can also be visualized in Genome Data Viewer, allowing users to visually explore similarities and differences in two assemblies.

W689: NCBI Genome Resources

Fast BLAST Tools for Biologists; MagicBLAST and other Next Generation Tools for Computational Biologists and Developers!

Ben Busby, NCBI/NLM/NIH, Bethesda, MD

Ben will discuss some new developments in BLAST that can help biologists identify orthologs more quickly. He will then move to command line BLAST tools such as MagicBlast that allow researchers to easily search through thousands of next generation sequencing (NGS) datasets! He will show examples of several bioinformatics pipelines built on the MagicBlast pipeline. He will then move to visualization tools for computational biologists and developers, starting with web-based .bam visualization, moving to sequence viewer embedding. Rounding out the discussion of Biological Data Science, he will present APIs that assist in large-scale comparative genomics!

W690: New Approaches for Developing Disease Resistance in Cereals

Stacking Multiple Stem Rust Resistance Genes at a Single Locus for Durable Resistance in Wheat

Ming Luo[1], Liqiong Xie[2], Sambasivam Periyannan[3], Rohit Mago[1], Burkhard Steuernagel[4], Brande Wulff[4], Brian J. Steffenson[5], T Lynne Reuber[6], Evans Lagudah[1] and Michael Ayliffe[1], (1)Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia, (2)Xinjiang University, China, (3)Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia, (4)John Innes Centre, Norwich, United Kingdom, (5)University of Minnesota, St. Paul, MN, (6)Two Blades Foundation, San Mateo, CA

Stem rust disease caused by the fungal pathogen Puccinia graminis f.sp tritici is a significant threat to global wheat production. The most cost effective way to control this disease is by genetic resistance. However major gene resistance is often readily overcome by pathogen evolution when resistance genes are deployed singularly. Combining major resistance genes is believed to extend their durability as multiple mutations are required in the pathogen to overcome this polygenic resistance. Major resistance genes can be combined by conventional breeding however this is a labour intensive process and resistance gene combinations are difficult to maintain in segregating families. In this study, we have used cloned stem rust resistance genes Sr22, Sr35, Sr45, Sr46 and Sr50 and the multi-pathogen adult plant rust resistance gene Sr55/Lr67/Yr46/Pm46/Ltn3 to produce binary vectors containing combinations of these genes. Constructs containing either 3, 4, 5 or 6 rust resistance genes were produced and transformed into bread wheat by Agrobacterium transformation. Molecular/genetic analysis demonstrated that some transgenic wheat lines contain all the resistance genes present in the binary vector used for transformation (ie. up to six) and these genes are inherited as a single locus in progeny plants. Transgenic plants are resistant to wheat stem rust disease with resistance co-segregating with the multigene transgenic locus.

W691: New Approaches for Developing Disease Resistance in Cereals

Integrating Genomic Selection in Breeding for Resistance to Rusts and Foliar Diseases in Wheat

Fnu Philomin Juliana, CIMMYT, El Batan, Mexico

Genomic selection is a promising technology that could increase genetic gains for quantitative disease resistance and help eliminate susceptible lines, before costly disease screening. To evaluate the potential integration of GS as a breeding tool, we tested genomic prediction for several diseases in CIMMYT's 1st year yield trials (YT) and 2nd year elite yield trials (EYT), from 2015-2016 and 2016-2017. While the YTs comprised about 9,000 lines, the EYTs were a subset comprising 1,092 lines. All lines were genotyped using genotyping-by-sequencing and the YTs were phenotyped for response to Ug99 stem rust (SR) race in Njoro, Kenya. The EYTs were phenotyped for SR in Njoro; yellow rust (YR) in Ludhiana, India; Fusarium head blight (FHB) in El Batan, Mexico; Septoria tritici blotch (STB) in Toluca, Mexico and spot blotch (SB) in Agua Fria, Mexico. The maximum within-nursery and across-nursery prediction accuracies were 0.74 and 0.60 for SR, 0.59 and 0.50 for YR, 0.42 and 0.21 for FHB, 0.50 and 0.18 for STB and 0.56 and 0.37 for SB, respectively. We also observed that at different selection intensities, GS could discard upto a maximum of 92% of the susceptible lines discarded by PS and select upto 61% of the resistant lines selected by PS, within nurseries. However, when selections were made across-nurseries, GS could discard 73.8-90.2% of the susceptible lines and select 24.5-61.6% of the resistant lines. While these results are promising, further efforts to improve prediction accuracies are crucial for the successful integration of GS in wheat disease resistance breeding.

W692: New Approaches for Developing Disease Resistance in Cereals

Structure-Guided Engineering of Synthetic Immune Receptors Against the Blast Fungus

Thorsten Langner[1], Abbas Maqbool[1], Izumi Chuma[2], Joe Win[3], Ryohei Terauchi[4], Mark Banfield[3] and Sophien Kamoun[5], (1)The Sainsbury Laboratory, Norwich, United Kingdom, (2)Graduate School of Agricultural Science, Kobe University, Kobe, Japan, (3)John Innes Centre, Norwich, United Kingdom, (4)Iwate Biotechnology Research Center, Kitakami, Japan, (5)The Sainsbury Laboratory, Norwich, England

Plant pathogens secrete a plethora of effector proteins to enable colonization of their hosts. These effectors interact with intracellular plant proteins to alter their function and promote infection. Plants are generally effective at fighting off pathogens and have evolved an effective immune system, including immune receptors of the nucleotide-binding, leucine-rich repeat proteins (NLR) class. However, NLRs tend to have a narrow recognition spectrum limiting their value in modern agriculture. Here, we present a strategy to improve NLR-mediated plant immunity using structural information of effector-target complexes.

The fungus *Magnaporthe oryzae* (syn. *Pyricularia oryzae*) is one of the most devastating plant pathogens causing blast disease on a wide range of monocot hosts, including rice, wheat, and millet. The *M. oryzae* effector AVR-Pik is recognized by the rice NLR pair Pik1/2 through binding to a heavy metal associated (HMA) domain that has integrated into Pik-1. We identified a sequence related effector, APikL2 (AVR-Pik like 2), which is conserved in nearly all *M. oryzae* isolates. Similar to AVR-Pik, APikL2 binds to HMA containing proteins and structural analyses revealed a common HMA-binding interface between these two effectors. However, APikL2 is not recognized by the NLR pair Pik1/2 and because it is widespread in *M. oryzae* is a high value target for blast disease resistance development. We combined sequence alignments and structure-based information derived from effector-target complexes to identify polymorphic residues around the effector-HMA binding interface that could define binding specificity of the NLR. We then introduced these residues into the HMA-domain of the Pik1 NLR to generate synthetic immune receptors that bind APikL2 and thus carry expanded effector binding spectra. Our work highlights how basic understanding of the biochemical and biophysical basis of pathogen-host interactions can be used to retool plant immunity and generate novel immune receptor functionalities.

W693: New Approaches for Developing Disease Resistance in Cereals

The Discovery of Decoy Domains in NLR Receptors Provides Novel Insight into Plant Immunity and Opens New Perspectives for Plant Protection

Thomas Kroj, INRA, Biology and Genetics of Plant-Pathogen Interactions Laboratory, Montpellier, France

Nucleotide-binding domain and leucine-rich repeat containing proteins (NLRs) are important receptors in plant immunity and allow specific recognition of pathogen effectors. Based on our work on the detection of the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by the rice NLR RGA5, we recently developed the hypothesis that some NLRs recognize effectors by integrated decoy domains This 'integrated decoy model' was further supported by work in other experimental systems and has been widely accepted. Comparative genomic analysis showed that NLRs carrying integrated decoy domains are frequent and widespread. We identified them in 31 land plants, from mosses to angiosperms, and they represent, on average, 7% of the NLRs.

By detailed structure-function analysis we further deciphered the molecular details of the binding of AVR-Pia and AVR1-CO39 to the integrated decoy domain of RGA5, a heavy metal-associated domain most related to the yeast copper chaperon ATX1 (RATX1 domain). This demonstrated that the direct RGA5-RATX1/effector binding is strictly required for effector recognition but only of moderate affinity and acts in concert with the association of the effectors to additional sites in RGA5. This combination of integrated decoy domains with additional independent effector-NLR interactions seems to confer robust effector recognition that is resilient to effector mutations. We will present first results on how knowledge on the molecular details of effector recognition by integrated decoy domains can be exploited for the engineering of the recognition spectrum of NLRs.

W694: New Approaches for Developing Disease Resistance in Cereals

Deep Learning for Image-Based Detection of Northern Leaf Blight in Maize

Chad DeChant[1], Tyr Wiesner-Hanks[2], Siyuan Chen[3], Ethan Stewart[2], Jason Yosinski[4], **Michael A. Gore**[2], Rebecca J. Nelson[5] and Hod Lipson[6], (1)Department of Computer Science, Columbia University, New York, NY, (2)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (3)Department of Mechanical Engineering, Columbia University, New York, NY, (4)Uber AI Labs, San Francisco, CA, (5)Plant Pathology and Plant-Microbe Biology Section, School

of Integrative Plant Science, Cornell University, Ithaca, NY, (6)Department of Mechanical Engineering and Institute of Data Science, Columbia University, New York, NY
Globally, it is estimated that more than 10% of crop yields are reduced by plant diseases each year. In the United States and parts of Canada, northern leaf blight (NLB), a fungal foliar disease of maize caused by *Setosphaeria turcica*, has progressively become more severe in the past 5 years, with economic losses in maize yield from NLB estimated at more than $1.9 billion in the most severe year. In maize, breeding for genetic resistance is the most effective and economical approach for control of NLB. When breeding for NLB resistance, the visual scoring of gray-brown necrotic NLB lesions at multiple time points throughout the growing season is essential, but this effort is very time-consuming and prone to discrepancies between different raters. In that light, our team of plant pathologists, plant geneticists, and computer scientists are working to develop a non-destructive, image-based phenotyping system that allows for rapid and accurate detection of NLB lesions (presence/absence) and estimation of diseased leaf area (DLA) under field conditions. In the first iteration of this system, a three-stage image analysis pipeline that integrated convolutional neural networks (CNNs)—deep learning models used for image recognition and classification—was developed through training on a large manually annotated image data set, then deployed to detect the presence/absence of NLB lesions in diverse field images collected with a hand-held camera. Overall, the system attained an accuracy of 96.7% on test set images held out from training. For the next iterations of the system, we are striving to enable NLB lesion detection and DLA estimation through deployment on aerial- and ground-based vehicles.

W695: New breeding technologies: Prospects and regulatory hurdles
Double-Strand Break Induced Genome Engineering in Plants: Past, Present, Future
Holger Puchta, Karlsruhe Institute of Technology, Karlsruhe, Germany
In the past we could show that sequence-specific nucleases can be used for inducing controlled changes into plant genome and to study double-strand break repair mechanisms. In the last years CRISPR/Cas system became a major tool and the *Streptococcus pyogenes* (Spy)Cas9 nuclease was used extensively for plant genome engineering. By the introduction of point mutations it is possible to transform this nuclease into a nickase that is inducing single strand breaks (SSBs) into DNA or a nuclease-dead DNA binding protein. Off-target effects might be avoided using a Cas9 nickase and two sgRNAs, to induce adjacent single strand breaks (SSBs) in opposite strands in the plant genome. This “paired nickase” strategy has a mutagenic potential comparable to the nuclease. Interestingly; sequence duplications are a prominent outcome of this approach, hinting to the possibility that the repair of adjacent SSBs is a major cause of sequence duplications during genome evolution of plants. We could also show that the nuclease of the Cas9 orthologue *Staphylococcus aureus* (SauCas9) is even more efficient in targeted mutagenesis in *A. thaliana* than SpyCas9. Using nuclease dead fluorescence fusions of both proteins we together with the group of Andreas Houben from the IPK Gatersleben were able to perform sequence specific visualization of repeated sequences in the Arabidopsis genome. The simultaneous use of different Cas9 orthologues will offer the opportunity to detect or control genetic information of plant cells on more complex levels.

W696: New breeding technologies: Prospects and regulatory hurdles
DNA-Free Genome Editing in Crops – GMO or Mutation
Janina Metje, Julius Kuehn-Institut, Quedlinburg, Germany
The CRISPR/Cas9 system has become a fast, easy and widely used genome editing tool and was entitled „breakthrough of the year 2015“ by Science journal. Still a wide application in crops is limited by constraints to plasmid-mediated delivery and regulatory restrictions. Typically, RNA-guided endonucleases (RGENs) are delivered into plant cells by transfection with plasmids or by *Agrobacterium tumefaciens* mediated T-DNA transfer. These methods induce stable expression of the CRISPR/Cas9 system in the host, which increases the chance of unwanted off-target effects. Furthermore, the system goes along with a possible integration of recombinant DNA and therefore the existence of transgenic plants [as intermediates]. Removal of that foreign DNA is not always possible e.g. in plants that reproduce asexually. Plants produced by conventional CRISPR/Cas9 systems will feasibly fall under the GMO law and are restricted by lack of acceptance and constraint and costly regulations.
Therefore, new genome editing methods are needed without the introduction of foreign DNA. In a new DNA-free genome editing system preassembled Cas9 protein-guide RNA ribonucleoproteins (RNPs) are directly delivered to the plant cell in a vector-free manner. RGEN RNPs targeted mutagenesis is highly efficient. RNPs are demonstrated to act immediately upon delivery and are degraded rapidly in the cells. The short activity period greatly decreased chance for off-target effects. Mutants obtained by this method are completely transgene free and are indistinguishable from naturally occurring genetic variations. Plants produced by DNA-free genome editing could be exempt from current GMO regulations.

W697: New breeding technologies: Prospects and regulatory hurdles
The Importance of Plant Breeding Innovation for the EU Seed Sector - Political, Regulatory and Communicational Needs
Petra Jorasch, European Seed Association, Brussels, Belgium
Plant breeders have always strived to create new variations of plant characteristics to provide solutions for disease and pest resistance, to achieve higher yields, to increase tolerance to environmental stress, and to breed new plant varieties that meet consumer expectations. The rediscovery of Mendel’s laws of heredity, in the early 1900s, turned the first plant breeding efforts from an art into science, and specialised farmer-breeders emerged, building a business concept on their efforts. From that point in time, scientific breakthroughs in agricultural and biological sciences have accelerated.
Governmental policy must be firmly based on sound scientific principles to avoid the risk of impeding innovation in plant breeding. Therefor seed industry takes the position that plant varieties developed through the latest breeding methods should not be subject to different or additional regulatory oversight if they could have been produced through traditional breeding methods or might also have been obtained from natural processes without human intervention. It is a major demand of the global seed industry to achieve alignment among governments on the criteria used to assess whether plants developed through plant breeding innovations should be subject to regulatory oversight under existing Genetically Modified Organism (GMO) regulations for plants.

Consistent policies needed!
Consistent policies among governments for products of the latest plant breeding methods, such as gene editing, would facilitate the development and uptake of advanced, innovative breeding applications by both industry and public breeders in developed and developing countries. Plant breeders need legal certainty so they can reliably plan their breeding programs, their product development and market potentials. Disproportionate regulatory hurdles mean higher costs, especially for registration and approval which limit the access of small and medium sized enterprises (SME) and public plant breeding institutions to the latest plant breeding innovation tools. Furthermore such government policies will impede the availability of a diversity of crops and varieties for farmers, including speciality crops and crops with niche markets.
Plant breeders Embrace the Power of Nature
The EU seed sector takes its responsibility in communicating about the benefits of plant breeding innovation. With our digital communication campaign "Embracing the Power of Nature" (www.plantbreeding.eu) we want to engage with a wider interested public as well as political decision makers on social media. We address the topic of food quality and how plant breeding can help to meet consumer expectations, how plant breeding can support to meet Europe's food demand and how resource efficient farming can benefit from plant breeding innovation. We invite all interested people to join the discussion by using the hashtag #EmbracingNature or #plantbreeding on social media!

W698: New breeding technologies: Prospects and regulatory hurdles

Will Genome Editing Edit European GMO Regulation?

Georg Leggewie, Federal Office of Consumer Protection and Food Safety (BVL), Berlin, Germany

The need for regulation of organisms treated by Genome Editing is presently being discussed among EU member states. In particular, the discussion focusses on to whether these organisms fall under the EU regulation of Directive 2001/18/EC and, hence, are to be treated like GMOs, or whether present legislation outside of this directive is sufficient to cover regulatory needs. Some parties also suggested an "in between" regulation that could yield in an assessment "light" for genome edited organisms. It has also been questioned whether the legal definition of a GMO is process or product based. The Federal Office of Consumer Protection and Food Safety (BVL) as German Competent Authority reiterates that law interpretation depends on a thorough analysis of the wording of the GMO definition. In addition, the teleological intentions of the law need to be identified and a systematic interpretation in the light of the provisions by superior law such as the Cartagena Protocol shall be taken into account. Single point mutations created by Genome Editing may not be subjected to the Directive either due to the definition of GMOs therein or by applying the exemption cause of Annex IB which refers to mutagenesis.
The highest institution regarding the interpretation of the law is the European Court of Justice (ECJ). The ECJ is about to deliver a leading case decision regarding several legal questions with respect to Genome Editing techniques. Among others, the two most important legal questions are: First, whether the scope of the exemption clause in the European genetic engineering law for organisms created by mutagenesis techniques includes Genome Editing and second, the conformity of this exemption clause with the precautionary principle. BVL expects that an ECJ decision will consider general principles laid down in international trading agreements which the EU is assigned to. Here, the principle of equal treatment of equal products in order to avoid an unjustified trading discrimination is most important.
A more detailed overview on crucial arguments being exchanged in the ongoing case will be discussed in the presentation.

W699: Next Generation Genome Annotation and Analysis

High-Resolution Comparative Analysis of Great Ape Genomes

Zev N. Kronenberg[1,2], Ian T. Fiddes[3], David Gordon[1], Shwetha Murali[1], Stuart Cantsilieris[1], Oliva S. Meyerson[4], Jason Underwood[1,5], Bradley J. Nelson[6], Mark J.P. Chaisson[7], Max L Dougherty[1], Katherine M. Munson[1], Alex R. Hastie[8], Mark Diekhans[9], Fereydoun Hormozdiari[10], Nicola Lorusso[11], Kendra Hoekzema[1], Ruolan Qiu[1], Karen Clark[12], Archana Raja[1], AnneMarie E. Welch[1], Melanie Sorensen[1], Carl Baker[1], Joel Armstrong[9], Tina A. Graves-Lindsay[13], Robert Fulton[13], Ahmet M. Denli[14], Emma Hoppe[1], Christopher Hill[15], Andy Wing Chun Pang[8], Joyce Lee[16], Susan Dutcher[13], Fred H. Gage[14], Wesley Warren[13], Jay Shendure[17], David Haussler[9], Valerie A. Schneider[18], Han Cao[16], Mario Ventura[11], Richard K Wilson[13], Benedict Paten[9], Alex Pollen[19] and Evan Eichler[20], (1)Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, (2)Phase Genomics, Seattle, WA, (3)University of California, Santa Cruz, CA, (4)Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Francisco, CA, (5)Pacific Biosciences, Menlo Park, CA, (6)Department of Genome Sciences, University of Washington School of Medi, Seattle, WA, (7)Computational Biology and Bioinformatics, University of Southern California, Los Angeles, CA, (8)Bionano Genomics, San Diego, CA, (9)Genomics Institute, University of California Santa Cruz and Howard Hughes Medical Institute, Santa Cruz, CA, (10)University of California, Davis, Davis, CA, (11)Department of Biology, University of Bari "Aldo Moro", Lecce, Italy, (12)National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD, (13)McDonnell Genome Institute at Washington University, St. Louis, MO, (14)The Salk Institute for Biological Studies, La Jolla, CA, (15)University of Washington School of Medicine, Seattle, WA, (16)BioNano Genomics, San Diego, CA, (17)University of Washington - Department of Genome Sciences, Seattle, WA, (18)National Center for Biotechnology Information (NCBI), Bethesda, MD, (19)Department of Neurology, University of California, San Francisco, San Francisco, CA, (20)University of Washington, Seattle, WA

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.

W700: Next Generation Genome Annotation and Analysis

GenSAS v5.1: A Web-Based Platform for Structural and Functional Annotation and Curation of Genomes

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

The Genome Sequence Annotation Server v5.1 (GenSAS, www.gensas.org) is a web-based annotation and curation platform that combines several common annotation tools into one easy-to-use, integrated resource. The user-friendly interfaces, with embedded instructions, guide users through the annotation process. GenSAS has annotation tools for eukaryotes and prokaryotes and supports model and non-model organisms. Users can upload a variety of evidence files to support the annotation process for their genome sequence. These include GFF3 files of aligned features and previous annotations; FASTA files of repeat, transcript, EST, or protein sequences; and gene models from Genbank. GenSAS also allows users to upload Illumina RNA-Seq reads, align the reads to the genome using TopHat, and use the data to train the gene model prediction program Augustus, which allows for more accurate gene models for eukaryotic genomes, especially non-model organisms. JBrowse and Apollo are integrated into GenSAS allowing structural annotation results to be easily viewed and manual curation to be performed. Users can share GenSAS projects with other users enabling collaborative or community wide curation. GenSAS also has a functional annotation step to assign protein functions and identify functional domains for the official gene set. After the annotation process is complete, the final step of the GenSAS pipeline generates the required files for publication which includes merging the manual annotations from Apollo into the final annotation.

W701: Next Generation Genome Annotation and Analysis

Combining RNA-Seq Data and Homology-Based Gene Prediction for Plants, Animals and Fungi

Jens Keilwagen[1], Frank Hartung[1], Michael Paulini[2], Sven O. Twardziok[3] and Jan Grau[4], (1)Julius Kühn-Institut, Quedlinburg, Germany, (2)European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom, (3)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (4)Institute of Computer Science, Martin Luther University Halle–Wittenberg, Halle, Germany

Genome annotation is of key importance in many research questions. The identification of protein-coding genes is often based on transcriptome sequencing data, ab-initio or homology-based prediction. Recently, it was demonstrated that intron position conservation improves homology-based gene prediction, and that experimental data improves ab-initio gene prediction.

Here, we present an extension of the gene prediction tool GeMoMa that utilizes amino acid sequence conservation, intron position conservation and optionally RNA-seq data for homology-based gene prediction. We show on published benchmark data for plants, animals and fungi that GeMoMa performs better than the gene prediction programs BRAKER1, MAKER2, and CodingQuarry, and purely RNA-seq-based pipelines for transcript identification. In addition, we demonstrate that using multiple reference organisms may help to further improve the performance of GeMoMa. Finally, we apply GeMoMa to four nematode species and to the recently published barley reference genome indicating that current annotations of protein-coding genes may be refined using GeMoMa predictions.

GeMoMa might be of great utility for annotating newly sequenced genomes but also for finding homologs of a specific gene or gene family. GeMoMa has been published under GNU GPL3 and is freely available at http://www.jstacs.de/index.php/GeMoMa.

W702: Next Generation Genome Annotation and Analysis

Defusion: A Tool to Improve Predictions of Tandemly Duplicated Genes Created by the MAKER Annotation Pipeline

Jie Wang[1], Dongyan Zhao[1], C. Robin Buell[2] and **Kevin L. Childs**[1], (1)Michigan State University, East Lansing, MI, (2)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

Genome sequencing of nonmodel organisms is growing rapidly, and it has become increasingly important to correctly annotate genes in an automated fashion. The MAKER pipeline provides a robust and scalable solution for genome annotation. However, accurate gene prediction by MAKER is dependent on high-quality and correctly aligned protein and transcript evidence. Exonerate is used by MAKER for aligning protein and transcript evidence. Unfortunately, with tandemly duplicated genes, exonerate often stretches protein and transcript alignments across the duplicated loci. MAKER interprets this evidence to mean that the tandemly duplicated genes are in fact a single locus, and one gene model is created that crosses the two loci. This is particularly problematic for plant species as genes involved in producing secondary metabolites are often tandemly arrayed. Here we developed a Python-based tool called deFusion to identify and locally annotate fused gene models generated by MAKER. Fused genes are recognized by finding sequence similarities between the 5' and 3' ends in MAKER gene models. By default, the midpoint of the fused gene model is suggested as the breakpoint, but human-curated customized breakpoints are also accepted. Fused loci are split, and each separated locus is locally re-annotated. Newly aligned evidence, new *ab initio* predictions and new MAKER gene models replace the evidence and models from the fused locus, and deFusion outputs a corrected transcript, protein and MAKER gff files. The deFusion tool has been successfully applied to fix fused gene models from two medicinal plant genomes: *Catharanthus roseus* and *Camptotheca acuminata*.

W703: Next Generation Genome Annotation and Analysis

ACE-Plus: Variant-Aware Gene Structure Prediction in Individualized Genomes

William H Majoros[1], Carson Holt[2], Michael Campbell[3], Doreen Ware[3], Mark Yandell[4] and Tim Reddy[5], (1)Duke University, DURHAM, NC, (2)University of Utah, Salt Lake City, UT, (3)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (4)Department of Human Genetics, University of Utah, Salt Lake City, UT, (5)Center for Genomic and Computational Biology, Duke University, Durham, NC

The accurate interpretation of genetic variants is critical for characterizing genotype-phenotype associations. Because the effects of genetic variants can depend strongly on their local genomic context, accurate genome annotations are essential. Furthermore, as some variants have the potential to disrupt or alter gene structure, variant interpretation efforts stand to gain from the use of individualized annotations that account for differences in gene structure between individuals or strains. We describe a method for identifying possible functional changes in gene structure that may result from sequence variants. Our method analyzes explicit haplotype sequences to detect gene-structure changes and their possible repercussions. Software implementing this method is freely available and can be readily applied to diverse species, making it a broadly useful tool for use in eukaryotic population-based re-sequencing projects, particularly for assessing the joint impact of all variants at a locus.

W704: Non-coding RNA

The Role of Noncoding RNA Pattern and Function in Cotton Fiber Cell Development

Xueying Guan, Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China

Cotton fiber is the most important sustainable fiber source for textile industry. It is a single cell organ derived from the epidermis of the cotton ovule or seed. To understand the molecular basis of plant cell differentiation pattern, the cotton fiber cell is a good model system. Non-coding RNA is emerging as one of the most important regulators for the gene expression in response to multiple biological transitions and environmental stimuli. We systematically investigate the non-coding RNA behavior in the fiber cell differentiation progress. The major data indicate both small RNA and long non-coding RNA (lncRNA) play critical roles in fiber cell development. First of all, the fiber cell generates a unique group of small RNA in the fiber initiation stage. For example, the miR828 and miR858 trigger the target gene GhMYB2 to generate tasiRNAs in fiber cell fate determination. Another MIXTA MYB transcriptional factor coding gene, GhMML3 can generate an antisense transcript on the 3' end of the gene loci. Together with the sense and antisense transcripts, a double strand RNA come into being to derive small RNAs. These secondary generated small RNAs interfere the cell fate determination in the mml3 mutant in stimulating the fiberless seed phenotype. On the other hand, the long non-coding RNAs are also found to take parts in the fiber cell differentiation by small RNA generation. We therefore conclude noncoding RNAs directly impact the fiber cell development in multiple aspects of molecular regulation.

W705: Non-coding RNA

Characterization of the Noncoding Bovine and Reindeer Rumen Papillae Transcriptome

Daniel Fischer, Ilma Tapio, Seppo Ahvenjärvi and **Johanna Vilkki**, Natural Resources Institute Finland, Jokioinen, Finland

Until now, transcriptome studies of the papillae of the rumen wall of ruminants like cattle or reindeer are rare. We present here a comparative analysis of the expression profiles of the rumen papillae based on six reindeer and eight cow samples. The transcriptomes of all samples have been sequenced with the Illumina HiSeq 3000 platform using stranded protocol with 2 x 150 bp reads. The reads were mapped against the Bovine (UMD_3.1) respective a recently published reindeer (http://gigadb.org/dataset/100370) genome. On the average, almost half of the mapped bovine reads mapped to unannotated regions of the bovine genome, indicating the need for better annotation of the genome. We analyzed long-non-coding RNA in the papillae by applying a pipeline for lncRNA-calling derived from a FAANG-Europe working group and characterize the data with respect to that. The expression profiles for non-coding transcripts are compared between reindeer and bovine. Until the publication of the reindeer genome, usually the bovine genome and annotation were used to analyze both species. Hence, with the publication of the reindeer genome and annotation, a special focus of this study is on the effect of the new, species-specific genome and its effect on the results of the characterization. For that, we analyze for the reindeer data also the differences between using the bovine genome as a reference versus using the new reindeer specific genome.

W706: Non-coding RNA

Annotation of Noncoding RNAs in Wheat Stem Sawfly and Wheat

Hikmet Budak, Montana State University, Bozeman, MT

Wheat Stem Sawfly (WSS) is one of the most important pests, causing yield and economic losses in crop plants. The lack of information about molecular mechanisms of WSS for defeating plant's resistance prevents application of effective pest control strategies therefore, it is essential to identify the genes and their regulators behind WSS infestations. Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) are recognized with their regulatory functions on gene expression, tuning protein production by controlling transcriptional and post-transcriptional activities. A transcriptome-guided approach was followed in order to identify miRNAs, lncRNAs, tRNA, and mRNA of WSS and their interaction networks. A total of 11 miRNA families was detected in WSS transcriptome together with the annotation of 1,251 novel mRNAs. The network between WSS miRNAs, lncRNAs, and mRNAs suggested miRNA-mediated regulatory roles of lncRNAs as competing endogenous RNAs. In the light of the previous evidence that small RNA molecules of a pathogen could suppress the immune response of host plant, we analyzed the putative interactions between larvae and wheat at the miRNA level. The analysis of tRNA gene content of WSS transcriptome revealed that the majority of tRNA gene families were represented by more than a single copy. A total of 159 putative tRNA genes were identified, 41 and 50 of which were encoded by actively-expressed mRNA and lncRNA transcripts, respectively. Overall, this study provides a profile of larva and adult WSS life stages in terms of coding and non-coding elements. These findings also emphasize the potential roles of wheat and larval miRNAs in wheat resistance to infestation and in the suppression of resistance which is critical for the development of effective pest management strategies.

W707: Non-coding RNA

Broadening the miRNA Catalogue in Livestock Species

Andreia J Amaral[1], Sylvain Marthey[2], Anne Hoffmann[3], Jacques Lagnel[4], Fiete Haack[5], Kisun Pokharel[6], Oana Palasca[7], Stefan E. Seemann[8], Anoop Arya[9], Luís Gama[10], Martien A.M. Groenen[11], Juha Kantanen[6], Richard P.M.A. Crooijmans[12], Elisabetta Giuffra[13], Peter F. Stadler[14], Christian Anthon[8,15], Ole Madsen[12] and Jan Gorodkin[7], (1)Faculdade de Medicina Veterinária, University of Lisbon, Lisboa, Portugal, (2)INRA, UMR GABI, Jouy-en-Josas, France, (3)University of Leipzig, Leipzig, Germany, (4)INRA PACA, Montfavet Cedex, France, (5)Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany, (6)Natural Resources Institute Finland, Jokioinen, Finland, (7)Center for non-coding RNA in Technology and Health, Department

of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark, (8)Center for non-coding RNA in Technology and Health, Department of Veterinary Clinical and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark, (9)AgroParisTech, INRA, Wageningen University, WAGENINGEN, Netherlands, (10)Faculdade de Medicina Veterinária, Universidade de Lisboa, Lisbon, Portugal, (11)Wageningen University & Research Center, Wageningen, Netherlands, (12)Wageningen University, Wageningen, Netherlands, (13)INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France, (14)Bioinformatics Group, Department of Computer Science University of Leipzig, Leipzig, Germany, (15)Center for non-coding RNA in technology and Health, Faculty for Health and Medical Sciences of University of Copenhagen, Frederiksberg, Denmark
MicroRNAs play a crucial role in the regulation of gene expression. Their action is crucial in many biological processes and functions, such as cell development and differentiation, and in response to disease. Moreover, it has been shown that polymorphisms in miRNAs can be linked to diseases and complex traits. An improved annotation of miRNAs in domestic animals is therefore required in order to acquire a comprehensive understanding of their impact on livestock traits. For this, we used 328 quality approved small-RNA-seq datasets available from public repositories for five livestock species (*Gallus gallus, Sus scrofa, Equus caballus, Ovis aries* and *Bos taurus*). The data was used to quantify miRNA expression in different tissues as well as to identify putative novel miRNA candidates. For *Bos taurus,* our analysis allowed an increase of 50% of the total miRNA catalogue currently available. Interestingly, for *Ovis aries*, we have identified a large number of putative novel miRNAs that share the seed with pre-existing miRNAs in other species. Currently we are studying the miRNA convergence at the functional and genomic level, which will be reported along with the list of novel miRNAs. Furthermore, we are also investigating the occurrence of other types of ncRNAs, e.g. tRNAs and snoRNAs in the analyzed data, some of which are known to be processed into miRNA. We have created a comprehensive ncRNA annotation for the studied species as well as interspecies pairwise and multiple genome-wide alignments. We believe these findings will further contribute to the understanding of the functional genome of the studied species.

W708: Non-coding RNA
RNAcentral: A Comprehensive Collection of Non-Coding RNA Sequences
Blake A Sweeney, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, United Kingdom

W709: Non-coding RNA
Evolution of Long Intergenic Non-Coding RNAs by Duplication in Plants
Sishuo Wang and **Keith Adams**, University of British Columbia, Vancouver, BC, Canada
Long intergenic non-coding RNAs (lincRNAs) have been identified on a large scale in several plant species. We conducted a systematic study of the evolution of lincRNAs by duplication in six plant species. In contrast to previous findings in animals, we found that some plant genomes are rich in duplicated lincRNAs and the proportion of duplicated lincRNAs greatly varies among different organisms. Distinct from protein-coding genes, only a small proportion of duplicated lincRNA pairs were derived from whole-genome duplication and interspersed duplicates are the dominant type of lincRNA duplicates. Duplicated lincRNAs exhibited much lower co-expression correlation coefficients than protein-coding genes indicative of extensive expression divergence. Paralogous lincRNAs showed extensive complementary and partitioned expression in a tissue-specific manner suggesting sub-functionalization of duplicated lincRNAs. Tandemly duplicated lincRNAs showed lower expression divergence than whole-genome duplicates and interspersed duplicates. In addition, the expression level divergence was correlated with DNA methylation level divergence, but not sequence divergence for duplicated lincRNAs, suggesting the role of DNA methylation and epigenetic factors in mediating expression divergence of lincRNA duplicates. This study provides insights into the evolution of lincRNAs by duplication in plants.

W710: Non-Seed Plants
Diatom Evolution through the Lens of the Largest Diatom Genome Sequenced to Date, *Psammoneis japonica*
Matthew Parks, Chicago Botanic Garden, Glencoe, IL
The hyper-diverse and ecologically important diatoms (Bacillariophyta) are unicellular phytoplankton with chimeric genomes derived potentially from two photosynthetic (red and green algae) and one heterotrophic precursor lineages. Complete genome sequences are currently available for four diatom species, representing two of four major morphological types (polar centrics and raphid pennates). Here, we present the genome sequence and analyses for *Psammoneis japonica*, which is a member of a critical third diatom clade, the araphid pennates. The *Psammoneis* assembly comprises approximately 25X PacBio and 100X Illumina coverage and is stringently filtered for contaminant sequences. With a resulting total assembly size of 91.2Mbp (N50=390kbp) and over 15,000 annotated genes, the *Psammoneis* genome is the largest diatom genome sequenced thus far. Sequencing of this genome enables a more comprehensive, comparative genomic view of diatom ecological (pelagic/benthic), morphological (centric/pennate, araphid/raphid) metabolic (carbon/nitrogen cycles) and reproductive (oogamy/anisogamy/isogamy, dioecy/monoecy) character evolution. General comparisons, including orthologous protein divergence, intron conservation and repetitive element content also provide further insight into rates and patterns of evolution that have shaped genome size and architecture across the diatoms.

W711: Non-Seed Plants
Plastid-Mediated Stress Response Signalling in the Algal Progenitors of Land Plants
Jan de Vries[1], Bruce A. Curtis[1], Sven B. Gould[2] and John M. Archibald[1], (1)Dalhousie University, Halifax, NS, Canada, (2)Heinrich Heine University Düsseldorf, Düsseldorf, Germany
Land plants evolved from streptophyte algae. Recent phylogenomics analyses have shown that of all extant streptophytes, Zygnematophyceae are most closely related to the algal land plant progenitor. Among the major questions revolving around the algal ancestors of land plants is what features aided their conquest of land. One of these likely was the ability to ward off terrestrial stressors, such as high light, drought, and severe cold. These stressors are known to strongly impact plastid biology and recent work on retrograde signalling highlights the plastid as a

major hub in land plant stress-response. We used comparative transcriptomics to investigate stress response signalling in one representative of each of the six major lineages of streptophyte algae, including the Zygnematophyceae *Zygnema*. Our data uncover the presence of stress signalling circuits known from land plants in various streptophyte algae. *Zygnema* stood out in having all genes necessary for utilizing the canonical abscisic acid signalling pathway – *the* stress phytohormone of land plants. Simultaneously, *Zygnema* invested a higher transcriptional budget into its plastid than any of the other five streptophyte algae. As seen in many of the other species investigated, it was found to encode a battery of retrograde signalling genes known from land plants. We conclude that stress signalling and plastid-nucleus communication were among the features that aided the algal land plant ancestor in its conquest of *terra firma*.

W712: Non-Seed Plants

A Comparative Metabolomic Study between Constitutively Desiccation Tolerant *Syntrichia ruralis* and Inducibly Desiccation Tolerant *Physcomitrella patens* during Drying

Kumudu Rathnayake, Univesity of South Dakota, Vermilion, SD and Koster, Karen L., University of South Dakota, USA; Oliver, Melvin J., USDA-ARS, University of Missouri, USA; Alexander, Danny C., Metabolon Inc, USA; Wone, Bernard W. M., University of South Dakota, USA

Vegetative desiccation tolerance, lost in most angiosperms during evolution, is an important trait in many bryophytes, and allows some plants to survive the loss of almost all cellular water and resume growth upon rehydration. Among the mosses, some are fully desiccation tolerant (DT) with constitutive mechanisms that allow them to survive drying, whereas others have inducible mechanisms. Understanding metabolic changes that accompany the desiccation response in mosses might inform the development of more drought tolerant crop plants. We compared metabolic profiles of gametophytic tissues in constitutively DT *Syntrichia ruralis* and inducibly DT *Physcomitrella patens* to understand their global metabolic responses during dehydration. We detected similar metabolic changes in both mosses during dehydration, but also species-specific perturbations. The abundance of intermediates involved in the glutathione pathway, such as gamma-glutamyl amino acids and 5-oxoproline, increased significantly in both moss species, suggesting that they have a potential protective role during drying. Endogenous ABA levels increased significantly in dehydrated *P. patens,* but were not reported for *S. ruralis*, suggesting that ABA plays an important role in the acquisition of desiccation tolerance in *P. patens* but not in the constitutively DT *S. ruralis*. The levels of many soluble sugars and tricarboxylic acid cycle metabolites increased significantly during dehydration in *P. patens* but not in *S. ruralis* are in accord with the constitutive protection strategy in *S. ruralis*. Collectively, the metabolic alterations observed in *P. patens* could contribute to the acquisition of desiccation tolerance by establishing cell protective mechanisms via osmotic adjustment and enhanced antioxidant systems.

W713: Non-Seed Plants

The Funaria Genome: A Chromosome-Scale Assembly.

Peter Szovenyi, University of Zurich, Zurich, Switzerland

Thanks to its attractive phylogenetic position and molecular genetic attributes the model moss *Physcomitrella patens* has become a powerful tool for developmental and molecular genetic research during the last decade. Although many of the advantageous properties of the bryophyte system have been used in studies on development and cellular function, several peculiar characteristics have not received much attention so far. In particular, evolutionary genomics and comparative developmental genomics of mosses are still in their infancy and are waiting to be exploited. Here we report the genome of *Funaria hygrometrica*, a relative of the model moss *P. patens*. *Funaria* and *Physcomitrella* represent end points of morphological and developmental complexities within the group of Funariid mosses and thus provide an ideal model system for comparative studies. Using a polished PacBio assembly and additional Chicago and HiC libraries we present our genome assembly consisting of 26 chromosomes with a total length of ~320 Mb. We show that many features of the *F. hygrometrica* genome are shared with those of *P. patens* but some striking differences exist with strong phenotypic consequences. We will also use our genome assembly to describe how chromosome-scale rearrangements and gene losses/gains have contributed to the evolution of the *P. patens* genome. Availability of the *F. hygrometrica* genome will open up new fields in non-vascular plant research by providing novel insights into the comparative genomics, developmental biology and evolutionary genomics of the Funariid mosses.

W714: Non-Seed Plants

The *Sphagnum* Microbiome: Describing the Complex Interactions between *Sphagnum* and their Symbiotic Bacteria

Adam Healey, HudsonAlpha Institute For Biotechnology, Huntsville, AL

Peat moss (generated from the *Sphagnum* genus) is responsible for sequestering approximately ⅓ of the world's terrestrial carbon, more than any other plant genus. *Sphagnum* also plays an important role in northern peatland ecosystems, due to their ability of occupy a wide variety of niche habitats. Their ability to grow and dominate these habitats can be attributed to the transformation of their environment into one that is unfavourable to other plants species, achieved through the cultivation of methanotrophic and nitrogen-fixing bacteria. Despite their importance and symbiotic relationship with *Sphagnum*, this microbiome has not been well-described, in part to due the difficulty in culturing these organisms as well as the *Sphagnum* ecosystem itself, in which many species occupy the same habitat making identification difficult. Here, using *de novo* generated metagenomic assemblies derived from whole-genome sequencing of *Sphagnum* diversity collections, we describe the *Sphagnum* microbiome, focusing on lineage specific differences among species, as well as those bacteria which comprise the core microbiome. With this knowledge, we hope to shed light on not only plant-microbe interactions within arctic peatlands, but also the niche occupation of inhospitable habitats by *Sphagnum* species.

W715: Non-Seed Plants

Unfurling Monilophyte Genomics: How Polyploidy, Transposable Elements, and the Alternation of Independent Generations Drive Fern Evolution

D. Blaine Marchant, University of Florida, Gainsville, FL

Ferns (Monilophyta) are notorious for possessing large genomes (average 1C = 12 pg) and numerous chromosomes (average $n = 59$). Historically, these features, which characterize all homosporous ferns, were attributed to repeated rounds of ancient polyploidy (whole-genome

duplication - WGD). However, this explanation for the origin of large genomes and many chromosomes, long considered emblematic of ferns, is unsubstantiated due to the absence of a sequenced fern genome. The lack of this crucial resource has not only hindered investigations of the evolutionary processes underlying the unusual genome characteristics of ferns, but has also impeded synthesis of genome evolution across land plants. Using the model fern species *Ceratopteris richardii*, or C-Fern, we obtained the first sequenced and assembled nuclear genome for a homosoporous fern. With this novel resource, we evaluated the possible roles of WGD and the expansion of transposable elements (TEs) in shaping fern genome evolution. In addition, we investigated the ramifications of having two independent life phases (gametophyte and sporophyte generations) at the genome, gene, and transcript level. Not only do these results address the evolutionary processes and genomic changes that have shaped the genomes of ferns, but the assembled data set will also facilitate new insights into evolutionary genomics of land plants as a whole.

W716: Oats, Wild and Cultivated
Genomic Selection Addresses Genotype-by-Environment Interaction
Lucia Gutierrez, University of Wisconsin - Madison, Madison, WI
Modern plant breeding involves evaluating the genetic merit of lines by discerning genetic from environment and noise components. Therefore, controlling micro (i.e. plant-to-plant variations due to field heterogeneity) and macro (i.e. genotype by environment interaction, GxE) environmental variability is fundamental for breeding success. The aim of this research was to compare methodological approaches to optimize resource allocation for plant breeding using genomic selection and to test them on a large MET for oats. We compared experimental design strategies based on both micro-environmental variation (local control of field heterogeneity with experimental designs) and macro-environmental variation (GxE), to increase prediction accuracy as well as strategies for predicting genotypic performance for local adaptation. Our results show that resources can be optimized within and across locations when field heterogeneity and GxE is modeled with genomic prediction tools. This would increase the selection gain that can be achieved with the same phenotyping resources.

W717: Oats, Wild and Cultivated
Targeted Sequence Analysis of Pathogenesis-Related Factors in Oat
Matthew J. Moscou, The Sainsbury Laboratory, Norwich, United Kingdom

W718: Oats, Wild and Cultivated
Development of a Transposon-Mediated Gene Tagging System in Oat
Mohannad Mahmoud[1], Rajvinder Kaur[1], Nicholas A. Tinker[2] and **Jaswinder Singh**[3], (1)McGill University, Ste Anne de bellevue, QC, Canada, (2)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (3)McGill University, Ste Anne de Bellevue, QC, Canada
Oat (*Avena sativa*) is a globally-important cereal crop species with documented benefits to human health. However, oat has lagged behind other cereals with respect to genomic resources, due in part to the complexity of its hexaploid genome. Among various functional genomic tools to characterize genes in plants, transposon-based insertional mutagenesis approach offers great potential, especially in plant species, which possess large genomes and genetic transformation is not a routine. In order to overcome gene and genome redundancy, a unique transposon-assisted activation tagging approach has been described in Arabidopsis, popular, rice and barley. Our aim in the current project is to introduce the *Ac/Ds* transposable elements in oat for the development of an activation gene tagging resource. Highly regenerative calli derived from mature oat seeds of oat cultivar 'Park', were successfully transformed with several *Ac/Ds* constructs, using a biolistic delivery system. Twenty-two independent transgenic events were obtained using two different antibiotic selection schemes. Our data indicate that co-transformation of *Ac* and *Ds* constructs led to 5% of primary *Ds* transpositions at T_o stage. Generation advance of T_1 plants containing both *Ac* and *Ds* elements from four different events established the transposition frequency between 11 to 30%. Movement of *Ds* element was confirmed by histochemical and molecular assays. Homozygous oat lines containing individual *Ac* and *Ds* elements are being hybridized for further re-mobilization of *Ds* transposons in the oat genome.

W719: Oats, Wild and Cultivated
Updating the Evolutionary Paradigm of Hexaploid Oat
Yong-Bi Fu, Plant Gene Resources of Canada, Saskatoon, SK, Canada
My presentation will update our research in oat evolution reflected in the maternal lineages of 25 Avena species. The maternal phylogenetic signals of these species were acquired from their chloroplast and mitochondria genomes using a multiplexed shotgun sequencing procedure. Phylogenetic analyses of the acquired organelle SNP data revealed a new maternal pathway of oat genome evolution involving three diploid species (*A. ventricosa*, *A. canariensis* and *A. longiglumis*) and two tetraploid species (*A. insularis* and *A. agadiriana*).

W720: Oats, Wild and Cultivated
A Reference Quality Assembly and Annotation of the *Avena atlantica* Genome
Peter J. Maughan[1], Rebekah Lee[1], Tim Langdon[2], Jessica Schlueter[3] and Rick Jellen[1], (1)Brigham Young University, Provo, UT, (2)IBERS, Aberystwyth University, Aberystwyth, United Kingdom, (3)University of North Carolina at Charlotte, Charlotte, NC
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 has garnered renewed interest for human consumption. Until now there has not been a published reference genome for any of the three oat sub genomes. Here we report a quality sequence assembly, annotation and hybrid optical map assembly of the A-genome diploid *Avena atlantica*. The hybrid assembly is composed of 3417 scaffolds, spanning ~3.7 Gb, with an N50 of 11.86 Mb and a BUSCO estimated completeness of 97.6%. It is hoped that this genome sequence can escalate research within the oat community.

W721: Oats, Wild and Cultivated

Plans for a Hexaploid Oat Genome, and Beyond

Nick Sirijovski and **Olof Olsson**, Lund University, Lund, Sweden

Relative to other cereals such as rice, barley and wheat, very little is know about the genetics of oat. Cultivated oat (*Avena sativa*) is a hexaploid comprised of three diploid genomes (AACCDD). It has a 1C genome of 21 chromosomes with a total size estimated to 13Gb. The large genome size and polyploidy has meant that deciphering the genetics of cultivated oat has lagged behind other cereals. Recently, oat has received much attention due to well documented health benefits of consuming this 'super food', which in turn has lead to increased production of oat-based novel foods and ingredients e.g. dairy alternatives, beta-glucan extracts, and even meat substitutes. With the fast paced development of next generation sequencing technologies, it has now become possible and affordable to undertake genome sequencing of hexaploid oat using short read technology. Herein we report on the status of the Swedish oat genome sequencing project, which is part of the newly inaugurated ScanOats research center in Lund, Sweden. In addition, we present our plans and preliminary data for non-destructive analysis of oat seeds using neutrons, X-rays and NMR.

W722: Organellar Genetics

Progress in Implementing Plastid Transformation in *Arabidopsis thaliana*

Pal Maliga[1], Qiguo Yu[1], Lisa Lamanna[1], Megan Kelly[1] and Kerry Lutz[2], (1)Rutgers University, Piscataway, NJ, (2)Farmingdale State College, Farmingdale, NY

We have reported high frequency plastid transformation in *Arabidopsis thaliana* based on hypersensitivity to spectinomycin in *ACC2* null mutants lacking a plastid-targeted acetylcoenzyme A carboxylase. As it was difficult to obtain fertile transplastomic plants in the hypersensitive Sav0 accession, we deleted the *ACC2* gene in the RLD and Ws accessions using the CRISPR/Cas9 system. RLD and Ws were chosen because they readily yield plants in tissue culture when exposed to plant growth regulators. Progress will be reported on plastid transformation in Arabidopsis with new vectors which do not impose a metabolic burden on the plants in *ACC2*-knockout RLD and Ws ecotypes.

W723: Organellar Genetics

Designer Pentatricopeptide Repeat Proteins as Regulatory Switches for Plastid Transgenes

Alice Barkan[1], Pal Maliga[2], Qiguo Yu[2], Margarita Rojas[3] and Rosalind Williams-Carrier[3], (1)University of Oregon, Eugene, OR, (2)Rutgers University, Piscataway, NJ, (3)University of Oregon

The manipulation of chloroplast genomes offers unique opportunities for the production of biofuels, bioplastics, and pharmaceuticals. Transgenes in the plastid compartment can be expressed at extraordinary levels and plastid operons are ideal for the expression of non-native biochemical pathways. However, to fully exploit transplastomics there is a need to develop robust methods to regulate transgene expression in order to reduce deleterious effects on plant growth. Recent advances in understanding how pentatricopeptide repeat (PPR) proteins activate chloroplast gene expression suggest novel approaches for the design of regulatory systems for transplastomic applications. PPR proteins are RNA binding proteins that activate plastid gene expression at the post-transcriptional level. They bind specific RNA sequences via a modular recognition mechanism involving an amino acid code that is quite well understood. These features predict that the regulated expression of engineered PPR proteins from the nuclear genome in conjunction with cognate PPR binding sites placed in appropriate positions near plastid transgenes can be used to control where and when plastid transgenes are expressed. I will describe the results of proof-of-concept experiments that explore strategies of this type.

W724: Organellar Genetics

Cytonuclear Interactions and Extremes Rates of Plant Organelle Genome Evolution

Daniel Sloan, Colorado State University, Fort Collins, CO

W725: Organellar Genetics

Protein Composition, Structure and Biogenesis of the Eukaryotic Carbon-Concentrating Organelle, the Pyrenoid

Martin Jonikas, Princeton University, Princeton, NJ

W726: Organellar Genetics

Microcompartments for RuBisCO: Compartmentalization for Improved Photosynthesis

Ben Long, Wei Yih Hee, Robert E. Sharwood, Benjamin D. Rae, Sarah Kaines, Yi Leen Lim, Nghiem D. Nguyen, Baxter Massey, Soumi Bala, Susanne von Caemmerer, Murray R. Badger and G. Dean Price, Australian National University, Canberra, Australia

Improvement of yield potential in food crops has hit a breeding road-block which could be solved through genetic engineering. Achieving this goal is also likely to lead to better use of water, more efficient use of valuable agricultural land, and help alleviate growing global demand for primary produce. One proposed strategy is to introduce CO_2-concentrating mechanisms (CCMs) from cyanobacteria into C_3 chloroplasts such that they can achieve higher photosynthetic rates, greater biomass, and increased yield. The cyanobacterial CCM utilizes active accumulation of CO_2 in the form of bicarbonate within the cell, where it diffuses into a specialized micro-compartment known as the carboxysome, for rapid conversion to CO_2 and fixation into organic compounds by the enzyme RuBisCO. Expression and assembly of carboxysomes within plant chloroplasts, however, is a highly complex metabolic engineering task in this crop improvement strategy. The carboxysomes of some cyanobacteria require the coordinated expression of as many as 13 genes, with protein stoichiometries ranging from tens to thousands of individual copies of some subunits. Here we present current progress and challenges toward the expression and assembly of functional alpha-type carboxysomes from the coastal marine cyanobacterium *Cyanobium* sp. PCC 7001 within tobacco chloroplasts, toward the improvement of photosynthesis.

W727: Ornamentals

Utilizing GBS and Transcriptome Analysis to Understand Plant Development Rate in Petunia

Yufang Guo, QiuXia C. Chen and **Ryan M. Warner**, Michigan State University, East Lansing, MI

The rate of vegetative node formation (development rate; the inverse of plastochron) strongly influences the timing of flowering for plants grown under inductive conditions, as flowering time is a function of how many nodes form before the transition to reproductive development occurs, and the rate at which those nodes form. Understanding the genetic control of development rate would aid breeding of cultivars with reduced crop production time, particularly under cool temperatures, which is a desirable breeding goal in petunia. The genetic determinants of vegetative development rate are complex and poorly understood. To facilitate elucidation of the control of development rate, two interspecific *Petunia* F_7 recombinant inbred line (RIL) populations, *P. integrifolia* × *P. axillaris* (the "IA" population), and *P. axillaris* × *P. exserta* (the "AE" population), were developed, phenotyped for development rate and other crop timing and quality traits across a range of temperatures, and genotyped using "tunable " genotyping-by-sequencing. Despite considerable marker segregation distortion and regions of suppressed recombination in each population, multiple quantitative trait loci (QTL) for development rate were identified in each population. Additionally, we identified differentially expressed genes in shoot apex tissues between fast- and slow-developing IA RILs, 13 of which mapped to within 1 centimorgen (cM) of a development rate QTL. These results will facilitate gene discovery to further elucidate genetic control of development rate and other crop timing and quality traits in petunia.

W728: Ornamentals

Mapping a New Black Spot Resistance Locus in Rose

Jason D. Zurn[1], David C. Zlesak[2], James Bradeen[3], Stan C Hokanson[4] and Nahla Bassil[1], (1)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (2)University of Wisconsin River Falls, Department of Plant and Earth Science, River Falls, WI, (3)University of Minnesota, Department of Plant Pathology, St. Paul, MN, (4)University of Minnesota, Department of Horticultural Science, St. Paul, MN

Rose black spot, caused by *Diplocarpon rosae*, is one of the most devastating foliar diseases of cultivated roses (*Rosa hybrida*). The pathogen is globally distributed and has the potential to cause large economic losses in the outdoor rose industry. Genetic resistance is the most economical disease management strategy for black spot and many breeding programs are focused on creating cultivars with durable resistance. The tetraploid cultivar Brite Eyes™ ('RADbrite') is resistant all *D. rosae* races except race 12. Because of this broad resistance, a 94 individual F_1 mapping population was developed by crossing Brite Eyes™ to the susceptible tetraploid 'Morden Blush'. The population was phenotyped with four races (8, 9, 10, and 11) and a genetic map was constructed using the WagRhSNP 68K Axiom array. The F_1 individuals were either resistant or susceptible to all races evaluated and segregated 1:1, suggesting resistance is mediated by a single locus. Preliminary mapping places the Brite Eyes™ resistance locus on a single linkage group in a 31.7 cM region delimited by RhMCRND_7069_318 and Rh12GR_258_2610. This linkage group is homeologous to chromosome 5 of the diploid integrated consensus map. Prior to this experiment, three resistance loci have been identified (*Rdr1, Rdr2,* and *Rdr3*). Both *Rdr1* and *Rdr2* are located on a chromosome 1 homeolog and the equivalent chromosomal location of *Rdr3* is unknown. However, races 3 and 9 are virulent on *Rdr3*, suggesting the resistance in Brite Eyes™ is novel. Future efforts will focus on developing a diagnostic test for marker assisted selection.

W729: Ornamentals

Rose Genomics and Genome Sequencing

M Bendahmane, Ecole Normale Supérieure, Lyon, France

Roses are of high symbolic value and have great cultural importance in many societies worldwide. The rose is well suited to be an original model organism for woody ornamental species as it has a relatively small genome size (560Mbp) and it has a short life cycle for a perennial woody plant. Several characters, such as recurrent blooming, flower morphogenesis, scent… are of high economic importance. During the past years, we generated a number of molecular, genomic and biotechnology tools such as an efficient genetic transformation protocol and a database of Rosa expressed genes with thorough annotation and an overview of gene expression patterns in a variety of tissues, conditions and developmental stages; the latest represented a valuable prerequisite to the rose genome sequencing. We have undertaken the genome sequencing of the diploid Rosa chinensis, an ancestor of modern roses that contributed several important characters. We generated and assembled a draft genome sequence for this cultivar. However, its relatively high heterozygosity hampered high quality genome assembly with 16,000 scaffolds and a N50 of about 230kb. To overcome this difficulty, we have generated and sequenced a homozygous tissue using R. chinensis as starting material. The availability of this homozygous material yielded high quality genome assembly with reduced number of contigs that could be assembled to pseudomolecules using genetic maps. The high-quality rose genome allowed expert gene annotations as well as the reconstruction of gene regulatory pathways associated with major rose traits. Together, these resources provide a solid foundation for understanding the mechanisms governing rose traits and their diversity. Latest developments will be presented and discussed.

W730: Ornamentals

Characterization of *GmSACPD* Gene Family in Soybean Reveals the Presence of Neofunctionalization Events

Naoufal Lakhssassi, Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL

The Stearoyl-acyl carrier protein desaturase (GmSACPD) gene family in soybean is composed of four members. Synteny analysis revealed that GmSACPD-A and GmSACPD-B were present in a new-duplicated segment, whereas GmSACPD-C was duplicated into GmSACPD-D member. The GmSACPD-B and GmSACPD-C genes have been previously reported to control soybean seed stearic acid content. However, it is not clear if the other GmSACPD members contribute to seed stearic acid content or have been neofunctionalized within the soybean genome to acquire new functions. In this study, a subset of an EMS mutagenized soybean population was screened to identify mutants within *GmSACPD-D* gene. Using a forward genetics approach, a nonsense and four missense *Gmsacpd-c* mutants were identified to contain not only high levels of seed, but also increased leaf and nodule stearic acid content. Soybean plants with GmSACPD-C mutations in non-conserved residues show an increase in stearic acid content while conserving healthy nodules. Interestingly, mutational analysis uncovers the impact of GmSACPD-C mutations in leaf and nodule structure and morphology. Using a reverse genetics approach, three missense *Gmsacpd-d* mutants were identified by TILLING. Fatty acid analysis reveals that all mutant lines contained the same level of seed stearic acid as the wild-type

Forrest. Phylogenetic analysis of sequenced genomes from 24 plant species indicates that SACPD genes were duplicated and derived from a common ancestor and a unique gene still present in Chlorophytic algae. Furthermore, gene structural analysis indicates intron loss events that occur during the transition from the relative of the aquatic ancestor of all land plants, chlorophytic algae, to the most ancestral land plant species. Moreover, GmSACPD-C showed cytoplasmic localization, whereas GmSACPD-D showed a mitochondrial-like localization patterns contradicting previous bioinformatics predictions. Our results demonstrate that GmSACPD-D member has been neofunctionalized to acquire a possible role in cyst nematode infection, to be temporary regulated in embryo development, and to exhibit an expression confined to ovule.

W731: Palm Genetics and Genomics
Oil Palm Breeding at a Turning Point
Sean Mayes, School of Biosciences, University of Nottingham, Nr Loughborough, United Kingdom

W732: Palm Genetics and Genomics
Construction of a SNP-Based Linkage Map and Identification of QTL Associated with Trunk Height in Oil Palm
Wirulda Pootakham, Nukoon Jomchai, Chutima Sonthirod and Sithichoke Tangphatsornruang, National Center for Genetic Engineering and Biotechnology, Pathum Thani, Thailand
Oil palm (*Elaeis guineensis*) is one of the most important oil bearing crops in the world. Marker-assisted selections have played a pivotal role in oil palm breeding programs as they reduce the amount of time and resources required to develop new cultivars. Rapid advancement in sequencing throughput together with an overall decrease in sequencing cost, next generation sequencing technologies have been applied to SNP identification in various plant species. We applied a genotyping-by-sequencing (GBS) approach for a large-scale SNP discovery and genotyping of a mapping population. We constructed the first SNP-based linkage map in oil palm, which spanned 1,429 cM and comprised 1,085 markers. We subsequently used the linkage map for the identification of quantitative trait loci (QTL) associated with trunk height. Three QTL affecting trunk height were detected on linkage groups 10, 14 and 15. Interestingly, the QTL governing stem stature identified on linkage group 14 was linked to two open reading frames encoding a putative gibberellin 2-oxidase and a putative DELLA protein GAI1, both of which have been implicated in plant height regulation via gibberellin homeostasis and signaling pathway. SNP markers associated with two major QTL were validated using the TaqMan assays in the population with similar genetic background. Additional validations were performed with individuals from other genetic background, and markers associated with *EgDELLA* appeared to be linked to trunk height phenotypes.

W733: Palm Genetics and Genomics
Conserved and Novel Mechanisms Underlie Oil Palm (*Elaeis guineensis Jacq.*) Fruit Abscission: A Model for Studying Fruit Abscission in the Palm Family (*Arecaceae*).
Timothy J. Tranbarger, Institut de Recherche pour le Développement, UMR DIADE, Quito, Ecuador
Recent research on the molecular and cellular mechanisms of fruit abscission of the tropical palm species *Elaeis guineensis* (African oil palm) and the function and original features of the fruit primary abscission zone (AZ^1) will be presented. The AZ^1 consists of approximately 10 cell layers in the boundary region between the pedicel (P) and mesocarp (M). During early fruit development, AZ^1 cells are distinguishable from P and M cells by their smaller size and periclinal cell division orientation. AZ^1 cell wall thickness increases earlier during development suggesting cell wall assembly occurs more rapidly in the AZ^1 than adjacent P and M cells. AZ^1 cells contain numerous intra-AZ^1 layer plasmodesmata (PD), but very few inter-AZ^1 layer PD, while nuclei are located adjacent to PD and are remarkably aligned within AZ^1 layer cells, and remain aligned and intact after abscission. These cellular features allow a high capacity for intra-AZ^1 layer signaling that may be important for AZ^1 development and function for fruit abscission. A decrease in methylesterified homogalacturonan (the main pectin component of the middle lamella) and the expression of two AZ^1 specific polygalacturonases are observed in AZ^1 cell layers during abscission. A comprehensive transcriptome analysis of the AZ^1, P and M tissues reveal specific AZ^1 expression combined with overlapping expression patterns between the AZ^1 and adjacent P and M tissues may be important for AZ^1 functional specificity. These results provide a platform to study fruit abscission within the palm family to understand how this fundamental plant process has diversified during evolution.

W734: Palm Genetics and Genomics
Resistance Gene-Mediated Recognition of the Oil Palm Bud Rot Pathogen
Sebastian Schornack, University of Cambridge, Sainsbury Laboratory (SLCU), Cambridge, United Kingdom
The filamentous oomycete pathogen *Phytophthora palmivora* infects and causes disease in many economically important crops and is the causal agent of oil palm bud rot. Despite its impact on oil palm cultivation, not much is known about resistance mechanisms to control *P. palmivora* infection. Small secreted pathogen proteins, termed effectors are important contributors to disease development but may also act as avirulence proteins to trigger gene-for-gene resistance in plants carrying the corresponding disease resistance (R) genes.
We therefore sequenced and de-novo assembled the transcriptome of a *P. palmivora* isolate during infection of *Nicotiana benthamiana*, a host plant and pathology model system. Using a newly developed secretome-prediction pipeline we identified all *P. palmivora* effector proteins.
We found that several effector genes were conserved among worldwide *P. palmivora* isolates suggesting essential roles during the infection process. In support, the core effector REX3 suppressed plant secretory processes by targeting cellular trafficking components.
To identify potential triggers of the plant immune system we surveyed effector homology to known *Phytophthora* avirulence proteins. When expressed together with a potato R-gene, some members of a candidate effector family trigger an immune response. Secondary structure aided computational prediction enabled us to identify a single amino acid polymorphism controlling the effector's avirulence activity. Plants carrying the potato R-gene are completely resistant to leaf infection. Taken together this work suggests that existing resistance resources may be considered to engineer *P. palmivora* resistance in oil palms.

W735: Palm Genetics and Genomics
Gene Expression Analysis of the Response of Oil Palm (*Elaeis guineensis*) Clones to *Phytophthora palmivora* Inoculations *in vitro*

Kelly Avila, Cenipalma, Barrancabermeja, Colombia, Rodrigo Andrés Avila, Cenipalma, bogota, Colombia and **Hernán Mauricio Romero**, Universidad Nacional de Colombia and Cenipalma, Bogota, Colombia

W736: Palm Genetics and Genomics

Facing the Challenges of Coconut Genetic Improvement: The Coconut Genome

Luc Baudouin, CIRAD, Montpellier, France, Alix Augusto Armero, Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Montpellier, France, Stéphanie Bocs, CIRAD, UMR AGAP, Montpellier, France, Dominique This, SupAgro Montpellier, Montpellier, France, Auguste Emmanuel Issali, CNRA, Port Bouet, Cote d'Ivoire and Yaodong Yang, CATAS, Wenchang, Hainan, China

Coconut is an integral part of the livelihood of millions of farmers in the coastal tropical areas and is facing an important mutation. While its place as a major cheap oil crop on the international market is slowly declining, promising markets are emerging, such as coconut water, virgin coconut oil, and sugar among others. Coconut is regaining its tradition role of multi-usage crop, leading to new opportunities. It is however facing serious threats such as widespread lethal diseases and insect pests. To meet these challenges, coconut genetic improvement should renew its practices and objectives. But it is a difficult task due to long generation duration, low planting densities and low prolificacy. Advances in coconut genomic studies will improve its efficiency in several ways: Neutral markers allow broadening the genetic base of selection; QTLs based on mapping populations or on whole genome association studies (GWAS) will reduce the time and the areas needed to establish a breeding program. Comparative genomics and transcriptomics provide an in-depth understanding of metabolic pathways involved in production, product quality and adaptation to biotic and abiotic stress. A coconut genome draft was published recently and will be converted into a reference sequence thanks to high-density linkage mapping. Other sequencing efforts have been undertaken, whose combined results will provide a preliminary basis for characterizing coconut genetic diversity at the gene level to be completed by more resequencing. A revival of coconut genetic improvement will depend, among other elements, on high quality phenotyping, in conjunction with large SNP sets, transcriptomics and comparative genomics.

W737: Palm Genetics and Genomics

Coconut Genetics and Genomics for Host Insect Resistance

Hayde F. Galvez[1], Darlon V. Lantican[1], Maria Luz J. Sison[1], Roanne R. Gardoce[1], Barbara L. Caoili[1], Alma O. Canama[1], Melvin P. Dancel[1], Anand Noel C. Manohar[1], Romnick A. Latina[1], Cris Q. Cortaga[1], Don Serville R. Reynoso[1], Michelle S. Guerrero[1], Susan M. Rivera[2], Ernesto E. Emmanuel[2], Cristeta A. Cueto[2], Consorcia E. Reano[1], Ramon L. Rivera[2], Maria Salve C. Vasquez[1], Robert Patrick M. Cabangbang[1], Edward Cedrick J. Fernandez[1], Susan R. Strickler[3] and Lukas A. Mueller[3], (1)University of the Philippines Los Baños, College, Laguna, Philippines, (2)Philippine Coconut Authority (PCA), Quezon City, Philippines, (3)Boyce Thompson Institute, Ithaca, NY

The Philippines is the second world supplier of coconut (*Cocos nucifera* L.) products. Used to be the topmost producer/supplier, the country however has been threatened with serious production constraints including the recent outbreak of coconut scale insect (CSI) in major coconut regions. To facilitate the development of insect resistant variety, advancements in genomics and related technologies are harnessed for their optimum integration in the coconut breeding program. Coconut NGS reads were generated by sequencing the whole gDNA of Catigan Green Dwarf (CATD) coconut variety using several sequencing platforms i.e. 50X Illumina Miseq, 15X PacBio SMRT, and Dovetail Chicago sequencing. Based on combined analysis, a total genome sequence length of 2.1 Gb consisting of 8,062 scaffolds with N50 value of 569 kb was assembled. The genome assembly and initial gene models were uploaded in a local genome database and utilized to develop DNA markers targeting candidate genes for insect resistance. The genome sequence of CSI was also characterized and a species-specific DNA marker system was developed. Employing both 'Choice' and 'No-Choice' tests, 73 core coconut germplasm and on-farm outstanding selections were assayed for host resistance against CSI. The coconut materials were also mined for point-mutations at the candidate gene sequences, which are being associated to the differential CSI host response. RNA-sequence data from these differential gene expressions are also being analyzed; output of which will be mapped back to the coconut genome for a targeted functional annotation. Through forward/reverse genetics correlation studies coupled with relevant genome and bioinformatics data, the mechanism of host insect resistance will be elucidated in selected coconut varieties.

W738: Palm Genetics and Genomics

Date Palm Sex Determination and the Basis of Dioecy in the Genus *Phoenix*

Maria Fernanda Torres[1], Lisa Sara Mathew[1], Yasmin A. Mohamoud[1], Andrew Clark[2], Karsten Suhre[1] and **Joel A. Malek**[3], (1)WCMC-Qatar, Doha, Qatar, (2)Cornell University, Ithaca, NY, (3)Weill Cornell Medical College, Doha, Qatar

The date palm is a dioecious tree of commercial importance in countries of the Middle East and North Africa among others. The tree is dioecious and it is impossible to physically distinguish the female date producing trees from the male trees until approximately six years after planting when flowering begins.

In our previous work to understand the genetic control of sex-determination in the date palm we identified regions of the genome linked to sex and showed that date palm employs an XY system. However, the sex-linked region spans multiple megabases and identification of the original mutations is challenging given recombination suppression in this large region. Others have shown that sex-determination likely arose prior to speciation in the genus. Therefore, to identify potential mutations original to sex-determination in the genus we have sequenced male and female individuals from all species in the genus. We then searched for short sequencers (kmers) linked to sex in all males. This was followed by BAC and phased, single-molecule sequencing of regions surrounding the identified kmers.

Here we show that, using genus-wide sequencing identified a small number of genes that appear original to sex-determination in *Phoenix*. The region appears syntenic to an oil palm genome scaffold, helping understanding the steps to dioecy. Functional analysis of a majority of these genes in other monocots has demonstrated their critical role in flowering. Our data suggest that Phoenix ancestors likely passed through gynodioecy prior to dioecy supporting a two-mutation model for sex-determination.

This study was made possible by grant NPRP-EP X-014-4-001 from the Qatar National Research Fund (a member of Qatar Foundation).

W739: Palm Genetics and Genomics

Marker Development for the Study at Micro- and Macro-Evolutionary Time Scales in Neotropical Palms

Marylaure de la Harpe[1], Jaqueline Hess[1], Oriane Loiseau[2], Nicolas Salamin[2], Christian Lexer[3] and **Margot Paris**[3], (1)University of Vienna, Vienna, Austria, (2)University of Lausanne, Lausanne, Switzerland, (3)University of Fribourg, Fribourg, Switzerland

With the rapid advances of next generation sequencing technologies, genome-wide patterns of recent speciation and adaptation have been successfully characterized in emblematic systems such as cichlid fish, Stickleback or *Heliconius* butterfly. In palms, and especially in *Geonoma* genus, genomic studies at the species and population levels are more difficult to address due to the lack of genomic markers suitable for the genotyping of high number of taxa that diverged recently (up to 18.5 My, Roncal et al. 2010). To fill this gap, we used a whole genome sequencing approach to develop target sequencing for 4'184 molecular markers of 1'300bp length in average, including 4'051 genes and 133 non-genic regions. These markers were chosen to cover a wide range of mutation rates allowing future studies at the genus, species and population levels. A special emphasis was given to the avoidance of large indels and duplications in the marker selection. The bait set was effective not only for Geonoma species, but also for species belonging to the 3 palm subfamilies tested (*Arecoideae*, *Ceroxyloideae* and *Coryphoideae*), with high mapping rates, specificity and efficiency. The number of high quality SNPs detected both at subfamily and at the population levels allowed successful analyses across micro- and macro-evolutionary time scales.

W740: Palm Genetics and Genomics

From Genome to Field: Sime Darby Deployment Technologies in Oil Palm

Sukganah Apparow, Sime Darby Technology Centre Sdn. Bhd., Serdang, Selangor, Malaysia

As the world's largest producer of certified sustainable palm oil, Sime Darby has an utmost focus in realizing the potential of its oil palm materials through development of genetic models for direct selection and *in silico* breeding. In 2016, 100 ha of high-yielding GenomeSelect™ oil palms were planted, involving the genotyping of 80,000 seedlings and 80 million genetic tests being conducted over a period of 10 months in order to select the best 18,000 for planting. We have established mass sampling techniques capable of collecting 40,000 fresh leaf samples and high-throughput SNP genotyping platforms with the potential to generate 5 million data points on a monthly basis. In-house bioinformatics tools and database have been developed for Big Data analysis and mobile device technologies synchronized to the internal database allows hassle-free selection of the high-yielding palms in the field. The in-house molecular marker deployment technologies have also allowed a reliable quality control system to be established to facilitate selection of parental palms based on genetic breeding values and to ensure correct parentage of the commercial *tenera* planting materials. Sime Darby Plantation will plant another 1,000 ha in 2018 and reach full replanting capacity in Malaysia using GenomeSelect™ materials by the year 2023. Sample tracking and quality control measures integrated throughout this pipeline from sampling to field selection that ensured the success of the GenomeSelect™ planting will be discussed along with their implementation challenges.

W741: Perennial Grasses

Tools for Optimizing Lignocellulosic Biomass Quality of Miscanthus

Luisa M Trindade, Wageningen UR Plant Breeding, Wageningen, Netherlands

W742: Perennial Grasses

***Lolium perenne* - a Model to Understand Self-Incompatibility in Perennial Grass Species**

Daniel Thorogood, Aberystwyth University, Aberystwyth, Ceredigion, United Kingdom

W743: Perennial Grasses

Summer Dormancy: A Mechanism for Cool-Season Perennials to Avoid Harsh Summer

Malay C. Saha, Noble Research Institute, Ardmore, OK

W744: Perennial Grasses

Genome-Wide Association and Transcriptome Analyses of Flowering Time in Switchgrass

Megan Sue Taylor, Purdue University, Department of Agronomy, IN, Carl-Erik I. Tornqvist, Department of Agronomy and DOE Great Lakes Bioenergy Research Center, University of Wisconsin - Madison, Madison, WI, Xiongwei Zhao, Purdue University, West Lafayette, IN, Paul Grabowski, USDA-ARS, Madison, WI, Michael Casler, USDA Dairy Forage Research Center, Madison, WI and **Yiwei Jiang**, Department of Agronomy, Purdue University, West Lafayette, IN

The timing of phase change from juvenile (vegetative) to adult with reproductive competence is a key factor influencing biomass yield of switchgrass. A decline in biomass yield is typically observed in switchgrass immediately following completion of flowering. The use of late-flowering switchgrass genotypes could be an effective way to increase biomass production in the northern USA. However, genetic mechanisms of flowering time are not well understood in switchgrass. The objectives of the study was to identify signals and candidate genes related to flowering time in switchgrass. The heading and anthesis dates of four populations (reciprocal crosses) derived from upland (early flowering) and lowland (late flowering) were recorded in two locations Lafayette, IN and Dekalb, IL. The heading and anthesis dates ranged from 176 to 224 and 191 to 254 days in Lafayette, IN and from 181 to 270 and 201 to 270 days in DeKalb, IL, respectively. Genome-wide association study (GWAS) detected three signals for heading and two signals for anthesis. Linkage mapping using one of the populations confirmed one signal for heading date located on chromosome 2B. The integrated GWAS and transcriptome analyses of individuals with contrasting heading and anthesis dates identified several genes that could be associated with flowering time including PSEUDO RESPONSE REGULATOR 5. Knowledge generated from the project will aid breeding programs in developing varieties of switchgrass that fully utilize the growing season and achieve high biomass yield.

W745: Perennial Grasses

Comparative Genomics of the Perennial Model Grass *Brachypodium sylvaticum* and its Annual Relative *B. distachyon*
Sean Gordon, DOE Joint Genome Institute, Walnut Creek, CA
Most of our knowledge of grass biology and responses to abiotic stress is based on studies of annual grasses like rice, wheat and barley. However, most of the grasses being developed as biomass crops are perennial. Annuals and perennials differ in many important physiological and developmental aspects, some of which may be particularly relevant to stress tolerance. Therefore, there is a pressing need for a tractable perennial grass model to study topics like abiotic stress tolerance and, in particular, to rapidly test transgenic approaches before moving into biomass crops. We previously established key tools that allow *Brachypodium sylvaticum* to be used as a perennial model grass including a highly efficient transformation protocol and inbred lines. Missing from this list is a high quality genome sequence, a prerequisite for a modern model organism. To fill this need, we sequenced the genome of *B. sylvaticum* to ~80x depth using PacBio long-read technology yielding an initial assembly of 358Mb of sequence in 1,118 contigs with a N50 contig length of 874Kb. These contigs were ordered and orientated into 9 pseudomolecules using a high-density genetic map created from 288 F2 individuals. The final reference genome is annotated with 36,927 gene loci. A deep RNA-Seq expression atlas covering 16 different tissues, including a subset of tissues sampled at different developmental stages and under a variety of abiotic stresses, was generated, allowing us to define gene regulatory networks within *B. sylvaticum*. A cross species comparison is being performed with the closely related small annual grass, *Brachypodium distachyon*.

W746: Perennial Grasses
Leaf Epicuticular Wax Load Segregation in the *Panicum virgatum* 4WCR Population
Jennifer Bragg, USDA-ARS, WRRC, Albany, CA, Lisa Chanbusarakum, USDA-ARS, WRRC, Albany, CT, Thomas Juenger, University of Texas at Austin, Austin, TX and Christian Tobias, USDA-ARS, Western Regional Research Center, Albany, CA
The C4 perennial grass *Panicum virgatum* (switchgrass) is a native species of the North American tallgrass prairie. This adaptable plant can be grown on marginal lands and serve as a renewable resource to both aid in soil and water conservation and provide biomass for forage or energy production. The two major switchgrass ecotypes, lowland and upland, present a range of desirable traits and underlying genetic diversity that can be mined for trait improvement breeding programs. Lowland adaptation to riparian zones versus upland acclimation to drier regions resulted in differences in tiller number and size, growth habit, flowering time, and waxy blue-green leaf color. The outbred 4WCR mapping population of 400 F_2 lines was created to take advantage of the diversity within two lowland (AP13 and WBC3) and two upland (DAC6 and VS16) tetraploid cultivars. Among these varieties, visual inspection suggests differences in epicuticular wax load, a phenotype associated with traits including heat and drought tolerance, UV protection, and defense against pathogens. The objective of our experiments was to identify the genomic variation underlying these differences. We used a quantitative colorimetric assay to measure epicuticular wax of leaf samples from the 4WCR F_0 founders, F_1 hybrids, and F_2 mapping population grown at four field sites with latitudes ranging from 30 to 42 °N. Preliminary analyses of these datasets using the 4WCR linkage map identify two high confidence QTLs associated with wax load and also reveal GxE effects.

W747: Plant and Animal Paleogenomics
Eutherian Chromosomes in the Light of Evolution
Jaebum Kim[1], Marta Farré Belmonte[2], Loretta Auvil[3], Boris Capitanu[3], Denis M. Larkin[2], Jian Ma[3] and **Harris A. Lewin**[4], (1)Konkuk University, Seoul, Korea, Republic of (South), (2)Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, United Kingdom, (3)University of Illinois at Urbana-Champaign, Urbana, IL, (4)UC Davis, Davis, CA
Chromosome rearrangements are a hallmark of genome evolution and essential for understanding the mechanisms of speciation and adaptation. Determining the types and chronological order of chromosome rearrangements over evolutionary time scales has been a difficult problem due primarily to the lack of high quality, chromosome-scale genome assemblies that are necessary for reliable reconstruction of ancestral genomes. In addition, for genome-wide comparisons that require resolving large numbers of rearrangements of varying scale, determining ancestral chromosomal states is challenging both methodologically and computationally. In my talk, I will present recent chromosome reconstruction results obtained using a new method developed by our chromosome evolution collaborative group, called DESCHRAMBLER, which uses as input syntenic fragments constructed from whole-genome comparisons of both high quality chromosome-scale and fragmented assemblies. We applied DESCHRAMBLER to sequenced genomes of 21 species that included representatives of 10 eutherian orders. Seven ancestral genomes leading to human were reconstructed, including the ancestor of all placental mammals. From these reconstructions, a detailed picture of chromosome rearrangements that occurred during ~105 million years of eutherian evolution was revealed. Our results provide an evolutionary basis for comparison of genome organization of all eutherians, and will facilitate greater understanding of the role of chromosome rearrangements in adaptation, speciation, and the etiology of inherited and spontaneously occurring diseases. With ongoing efforts to sequence many vertebrate genomes, it will be possible to extend reconstructions deeper into evolutionary time, and thus provide a more detailed picture of chromosome evolution in other vertebrate classes. Ultimately, it should prove possible to determine the ancestral vertebrate karyotype with high confidence order and orientation of syntenic fragments.

W748: Plant and Animal Paleogenomics
Comparative Analysis of Mammal and Angiosperm Phylogenomic Synteny Networks
Tao Zhao, Wageningen University, Wageningen, Netherlands and M. Eric Schranz, University of Amsterdam, Amsterdam, Netherlands
Comparative phylogenomic synteny (genomic context) analysis holds great promise for the inference of gene and genome evolutionary history. Utilizing the extensive available whole-genome resources, we have built complete microsynteny (local conserved gene order) networks for all genes of 87 mammalian and 107 angiosperms genomes, respectively. Thus, we can directly compare genome dynamics of these two major clades that have evolved and radiated during the last ~170 million years. To interpret the entire synteny network, we exploited network statistical parameters (i.e. average clustering coefficient, retention percentage, cluster sizes) to characterize and quantify various evolutionary features (i.e. conservation vs diversity) of gene families in a phylogenomic context. In addition, we dissected the composition and size

distribution of all synteny clusters, which provide intriguing insights into the differing genomic architectures and dynamics of mammals and flowering plants. Sufficient representative genomes for synteny network construction in this study provide us clearer phylogenetic profiling patterns of synteny clusters. We will highlight several representative examples of lineage-specific clusters (i.e. unique genomic changes) that signal potential links between genomic context variation and the evolution of lineage-specific phenotypic traits.

W749: Plant and Animal Paleogenomics
Dissecting Evolution and Disease using Comparative Vertebrate Genomics – Power from 200 Mammals
Kerstin Lindblad-Toh, Uppsala University, Uppsala, Sweden
With the generation of more than 100 vertebrate genome sequences (including ~50 mammals) in less than 25 years, the key question arises of how these resources can be used in new or ongoing projects. In the past, this diverse collection of sequences from humans as well as model and non-model organisms has been used to annotate the human genome and to increase the understanding of human disease.
In the future, comparative vertebrate genomics in conjunction with additional genomic resources will yield insights into the processes of genome function, evolution, speciation, selection and adaptation, as well as into the quantification of species diversity. Here I will discuss principles and key results from earlier vertebrate comparative genetics and genomics projects. I will also present the ongoing 200 mammals project.
For the 200 mammals project we have selected as broad a set of placental mammals as possible, pragmatically including factors such as availability of DNA at sufficient quality, with research communities interested in the organisms. As the sequencing has been performed by Illumina short reads, species could be selected even if the DNA quality was not optimal. For a diverse subset we have added long-range contiguity by Dovetail. The genome alignments are reference free and can therefore serve to annotate any of the 200 mammals. This project strives to annotate the human genome with single base constraint. It will also inform vertebrate biology and evolution and generate resources that could be important for conservation genomics.

W750: Plant and Animal Paleogenomics
The Impact of Ancient Polyploidy on Genome Evolution in Poales and other Monocots
Michael McKain[1,2], Haibao Tang[3], Joel McNeal[4], Saravanaraj Ayyampalayam[5], Claude W. dePamphilis[6], Thomas J Givnish[7], J. Chris Pires[8], Dennis Wm. Stevenson[9] and Jim Leebens-Mack[5], (1)The University of Alabama, Tuscaloosa, AL, (2)University of Alabama, Tuscaloosa, AL, (3)Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fuzhou, AZ, China, (4)Kennesaw State University, Kennesaw, GA, (5)University of Georgia, Athens, GA, (6)The Pennsylvania State University, University Park, PA, (7)Department of Botany. UW-Madison, Madison, WI, (8)Division of Biological Sciences, University of Missouri, Columbia, MO, (9)New York Botanical Garden, Bronx, NY
Comparisons of flowering plant genomes reveal multiple rounds of ancient polyploidy characterized by large intragenomic syntenic blocks. Three such whole-genome duplication (WGD) events, designated as rho (ϱ), sigma (σ), and tau (τ), have been identified in the genomes of cereal grasses. In order to investigate how WGDs have influenced species diversification, evolutionary innovations, and genome composition such as the GC profile of protein-coding sequences, the precise timing of these events must be inferred. Phylogenomic analysis of protein-coding genes from sequenced genomes and transcriptome assemblies of 35 species, representing all families in Poales, were used to date these WGD events. The rho event was found to have occurred prior to the diversification of Poaceae while sigma occurred prior to the diversification of the order Poales but after the Poales-commelinid split. The more ancient tau event was placed prior to the divergence of Asparagales and other monocots. As part of this study, the efficacy of phylogenetic signal from WGD was compared using putative paralogs derived from syntenic blocks, synonymous substitution frequency plots, and gene trees. The gene tree-based approach identified the largest number of putative WGD paralogs with ~48% more per WGD than the syntenic block approach. Gene families exhibiting high GC content are underrepresented among those with duplicate genes that persisted following these genome duplications. Ultimately, genome duplications had little overall influence on lineage-specific changes in the GC content of coding genes contrary to previous hypotheses.

W751: Plant and Animal Paleogenomics
Ancestral Genome Reconstruction on Whole Genome Level
Jijun Tang, University of South Carolina, Columbia, SC
Jijun Tang. Department of Computer Science and Engineering University of South Carolina, Columbia, SC 29208, USA
Comparative genomics, evolutionary biology, and cancer researches require tools to elucidate the evolutionary trajectories and reconstruct the ancestral genomes. Various methods have been developed to infer the genome content and gene ordering of ancestral genomes by using such genomic structural variants. In this talk, we will review the principles and algorithms of these approaches that can reconstruct the ancestral genomes on the whole genome level. We will discuss the advantages and limitations of these approaches in dealing with various genome datasets, evolutionary events, and reconstruction problems. We select four most famous and powerful approaches from both distance/event-based and homology/adjacency-based categories to analyze and compare their performances in dealing with different kinds of datasets and evolutionary events. Based on our experiment, GASTS has the best performance in solving the problems with equal genome contents that only have genome rearrangement events. The web server MLGO achieves the best performance in solving the problems with unequal genome contents that have all possible complicated evolutionary events

W752: Plant and Animal Paleogenomics
Reconstructing Wheat Evolutionary History
Caroline Pont, INRA, Clermont-Ferrand, France
Polyploidization have been reported as a major evolutionary force during plant paleohistory. Following the triplication reported in *Brassiceae* ~10 million years ago, and at the basis of rosids ~100 million years ago, bias in organisation and regulation, known as subgenome dominance, has been reported between the three post-polyploidy compartments referenced to as less fractionated (LF), medium fractionated (MF1) and more fractionated (MF2), that have been proposed to derive from an hexaploidization event involving ancestor intermediate of 7-14-21

chromosomes. Modern bread wheat experienced similar paleohistory during the last half million year of evolution opening a new hypothesis where the wheat genome is at the earliest stages on the road of diploidization through subgenome dominance driving asymmetry in gene content, gene expression abundance, transposable element content as dynamics and epigenetic regulation between the A, B and D subgenomes.

W753: Plant Chromosome Biology

Molecular Organization of Mitotic Metaphase Chromosomes in Higher Plants

Tomas Beseda[1], Martin Mascher[2], Beata Petrovska[1], Axel Himmelbach[2], Hana Simkova[3], Marek Sebela[4], Nils Stein[2] and **Jaroslav Dolezel**[1], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany, (3)IEB, Olomouc, Czech Republic, (4)Palacky University in Olomouc, Olomouc, Czech Republic

Eukaryotic genomes undergo distinctive structural changes during cell cycle. A faithful transmission of hereditary information to daughter cells and progenies is facilitated by the formation of compacted rod-shaped mitotic chromosomes. However, it is not clear how DNA is organized in the 3D chromosome space to fit a minute volume, and only a few chromosomal proteins are known. The progress in DNA sequencing technologies, computational modelling and proteomics provides new opportunities to reveal the mystery. The first insights into the 3D organization of DNA in human mitotic chromosomes were obtained and it was found that a large portion of mitotic chromosomes is not composed of chromatin, but other proteins. While highly synchronized populations of mitotic cells could be used in unicellular eukaryotes and metazoans, high degree of mitotic metaphase synchrony is hard to achieve in flowering plants. In order to circumvent this problem, we used flow cytometry to purify chromosomes isolated from synchronized meristem root tip cells. The chromosomes have well preserved morphology and can be obtained in large quantities. The use of Hi-C on purified chromosomes of barley revealed a helical arrangement of nested loops with ~40Mb DNA per one helical turn. This corresponds to fifteen to twenty turns per chromosome and our result supports the models proposed after the observation of artificially decondensed human mitotic chromosomes. Proteomic analysis of purified barley chromosomes led to the identification of large set of proteins, significantly expanding the number of proteins known to contribute to the formation and function of plant mitotic chromosomes.

W754: Plant Chromosome Biology

Live Cell CRISPR-Imaging in Plants

Andreas Houben, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and Steven Dreissig, Simon Schiml, Patrick Schindele, Oda Weiss, Twan Rutten, Veit Schubert, Solmaz Khosravi, Solmaz Khosravi, Michael Florian Mette, Holger Puchta, Andreas Houben

Elucidating the spatio-temporal organization of the genome inside the nucleus is imperative to understand the regulation of genes and non-coding sequences during development and environmental changes. Emerging techniques of chromatin imaging promise to bridge the long-standing gap between sequencing studies which reveal genomic information and imaging studies that provide spatial and temporal information of defined genomic regions. Here, we demonstrate such an imaging technique based on two orthologues of the bacterial CRISPR-Cas9 system. By fusing eGFP/mRuby2 to the catalytically inactive version of *Streptococcus pyogenes* and *Staphylococcus aureus* Cas9, we show robust visualization of telomere repeats in live leaf cells of *Nicotiana benthamiana*. By tracking the dynamics of telomeres visualized by CRISPR-dCas9, we reveal long range telomere movements of up to 2 μm within 30 minutes during interphase. Furthermore, we show that CRISPR-dCas9 can be combined with fluorescence-labelled proteins to visualize DNA-protein interactions *in vivo*. By simultaneously using two dCas9 orthologues, we pave the way for imaging of multiple genomic loci in live plants cells. CRISPR-imaging bears the potential to significantly improve our understanding of the dynamics of chromosomes in live plant cells.

W755: Plant Chromosome Biology

The Role of Retrotransposons in Maize Centromeres

Gernot Presting, Kevin Schneider, Daniel Laspisa and Vishal Negi, University of Hawaii, Honolulu, HI

Centromeric retrotransposons (CR) of maize target centromeres during integration into the genome. Little is known about their role in centromere function. A detailed map of the maize centromeres is essential to better understand the integration and turnover rates of CR elements. To this end we have manually edited 132 Mb of the newest maize reference genome B73 RefGen_v4 covering the ten centromeres. We added over 9 Mb of sequence, including over 1 Mb of CR sequence. The improved centromere sequences confirm earlier observations, based on the lower quality B73 RefGen_v2 genome, that CR2 elements insert into each centromere approximately once per century. The role of these retrotransposons in centromere function is under investigation.

W756: Plant Chromosome Biology

Genus *Cuscuta* as a Model to Study Transition between Monocentric and Holocentric Chromosomes

Jiri Macas[1], Pavel Neumann[1], Petr Novak[1], Tae-Soo Jang[1], Veit Schubert[2], Jana Cizkova[3], Sonja Klemme[1], Jaroslav Dolezel[3] and Andreas Houben[2], (1)Biology Centre CAS, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, (3)Institute of Experimental Botany, Olomouc, Czech Republic

The genus of parasitic plants *Cuscuta* (*Convolvulaceae*) is unique in containing species differing in their centromeric organization, with holocentrics being found in the subgenus *Cuscuta,* while the other related subgenera, *Grammica* and *Monogyna*, comprise species with monocentric chromosomes. Based on the genus phylogeny, species with holocentric chromosomes most likely originated from monocentric ancestor(s), providing an opportunity to study this transition in a set of closely related species. In this talk, we will present initial results from our comparative study of genome size, repetitive DNA and chromosome structure in *Cuscuta* and discuss these features with respect to the differences in centromere organization.

W757: Plant Chromosome Biology

Development of Wheat-*Haynaldia villosa* Alien Chromosome Lines and their use in Gene Mining and Wheat Breeding
Xiue Wang, Peidu Chen, Haiyan Wang, Haojie Sun, Keli Dai, Ruiqi Zhang, Heng Zhang, Zongkuan Wang, Aizhong Cao, Jin Xiao, Liping Xing and Shouzhong Zhang, Nanjing Agricultural University, Nanjing, China
Wild relatives provide rich gene resources for wheat breeding. *Haynaldia villosa* (2n=14, genome VV), is a diploid wild species and has proved to be resistant to several wheat diseases, such as powdery mildew, wheat yellow mosaic etc. The development of alien translocation lines conferring useful genes is the most effective way for the utilization of alien genes. In Cytogenetics Institute if Nanjing Agricultural University, a research platform for the induction of alien chromosome structural variation and for effective identification of alien chromatin has been established. A wheat-*H. villosa* alien translocation pool was constructed and the their chromosome constitution was characterized. Genes conferring resistances to powdery mildew, wheat yellow mosaic virus, strip rust as well as loci controlling grain quality has been assigned specific regions of *H. villosa* chromosomes. The whole arm translocation lines carrying useful genes have been released and utilized in breeding programs. Using the T6VS/6AL translocation carrying the powdery mildew resistance gene *Pm21* and the strip rust resistance gene *Yr26*, about 30 wheat varieties have been developed and commercially released in China.

W758: Plant Chromosome Biology
Chromatin Packing and Interactions in Rice
Lei Gong and Bao Liu, Northeast Normal University, Changchun, China
The non-random chromatin packing in the nucleus plays important roles in the regulation of gene expression and genome function. Here, we present a Hi-C analysis of the global chromatin interaction patterns in rice (*Oryza sativa* L.), a model species of monocot as well as a staple crop. We characterized both global and local chromatin interaction patterns using Hi-C assay. Based on the Hi-C interaction map, we explored the hierarchical chromatin structural features. It is concluded that genomic composition, epigenetic modification, and transcriptional activity could collaboratively achieve both global and local chromatin packing in rice.

W759: Plant Cytogenetics
The FANCM Helicase is Required for Crossover Formation in Wheat
Stuart Desjardins, University of Leicester, Leicester, United Kingdom

W760: Plant Cytogenetics
Sequencing of Single Pollen Nuclei Reveals Meiotic Recombination Events at Megabase Resolution
Andreas Houben, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and Steven Dreissig, Jörg Fuchs, Axel Himmelbach, Martin Mascher, Andreas Houben
Meiotic recombination is a fundamental mechanism to generate novel allelic combinations which can be harnessed by breeders to achieve crop improvement. The recombination landscape of many crop species, including the major crop barley, is characterized by a dearth of recombination in 65% of the genome. In addition, segregation distortion caused by selection on genetically linked loci is a frequent and undesirable phenomenon in double haploid populations which hampers genetic mapping and breeding. Here, we present an approach to directly investigate recombination at the DNA sequence level by combining flow-sorting of haploid pollen nuclei of barley with single-cell genome sequencing. We confirm the skewed distribution of recombination events towards distal chromosomal regions at megabase resolution and show that segregation distortion is almost absent if directly measured in pollen. Furthermore, we show a bimodal distribution of inter-crossover distances, which supports the existence of two classes of crossovers which are sensitive or insensitive to physical interference. We conclude that single pollen nuclei sequencing is an approach capable of revealing recombination patterns in the absence of segregation distortion.

W761: Plant Cytogenetics
Chromosome Rearrangements Caused by Double Monosomy in Wheat-Barley Group-7 Substitution Lines
Tatiana V. Danilova, Bernd Friebe and Bikram S. Gill, Kansas State University, Manhattan, KS
Interspecific hybridization is one of the driving forces in plant speciation, producing allopolyploids or diploids with rearranged genomes. The process of karyotype reshaping following homoploid interspecific hybridization has not been studied experimentally. Interspecific hybridization is widely used in plant breeding to increase genetic diversity and introgress new traits. Numerous introgression stocks were developed for hexaploid wheat *Ttiticum aestivum* L. (2n=6x=42, genome AABBDD). Double monosomic lines, containing one alien chromosome from the tertiary gene pool of wheat and one homoeologous wheat chromosome represent a simplified model for studying chromosome rearrangements caused by interspecific hybridization. The pairing of a chromosome from the tertiary gene pool with a wheat homoeologue is restricted by the activity of the wheat *Ph1* gene, thus rearrangements caused by chromosome breakage followed by the fusion of the broken arms can be expected. We analyzed chromosome aberrations in four sets of lines, that originated from double monosomics of barley (*Hordeum vulgare* L.) chromosome 7H and wheat group-7 chromosomes with dicentric or ring chromosomes. The dynamics of wheat-barley dicentric chromosomes during plant development was followed, and increased diversity of rearrangements observed. Besides the targeted group-7, other wheat chromosomes were involved into rearrangements, as chromosomes broken at centromeric region fused with other chromosomes. In some cells multi-centric chromosomes were observed. The structure and dosage of the introgressed barley chromatin were changed. The transmission of the rearrangements to the progenies was analyzed. The observed aberrations emphasize the importance of cytogenetic screening in gene introgression projects.

W762: Plant Cytogenetics
The Sum is Greater than its Parts: Meiotic Adaptation Mediated by Two-Way Gene Flow between *Arabidopsis arenosa* and *A. lyrata*
Levi Yant, John Innes Centre, Norwich, United Kingdom

W763: Plant Cytogenetics

Meiotic Stability in Allohexaploid Brassica Hybrids

Roman Gaebelein, Justus Liebig University Giessen, Giessen, Germany

The genus *Brassica* comprises some of the most important vegetable and oil crops worldwide. Many of these crops, such as *Brassica napus*, are allopolyploid, i.e. contain two genomes from different progenitor species. Studying mechanisms of polyploid formation in the genus *Brassica* not only broadens our understanding of polyploid speciation but can also benefit plant breeding, creating new crop types to increase and maintain yields for the future. We produced a number of allohexaploid hybrids by crossing *B. napus* ($A^nA^nC^nC^n$) with *B. carinata* ($B^cB^cC^cC^c$) and crossing the F_1 Hybrids ($2n = A^nB^cC^nC^c$) to *B. juncea* ($A^jA^jB^jB^j$). Many of the trigenomic tetraploid F_1 hybrids produced unreduced gametes ($n = A^nB^cC^nC^c$) that led to the successful production of hybrids of the desired karyotype ($2n = A^nA^jB^cB^jC^nC^c$). Advanced generations of these hybrids have shown a significant increase in seed set, suggesting increased genome stability. Furthermore, analysis of chromosome pairing during diakinesis and metaphase I showed increasingly diploidised chromosome behaviour in plants with particularly high seed yield. Marker information from the new *Brassica* 90K Illumina Infinium SNP genotyping array was also used for association analysis in order to link fertility data to genomic regions involved in the regulation of meiotic pairing. If allohexaploid *Brassica* hybrids can be established as a stable species they could be used as a new oil crop, benefiting from a diverse pool of genes for improved resistance and quality traits. This research also sheds light on how meiotic regulation may be involved in the evolution of new allopolyploid species.

W764: Plant Cytogenetics

A Combinatorial Mechanism for the Removal Interlocks in Meiotic Chromosomes

Chris Franklin[1], Marina Martinez-Garcia[2] and Eugenio Sanchez-Moran[1], (1)University of Birmingham, Birmingham, United Kingdom, (2)Harvard Medical School, Boston, MA

During the zygotene stage of meiosis chromosome synapsis and homologous recombination frequently lead to the formation of structural interlocks between entangled chromosomes. As the persistence of interlocks in metaphase I would be highly deleterious to chromosome segregation they must be removed by pachytene. Cytological studies in several species have revealed breaks in the chromosome axes and SC at zygotene associated with entanglements. By pachytene, these axial gaps are no longer observed. Hence it is proposed that formation of a transient chromosomal break allows resolution of the interlock. It is suggested that this might be accomplished by a type II topoisomerase, although, this would require some modification of its normal activity (Zickler and Kleckner, 1999). A second proposal invokes chromosome movement towards the end of the other bivalent, combined with de- and re-polymerisation of the SC and telomere detachment from the nuclear envelope to liberate the trapped chromosome (Rasmussen and Holm, 1980). However, experimental evidence has been lacking. Analysis of a hypomorphic *topII* mutant and a meiosis specific *topII* RNAi knock-down of *Arabidopsis thaliana* using immunocytochemistry and structured illumination microscopy (SIM) has now enabled us to confirm a central role for the protein in interlock resolution. Furthermore, analysis using a nucleoporin *nup136* mutant, which affects chromosome movement, reveals that although TOPII activity is required for the removal of some interlock structures, for others chromosome movement is also necessary. Thus our study suggests that at least two mechanisms participate in interlock removal.

MMG was supported by the 'COMREC' network FP7 ITN-606956 grant.

W765: Plant Dormancy Workshop

Gene Expression in the Floral Buds of Sweet Cherry Responds to Both Day Length and Temperature Including a Reset of the Cell Cycle.

Paul A. Wiersma and Denise Neilsen, Agriculture AgriFood Canada, Summerland, BC, Canada

The winter dormancy characteristics of temperate fruit species such as sweet cherry limit their effective production to a narrow range of climates. The growing region for sweet cherry in British Columbia, Canada has changed significantly over the last five decades including new zones further north and up-slope from the traditional orchards. There is considerable interest in projecting changing crop suitability into the future and we have combined a series of dormancy, cold hardiness, phenology and in season fruit growth models to explore changing locations for sweet cherry production. A key component of this process is the use of temperature driven models for dormancy induction, completion of endodormancy and ecodormancy. To complement the extensive observations over six years of bud cold-hardiness, bud break potential and forcing requirements in the green-house we have examined gene expression in the floral buds of these trees over the same period. Weekly samples of dissected buds from September to April for the years 2013-14 and 2016-17 were used for the construction of RNA-Seq libraries and 10-20 M reads per library obtained by Illumina sequencing. Additional analysis and corroboration of the RNA-Seq data was done with real-time RT PCR. Several distinct patterns of gene expression were observed. A spike of expression in a number of genes including transcription factors occurred at the time of optimal bud break in both years. Temperature models correctly place this activity at different times in the late fall depending on the year. Other genes follow a U-shaped expression pattern with higher levels in early fall and early spring and minimal expression corresponding to the period of ecodormancy in the coldest part of winter. These genes include several key components of the cell cycle. For every year examined this pattern followed the same course for initiation beginning at approximately 10 h of daylight and did not correspond to the different temperature profiles occurring in those years. Sweet cherry appears to respond to both chilling and short day conditions.

W766: Plant Dormancy Workshop

Mining of the Candidate Genes involved in Regulating the Chilling Requirement Fulfillment and Bud Dormancy Release of *Paeonia lactiflora* 'Hang Baishao', a Germplasm with the Superb Adaptability during Warm Winters

Jiaping Zhang, Yiping Xia, Danqing Li and Dong Zhang, Zhejiang University, Hangzhou, China

Herbaceous Peony is a world-famous flower and mainly planted in temperate or cool areas. Warm winter in subtropical regions severely hinders "the Southward Plantation of Herbaceous Peony" in the Northern Hemisphere. Studies on the dormancy, chilling requirement (CR) and relevant molecular mechanisms of peony need to be performed. Based on the chilling treatments, the optimal CR of Paeonia lactiflora Pall. "Hang Baishao" was 672.00 chilling hours or 856.08 chilling units for achieving the superior sprouting and flowering performances. During the

bud dormancy, the ratio of IAA/ABA fluctuated and starch content dropped, while the soluble sugar content and peroxidase activity rose steadily. Transcriptome sequencing were performed for the 'Hang Baishao' buds during the dormancy and sprouting. The differentially expressed genes (DEGs) could be divided into three categories, i.e. DEGs related to environmental response, metabolism, and cell growth. The "difference" in the expression patterns of SOC1 and WRKY 33 between two winters, and the "difference" of CR fulfillment periods also between these two winters, represented the interesting congruent relationships. Therefore, they are likely involved in determining the CR fulfillment period of 'Hang Baishao'. The genes related to phytochrome (PHY), heat shock protein (HSP), osmotin (OSM), dehydrin (DHN), auxin-repressed protein (ARP), repressor of GA (RGA), GA20 oxidase (GA20ox), peroxidase (PER), cyclin (CYC) and expansin (EXP) expressed actively. This study could contribute to the knowledge of the mechanisms that regulate CR and dormancy characteristics, and may be beneficial for breeding new peony cultivar that have low CR for horticulture use in subtropical regions.

W767: Plant Dormancy Workshop

Single-Base Resolution Methylome of Floral Bud Sweet Cherry (*Prunus Avium* L.) Varieties during Cold Accumulation in the Dormancy Period

Karin Rothkegel[1], Javier Cáceres-Molina[1], Tomás Carrasco-Valenzuela[1], Claudio Meneses[1,2] and Andrea Miyasaka de Almeida[1], (1)Universidad Andrés Bello, Centro de Biotecnología Vegetal, Santiago, Chile, (2)FONDAP, Center for Genome Regulation, Santiago, Chile

Epigenetic modifications can provide information about the connection between genotype and phenotype variation due to environmental conditions. In temperate perennial fruit species like sweet cherry, prolonged exposition to cold temperatures is required for dormancy release and flowering. In addition to sequence-based genetic diversity, epigenetic variation is believed to contribute to different chilling requirements and phenotypic plasticity among varieties. DNA methylation is an epigenetic modification that in plants occurs in three different contexts (CpG, CHG and CHH; H= non-guanine residue), and plays a key role in gene expression regulation. Aiming to identify regions with differential methylation, we performed whole genome sequencing of bisulfite-treated DNA from floral buds of 'Kordia' and 'Royal Dawn' at four time points of cold accumulation during dormancy. In average, 78.9% and 63.1% of total reads mapped to the 'Kordia' and 'Royal Dawn' reference genome respectively. From these, 41.6-50% corresponded to unique alignments and 21-36.9% of the reads did not align under any condition. Approximately 27% of total cytosines were methylated in all the conditions, corresponding to 38-41%, 33-34% and 25-28% of specific methylations at the CpG, CHG and CHH sites, respectively. Additionally, we observed changes in the frequency of differentially methylated regions (DMRs) during cold accumulation'. Identification of specific loci in these regions is in progress for future validation, in an attempt to describe epigenetic marks that may be contributing to differential chilling requirement. Finally, validated DMRs could be used as epigenetic biomarkers of cold accumulation in floral buds during dormancy in sweet cherry.

Funding: Consortium Biofrutales-CORFO 13CTI21520-SP05 and FONDEF G09I1008

W768: Plant Dormancy Workshop

A DNA Demethylase Overexpression Associated with the Biosynthesis and Accumulation of Flavonoids, and with Early Apical Bud Maturation in Poplar

Daniel Conde[1], Alicia Moreno-Cortés[2], Christopher Dervinis[1], José Manuel Ramos-Sánchez[2], Matias Kirst[1], Mariano Perales[2], Pablo Gónzalez-Melendi[2] and Isabel Allona[2], (1)University of Florida, Gainesville, FL, (2)Universidad Politécnica de Madrid, Madrid, Spain

Trees growing in boreal and temperate regions synchronize their growth with seasonal climatic changes in adaptive responses that are essential for their survival. In these regions, deciduous and periodic growth habits evolved into a single trait known as winter dormancy. Winter dormancy begins with growth cessation through the arrest of meristem activity and subsequent enclosure of the apical meristem into a dormant winter bud, a process called bud set. We identified a DEMETER-like (*CsDML*) cDNA from a winter-enriched cDNA subtractive library in chestnut (Castanea sativa Mill.), an economically and ecologically important species. We performed phylogenetic and protein sequence analysis, gene expression profiling, and 5-methyl-cytosine methylation immunodetection studies to evaluate the role of *CsDML* and its homolog in poplar, *PtaDML6*. We found that the signals that trigged bud dormancy in trees (short days and cold temperatures) induced *CsDML* and *PtaDML6* gene expression. Overexpression of *CsDML* in transgenic hybrid poplar accelerated short-day-induced bud formation. Buds of transgenic plants acquired the red-brown coloration typical of plants undergoing bud set, earlier than wild-type plants. This occurred alongside with the up-regulation of flavonoid biosynthesis genes and accumulation of flavonoids in the shoot apical meristem and bud scales. Our data shows that the *CsDML* gene induces bud formation needed for the survival of the apical meristem under the harsh conditions of winter.

W769: Plant Dormancy Workshop

Seasonal Dynamics of Bud Dormancy, the Cell Cycle and Metabolism in Temperate-Grown Grapevine

Santiago Signorelli[1], Yazhini Velappan[1], Dina Hermawaty[1], Patricia Agudelo-Romero[2], John A Considine[2] and **Michael J Considine**[2], (1)University of Western Australia, Perth, Australia, (2)University of Western Australia, Crawley, Australia

It is classically understood that the depth of bud dormancy in woody deciduous perennials increases during autumn and is quantitatively relieved by chilling during winter, preceding bud burst. Some earlier studies of the seasonal dynamics of bud dormancy in grapevine, however challenge this view. Here we have carried out extensive physiological analysis of grapevine bud dormancy, cell division and metabolism in commercially grown vines in the southwest of Western Australia. These data demonstrate conclusively that bud dormancy (sensu stricto) is largely transient, with an onset in late summer, and substantially relieved by mid-autumn, well-before any available models predict chilling accumulation. The dynamics of bud dormancy were not directly related to either the cell cycle or respiratory indicators. Transcriptome analysis is underway to investigate the relationships of dormancy to predicted and unknown regulators of grapevine ecophysiology.

W770: Plant Dormancy Workshop

A Novel Biclustering Method for Time Course Gene Expression Data and Its Application in Bud Dormancy Induction in Grapevine

Juan Xie, Anne Fennell and Qin Ma, South Dakota State University, Brookings, SD
Bud dormancy in grapevine is an adaptive strategy for the survival of drought, high and low temperatures and freeze dehydration stress that limit the range of cultivar adaptation. A better understanding of the biological mechanisms involved in bud dormancy is needed to promote advances in grape cultivars selection, breeding and improvement. In this big data era, the abundance of gene expression data sets provides an unprecedented opportunity to elucidate the underlying dormancy mechanism. Gene expression data can be used to identify genes with similar expression patterns across multiple conditions/time-points, i.e., co-expressed genes (CEGs). These CEGs are biologically meaningful and computationally significant local patterns, and are essential to infer higher-level functional machineries, e.g., regulatory and metabolic pathways. CEGs can be identified through biclustering, a data mining technique that allows simultaneous clustering of genes and conditions. A bud dormancy microarray time series with 2 genotypes, 2 photoperiod treatments and 6 replicates at each of 7 timepoints was used for biclustering analysis. A time series version of QUBIC was used to conduct biclustering analysis and identify CEGs that having a particular pattern. Specifically, we designed a model to represent expression patterns which maintains the tendency and level of expression change, but tolerate various time scale. For the identified correlated expression patterns, time shifting was allowed when comparing expressions patterns in same time course. The new method is capable of detecting expression correlation with time delay, thus can provide more precise co-expressed gene set for bud dormancy. Differential expression analysis was conducted on selected modules to identify DEGs that are specific to the dormancy responsive *V.riparia*.

W771: Plant Interactions with Pests and Pathogens
Wheat Hessian Fly Interactions: A Duel till Death
Subhashree Subramanyam, Purdue University, West Lafayette, IN, Jill A. Nemacheck, USDA ARS, West Lafayette, IN and Christie Williams, USDA-ARS at Purdue University, West Lafayette, IN
A duel between wheat (*Triticum aestivum*) and its major dipteran insect pest, the Hessian fly (*Mayetiola destructor*) elicits one of two interactions: incompatible (plant wins and larvae die), and compatible (larvae win and plant dies). During an incompatible interaction, the wheat plant surveillance mechanism detects the larval salivary effectors with an appropriate *R* gene triggering a gene-for-gene recognition that in turn activates plant defenses rendering the host plant resistant. This recognition event triggers expression of defense-response genes, changes in surface wax composition, accumulation of antifeedant proteins, as well as controlled host-cell permeability that aids in delivery of antinutrients, all leading to larval death. In contrast, during a compatible interaction, the salivary effectors from virulent larvae essentially hijack the host plant system by suppressing defense response, and altering the metabolic pathways leading to physiological changes at the feeding site (crown tissue) that provide the developing larvae a diet rich in essential nutrients making the wheat plant susceptible. Our recent investigations using Next generation RNA-Sequencing technology and quantitative real-time PCR expression studies in wheat and Hessian fly have revealed several differentially expressed genes and associated metabolic pathways providing molecular insight into plausible resistance and susceptibility mechanisms. Development of the Hessian fly *in planta* translocation (HIT) feeding assay opened up research avenues to test the effects of various defense and insecticidal proteins on this obligate parasite of wheat. In addition, we have explored the use of model grass genome and nonhost of Hessian fly, *Brachypodium distachyon*, for functional characterization of candidate defense genes for development of molecular tools to overcome economic devastations caused by this insect pest.

W772: Plant Interactions with Pests and Pathogens
Map Based Cloning of Hessian Fly Resistance Gene H13 in Wheat
Bikram S. Gill, Kansas State University, Manhattan, KS

W773: Plant Interactions with Pests and Pathogens
Turning Weakness into Strength: From Susceptibility Genes to Resistance in Solanaceous Crops
Anne-Marie A. Wolters[1], Kaile Sun[2], Annelies E.H.M. Loonen[1], Evert Jacobsen[1], Richard G.F. Visser[1] and Yuling Bai[1], (1)Plant Breeding, Wageningen University & Research, Wageningen, Netherlands, (2)Henan Agricultural University, Henan, China
Until recently, breeding efforts for resistance against pathogens in crops have mainly focussed on detection, mapping and cloning of single dominant genes. Most of these genes belong to the class of NB-LRR (Nucleotide-Binding, Leucine-Rich Repeats) genes. These genes usually confer resistance against specific strains or isolates of the pathogen, and are therefore narrow-spectrum resistance genes. Furthermore, this type of resistance can easily be broken by the emergence of new isolates. Because of this, the resistance conferred by NB-LRR type genes is often not durable. One way to tackle this problem is to stack several resistance genes against the same pathogen in one cultivar. Another way to achieve broad-spectrum durable resistance is to identify and silence or mutate so-called susceptibility genes (S-genes). Susceptibility genes can be defined as genes that encode proteins required for the pathogen to successfully invade and develop on the plant. When the function of such genes is impaired, the pathogen cannot grow on the plant and resistance is obtained.
A classic example of an S-gene is the Mildew Locus O (MLO) gene in barley. Mutant alleles of this gene and orthologous genes in other plant species confer durable resistance to powdery mildew. S-genes for several fungal, oomycete and bacterial pathogens have been identified and cloned from Arabidopsis. Several, but not all, of these genes show effects on plant development when mutated.
We are interested in the identification and analysis of S-genes for several pathogens in Solanaceous crops, such as tomato, potato and pepper. We have silenced or mutated orthologs of Arabidopsis S-genes, and obtained transgenic plants resistant against tomato powdery mildew, *Phytophthora infestans* and/or *Botrytis cinerea*. In my presentation I will elaborate on our results.

W774: Plant Interactions with Pests and Pathogens
Disease-Specific Genes Expressed in Compatible Plant-Pathogen Interactions: Examples from Tomato Leaves
Johannes Fahrentrapp and Fabio Rezzonico, Zurich University of Applied Sciences, Wädenswil, Switzerland
Abiotic and biotic disturbances induce gene expression level changes that can be measured and potentially be used for the detection of early stress symptoms *in planta*. If differentially regulated as response to a specific stress, these genes may be used as indicators allowing discrimination of its original inducer. For biotic stresses, examples of such specific indicators are resistance genes of resistant plant varieties

functioning as specific receptors of pathogen effectors. In compatible plant species disease-specific genes were identified for instance as so-called susceptibility genes – i.e., genes that are targets of pathogen effectors making the plant accessible for the pathogen. In our study we aimed at identifying differentially-regulated genes that are specifically induced by one pathogen but not by others. In disease-comparative studies in tomato leaves, we identified 68 genes specifically regulated by either *Botrytis cinerea* or *Phytophthora infestans* at 24 hours post inoculation. These genes are not differentially expressed in leaves infected with *Oidium neolycopersici*. Following single-drop infections, we further investigated the spatial expression patterns of *B. cinerea*-upregulated genes throughout healthy leaf tissue. At 24 hpi the expression of two out of four upregulated genes did not vary significantly between sampling and infection location on the same tomato compound leaf indicating a leaf-systemic pathogen induced regulation.

W775: Plant Interactions with Pests and Pathogens

Botrytis and Native Grape Yeasts – Not All Interactions are Created Equal

Patricia Okubara, USDA ARS, Pullman, WA

Native yeasts are of increasing interest to grape growers and winemakers in Washington State because of their potential to contribute to vineyard health and wine quality. In this pilot project, we used eleven strains of native yeasts and nine isolates of the Botrytis bunch rot pathogen, all obtained from Washington grapes. The yeasts included *Candida saitoana, Curvibasidium pallidicorallinum, Metschnikowia chrysoperlae, Metschnikowia pulcherrima, Meyerozyma guilliermondii*, *Wickerhamomyces anomalus* and *Aureobasidium pullulans*. Our objectives were to assess the biocontrol activities of each yeast, and to determine how the pathogen isolates responded to them. In laboratory assays, the native yeasts rapidly and extensively colonized artificial wounds on Thompson Seedless berries, showing about a 10,000-fold increase in population density per wound site within two days of introduction. Eight of the nine *Botrytis* isolates produced substantial symptoms on grape berries, generating disease severity ratings of 5.1-5.7 on a scale of 0 (healthy) to 7 (shrunken berry covered by fungus). However, each of the yeasts varied in their ability to suppress symptoms induced by four selected pathogen isolates when tested on grape berries. Our findings show that interactions on the berry depend on both yeast strain and *Botrytis* isolate. Biocontrol activity on the berry was not always reflected in inhibition assays on synthetic medium, suggesting that niche or nutrient competition was a likely biocontrol mechanism in planta. Finally, nutrient utilization differed among yeast species, indicating that co-deployment of more than one yeast might be a potential strategy for control of Botrytis bunch rot in the vineyard.

W776: Plant Interactions with Pests and Pathogens

Dissection of Gene Expression Networks Involved in Resistance to _Verticillium longisporum_ in _Brassica napus_

Harmeet Singh Chawla, Department of Plant Breeding, Justus Liebig University, GIESSEN, Germany, Christian Obermeier, Plant Breeding, Giessen, Germany, Andreas von Tiedemann, Department of Crop Sciences, Plant Pathology and Crop Protection Division, Georg August University, Goettingen, Germany and Rod Snowdon, Justus Liebig University, Giessen, Germany

Verticillium longisporum is a soil-borne hemibiotrophic fungal pathogen.It can survive in soil for several years by producing microsclerotia. These microsclerotia germinate in the soil and produce fungal hyphae, that can further penetrate into the epidermal cells of the roots, eventually entering the xylem elements of the host plant.It has been described that blocking of xylem vessels by fungal structures lead to obstruction in water and nutrient transportation, thereby causing chlorosis, leaf yellowing, premature ripening and senescence in oilseed rape. However, under field conditions, these symptoms become visible very late in the growing season hence making Verticillium an invisible threat to the production of oilseed rape in Europe. Furthermore, non-availability of fungicide based control make it even harder to control this pathogen under the natural conditions. Therefore, identification of host resistance is a desirable and sustainable approach to disease management. However molecular mechanism underlying Verticillium resistance is poorly understood. In order to dissect the gene expression networks involved in resistance to *Verticillium longisporum* in *Brassica napus,* we did RNA sequencing for two double haploid lines (from a bi-parental population) exhibiting a contrasting resistance reaction to Verticillium infection. By identifying the differentially expressed genes within QTL regions for Verticillium resistance, we were able to recognise some of the key candidate genes involved in the resistance reaction. An in-depth understanding of the resistance mechanism will be instrumental in the genetic improvement of the currently available oilseed rape cultivars.

W777: Plant Molecular Breeding

Characterization of a QTL and Candidate Gene (*LsGA2ox2*) Associated with Light Regulation of Lettuce Seed Germination

Heqiang Huo[1], Mohan Niroula[2], Theresa Hill[2], Juncheng Li[3], Linhchi Dinh Nguyen[3], Maria Jose Truco[4], Richard Michelmore[4] and Kent Bradford[5], (1)University of Florida, Apopka, FL, (2)Department of Plant Sciences, Seed Biotechnology Center, University of California, Davis, California, USA, Davis, CA, (3)Mid-Florida Research and Education Center, University of Florida,, Apopka, FL, (4)Genome Center, University of California, Davis, Davis, CA, (5)University of California, Davis, CA

Lettuce seed germination is promoted by light through activation of gibberellin (GA) biosynthesis. A quantitative trait locus for inhibition of lettuce seed germination by the darkness was mapped to chromosome 7 using a RIL population derived from a dark-germinating cultivar (*Lactuca sativa* cv. “Salinas”) and a light-requiring accession of *L. serriola* (US96UC23). The *LsGA2ox2* gene encoding a GA-inactivation enzyme was located in this QTL region and exhibited distinct expression patterns in Salinas and US96UC23 seeds imbibed in the dark. Nine of twelve SNPs between Salinas and US96UC23 (UC) were within a 93 bp window in the promoter of *LsGA2ox2*, and the same 9 SNPs were conserved in light-requiring Grand Rapids (*L. sativa*) and dark-germinating W48 (*L. serriola*). Expression of *ProUCGA2ox2::UCGA2ox2* in Salinas lines resulted in strong inhibition of seed germination in the dark. Site-directed mutagenesis of three SNPs associated with an abscisic acid-responsive motif in the *UCGA2ox2* promoter resulted in failure to complement *atga2ox2* mutants, suggesting that these SNPs are critical for upregulation of *LsGA2ox2* in seeds imbibed in the dark. Knockout of *LsGA2ox2* through CRISPR/Cas9 in Grand Rapids or RNAi silencing of *LsGA2ox2* in US96UC23 resulted in stem elongation and enhanced seed germination in the dark that are associated GA upregulation. The combination of genetic mapping, mutant complementation, and conserved SNPs implicates degradation of GA by *LsGA2ox2* in the inhibition of germination by darkness. Our results may facilitate breeding of lettuce varieties with more uniform germination of pelleted seeds and improve lettuce stand establishment and yield.

W778: Plant Molecular Breeding

The P450 Gene CYP749A16 Is Required for Tolerance to the Sulfonylurea Herbicide Envoke in Cotton (*Gossypium hirsutum* L.)

Gregory Thyssen, Cotton Chemistry and Utilization Unit, USDA-ARS-SRRC, New Orleans, LA, Marina Naoumkina, Cotton Fiber Biosciences Unit, USDA-ARS-SRRC, New Orleans, LA, Jack McCarty, Genetics & Sustainable Agriculture Research Unit, USDA-ARS, Mississippi State, MS, Johnie N Jenkins, USDA-ARS, Mississippi State, MS and David D. Fang, Cotton Fiber Biosciences Research Unit, USDA-ARS-SRRC, New Orleans, LA

Weed management is critical to global crop production and is complicated by rapidly evolving herbicide resistance in weeds. New sources of herbicide resistance are needed for crop plants so that applied herbicides can be rotated or combined to thwart the evolution of resistant weeds. The diverse family of cytochrome P450 proteins has been suggested to be a source of detoxifying herbicide metabolism in both weed and crop plants, and greater understanding of these genes will offer avenues for crop improvement and novel weed management practices. Here, we report the identification of CYP749A16 (Gh_D10G1401) which is responsible for the natural tolerance exhibited by most cotton, *Gossypium hirsutum* L., cultivars to the herbicide trifloxysulfuron sodium (CGA 362622, commercial name Envoke). The 1-bp frameshift insertion in the third exon of CYP749A16 results in the loss of cotton plant's resistance to Envoke herbicide. The DNA marker designed from this insertion perfectly co-segregated with the phenotype in 2145 F_2 progeny of a cross between the susceptible cultivar Paymaster HS26 and tolerant cultivar Stoneville 474, and in 550 recombinant inbred lines of a multi-parent advanced generation inter-cross population. Marker analysis of 382 additional cotton cultivars identified twelve cultivars containing the 1-bp frameshift insertion. A greenhouse experiment demonstrated a perfect match between marker genotypes and phenotypes in 188 plants from the selected twelve cultivars. Virus induced gene silencing of CYP749A16 generated sensitivity in the resistant cotton cultivar Stoneville 474. Taken together, we conclude that CYP749A16 is required for Envoke herbicide resistance in cotton.

W779: Plant Molecular Breeding

The Axiom Pear 70K Genotyping Array and Its Use in the Characterization of the USDA Pear Germplasm Collection

Sara Montanari[1], Luca Bianco[2], Michela Troggio[2], Brian J Allen[1], Nahla Bassil[3], Joseph Postman[3], Katherine M. Evans[4], A. Dhingra[5] and David B. Neale[6], (1)Department of Plant Science, University of California, Davis, Davis, CA, (2)Research and Innovation Centre, Edmund Mach Foundation, San Michele all'Adige, Italy, (3)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (4)Washington State University, Wenatchee, WA, (5)Molecular Plant Sciences, Washington State University, Pullman, WA, (6)Department of Plant Sciences, University of California, Davis, Davis, CA

A source of diversity and the development of genomic tools, such as reference genomes and molecular markers, are fundamental resources in plant breeding. In this project, we developed a high-density and high-efficiency single nucleotide polymorphism (SNP) array for pear, with the double objective of conducting genetic diversity and genome-wide association studies. The USDA National Clonal Germplasm Repository (NCGR) in Corvallis, OR, maintains a rich collection of more than 2,300 clonal pear accessions, including nearly every known *Pyrus* species. We re-sequenced 55 of these accessions and designed an Affymetrix Axiom Pear 700K Genotyping Array. By using this array to screen 288 highly diverse pear samples we were able to identify the most robust and informative SNPs to include in the commercial Axiom Pear 70K Genotyping Array, currently the densest SNP array available for pear. We further validated this array by genotyping the remaining NCGR pear accessions: more than 90% of the SNPs were classified as high quality and polymorphic (PolyHighResolution). This large dataset is being used to evaluate the genetic diversity of this pear germplasm collection and it will provide useful information for pear breeding.

W780: Plant Molecular Breeding

Functional Genomics of North American Soybean Breeding History

Yong-Qiang (Charles) An, USDA/ARS PGRU, Saint Louis, MO

W781: Plant Molecular Breeding

Development of New Restore CAPS Markers Linked to the Male Sterility in *Brassica napus*

Shinje Kim, Fungi and Plants Co.,Ltd, Chungcheongbuk-do, Korea, Republic of (South)

W782: Plant Molecular Breeding

Breeding Strategies Using Genomic Selection Increase Genetic Gain in Wheat Breeding Programs

Biructawit B. Tessema[1], Huiming Liu[1], Anders Christian Sørensen[1], Jeppe R. Andersen[2] and Just Jensen[1], (1)Molecular Biology and Genetics, Aarhus University, Tjele, Denmark, (2)Nordic Seed, Galten, Denmark

The traditional wheat breeding programs have been running for several years yet the genetic gain has been very limited. However, the use of genomic information for a selection criterion can increase genetic gain. This study was set to see how much genetic gain can be increased by implementing genomic selection on traditional wheat breeding program. In addition, we investigated the effect of genetic correlation between different traits on genetic gain. A series of wheat breeding programs that run simultaneously for 30 years was simulated using stochastic simulation, meaning each year a new breeding program starts with a cross of 60 parental lines followed by six generations of selfing. Selection was performed on three different generations. At F2, phenotypic selection was performed on breeder's visual preference. At F5 and F6, either phenotype or Genomic Estimated Breeding Value (GEBV) was used to select on yield. Yield at F5 and F6 was considered as different traits because they differ in plot size, population density, and number of plot replications. Plot heritability of these traits were 0.1, 0.2, and 0.3 while the economic values were 0, 0, and 1. In addition, we simulated different levels (0.3, 0.5, 0.7, 0.9) of genetic correlation between F2 and F5 as well as between F5 and F6. The varied selection criterion and varied genetic correlations make a total of 16 scenarios. GEBV as a selection

criterion significantly increased genetic gain by 10% compared to phenotype. Besides, the genetic gain was higher with the higher genetic correlation between traits.

W783: Plant Phenotypes
Quantifying Variation of Crop Resilience Under Temperature Stress
Malia Gehan, Donald Danforth Plant Sciences Center, 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 associated with the development of a hyperspectral and 3D imaging platform aimed at the identification of early abiotic stress response.

W784: Plant Phenotypes
Schapman Place Holder
Scott C. Chapman, CSIRO and The University of Queensland, St Lucia, Australia
replace with "real" abstract :-)

W785: Plant Phenotypes
Data Standards for Plant Phenotyping: MIAPPE and its Implementations
Cyril Pommier, URGI, INRA, Université Paris-Saclay, Versailles, France
The Minimal Information About Plant Phenotyping Experiment standard construction has been initiated four years ago. It has been constructed from the experience of expert from European infrastructures and institutes like Elixir, Emphasis, INRA, WUR, iBet, IPK, EBI and PAN. It is a checklist that formalize and document the minimal metadata necessary to ensure long term FAIRness (Findable, Accessible, Interoperable, Reusable) of field or greenhouse datasets, including high througputs.
This list has been implemented in several databases like GnpIS or eDale, in a file format, ISA Tab, in a web service, the Breeding API and an RDF implementation is under construction. We will review those implementations and detail the plans for the future evolutions of the standard.

W786: Plant Phenotypes
An Industry Perspective on High-Throughput Phenotyping
Sandra M. Dunckel, LongReach Plant Breeders, Adelaide, Australia
The fast development of new technologies and the integration of different fields of study, make it a very exciting time to work in agriculture. Drones, originally a military technology, have become increasingly popular in numerous fields. It is now possible to measure important traits such as plant height, biomass, and flowering time from the bird's eye view. The fast development of consumer based drones has opened the door to literally view agriculture from a different perspective. As with all new technologies, scientists are faced with significant challenges adapting them in their research programs. With many different platform, sensor and data processing pipeline options, the right choice is not always obvious. What kind of data are we after and how will it be incorporated in a commercial breeding program? Is a drone, ground-based phenomobile, or a combination of both the best choice? Which sensors are really needed and what is just nice to have? Join me on the hunt of an easy to setup, no-fuss high-throughput phenotyping system across the industry.

W787: Plant Phenotypes
NAPPN: Where are we Now and where are we going Next?
Argelia Lorence, Arkansas State University, Jonesboro, AR
The North American Plant Phenotyping Network (NAPPN) is an association of researchers, developers and consumers of plant phenotyping technologies across all organizational dimensions. The first meeting of the NAPPN occurred toward the end of 2016. Based on community input, the organization convened again as a General Assembly at PHENOME 2017 and an "ad hoc Board" was formed, formulated provisional bylaws, and is currently in the process of conducting the election of an Executive Board (intended to replace the ad hoc Board). In the coming year, the General Assembly will again convene at PHENOME. Potential next steps will involve seeking out support for research coordination, creating mechanisms for sharing resources such as equipment and facilities, and finding novel ways to enable trans-disciplinary research by organizing and publicizing opportunities for researchers across disciplines to learn from each other. To learn more about the NAPPN or to get involved, visit http://nappn.plant-phenotyping.org/. NAPPN relies on and encourages open communication and welcomes participation by individuals from diverse backgrounds, areas of study, and organization types. Members share involvement and interest in plant phenomics and are welcome from all ranks and levels of training, stature, and expertise.

W788: Plant Phenotypes
Jschnable Place Holder
James C Schnable, University of Nebraska-Lincoln, Lincoln, NE
replace with "real" abstract :-)

W789: Plant Reproductive Genomics
The Persimmon Genome Unveils Lineage-Specific Paleoduplication Events Driving Diversification of Sexual Systems

Takashi Akagi[1,2], Kenta Shirasawa[3], Hideki Nagasaki[3], Hideki Hirakawa[3], Ryutaro Tao[2], Luca Comai[4] and Isabelle M. Henry[4], (1)JST-PRESTO, Saitama, Japan, (2)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (3)Kazusa DNA Research Institute, Kisarazu, Japan, (4)Plant Biology and Genome Center, UC Davis, Davis, CA
Sexual polymorphism, a main strategy to maintain genetic diversity within a species, has long been a major focus in biology. In contrast to the situation in higher vertebrates, the evolution of sexual systems in plants likely occurred independently in multiple sexual lineages and only few of the mechanisms underlying these transitions have been unveiled. Here, we present a full draft of the genome sequence of dioecious persimmon (*Diospyros* spp.), a species nested within the Ericales order, and mined it for signs of genome evolution associated with persimmon-specific traits. Our analyses revealed a lineage-specific whole genome duplication event associated with strong positive selection on a small subset of the duplicated genes, which include *MeGI,* a key sex determinant in persimmon species. Evolutionary and physiological analyses indicated that *MeGI* underwent neofunctionalization via positive selection after duplication from its sister gene, *Sister of MeGI* (*SiMeGI*), and acquired a new function as a repressor of male organ development. These findings exemplify how whole genome duplication events can contribute flexible genetic material whose variation can be selected for development of new sexual systems.

W790: Plant Reproductive Genomics
Sex Chromosome Evolution in Liverworts.
Peter Szovenyi, University of Zurich, Zurich, Switzerland; University of Zurich, Institute of Systematic Botany, Zurich, Switzerland
In the majority of organism, sex is determined by the segregation of a special set of chromosomes, the sex chromosomes. During the last 10-20 years a wealth of information has been gained on the structure and organization of sex chromosomes, nevertheless, much remains to be learned about their origin and evolution. Evolution, origin and structure of sex chromosomes is even less understood in systems in which sex is expressed in the dominant haploid phase. We use liverworts, a monophyletic group of bryophytes with a haploid-dominant life cycle to study the origin and evolution of sex chromosomes in haploid dioecy. Liverwort species are primarily dioecious, and it is hypothesized that dioecy is ancestral to the group and sex chromosomes have originated several hundred million years ago. Furthermore, in contrast to diploid-dominant systems, sex chromosomes under haploid dioecy are expected to be influenced by similar evolutionary forces and should follow similar evolutionary trajectories. Nevertheless, size and number of sex chromosomes varies tremendously among liverwort species with some chromosomes being larger while others smaller than their autosomal complements. Using a combination of single-molecule and short-read sequencing technologies we show that the hypothesis assuming similar evolutionary trajectories for both sex chromosomes is premature. Furthermore, we present evidence that composition of liverwort sex chromosomes is a more dynamic than expected and that the hypothesis concerning the evolutionary trajectories of the sex chromosomes has to be reconsidered.

W791: Plant Reproductive Genomics
MicroRNA Control of Flowering Regulation in Tropical/Subtropical Tree Crops
Muhammad Umair Ahsan[1], Alice Hayward[1], Christine Beveridge[2] and Neena Mitter[1], (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia, (2)School of Biological Sciences Faculty of Science, The University of Queensland, St Lucia,, Brisbane, Australia
The horticultural tree life cycle can be differentiated into two distinct growth stages - vegetative and reproductive. Unlike annual crops, horticultural tree crops enter into an annual flowering cycle once they attain reproductive phase. Herein, they undergo various developmental changes including flowering, fruit set, fruit development, fruit ripening, fruit drop and vegetative growth. The annual crop cycle is critical for horticultural crops like Persea americana (avocado) and Macadamia integrifolia (macadamia) as their productivity depends upon successful flowering and fruit development. It is therefore important to understand flowering and the factors involved in the crop cycle in these horticultural trees. Various endogenous and exogenous factors are involved in flowering regulation. Among them a class of small non-coding RNAs known as microRNAs play a vital role. We investigated the expression of flowering-associated miRNAs in avocado and macadamia at distinctive time-points during their annual crop cycle. We looked at the expression of two miRNA that are known act antagonistically during flowering; the floral inhibitor miR156, and the floral promoter miR172, which regulates APETALA2 (AP2)-like transcription factors. Preliminary results from avocado showed no significant difference in the expression of miR156 during the crop cycle, while significant differences were observed for miR172. Abundance of miR172 was higher during flowering and initial fruiting but lower at fruit ripening and drop/harvest. These results will help us in understanding the complex crop cycle and flowering regulation in horticultural trees.

W792: Plant Reproductive Genomics
Genetic Analysis of Flowering Time Regulation in Sugar Beet and Related Species of the Family Offering New Perspectives for Breeding
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
Sugar beet is a biennial species where floral transition is initiated after exposure to extended cold. It belongs to the Amaranthaceae plant family together with spinach, Quinoa and Amaranth. Key regulators for bolting time control have been cloned. Two homologs of the Arabidopsis *FT* gene have contrasting effects on flowering time in beet. *BvFT1* is a floral repressor whereas *BvFT2* induces bolting and flowering. *BTC1* and *BvBBX19* are both acting upstream of the *BvFT's* perceiving signals form the photoperiod pathway. Single nucleotide mutations in either *BTC1* or *BvBBX19* have a strong impact on the function of the encoded protein turning an annual into a biennial genotype. Our recent activities are focused on the interaction between BTC1 and BvBBX19 and their transcriptional regulation of both *BvFT* genes. We reason that the *BTC1* / *BvBBX19* module established a function like the *CO* gene from Arabidopsis because a functional *CO* ortholog is missing in beet. We selected double mutants (*btc1* and *Bvbbx19*) from segregating beet offspring which displayed strong non-bolting phenotypes corroborating an epistatic gene model where both proteins jointly control their downstream targets. These studies path the way for breeding vegetative crops in which orthologs of the floral activator *FT* are completely downregulated even after cold treatment giving rise to 'never bolting' crops. Putative

applications are winter beets sown before winter or vegetables like lettuce or Chinese cabbage which have a strong tendency for early bolting if they experience cold temperatures in early spring cultivation.

W793: Plant Reproductive Genomics
The Role of Ovate Family Proteins in Regulating Shape of Tomato Fruit
Esther Vanderknaap, University of Georgia, Athens, GA
The final shape and size of fruits result from coordinated cell proliferation and expansion along different axes of growth. The shape of many elongated and pear-shaped tomato varieties is controlled by a naturally occurring premature stop mutation in *OVATE*, a member of the Ovate Family Proteins (OFPs) class. Histological analyses demonstrated that the mutation results in elongated shape associated with an altered cell division pattern. Mapping of the suppressors of the ovate (*sov*) led to the identification of another member of the family, *SlOFP20*, as the candidate gene underlying *sov1*. A synergistic interaction was found between *ovate* and *sov1* in controlling fruit elongation, which suggests that OVATE and SlOFP20 are in the same developmental pathway. Yeast 2 Hybrid experiments showed that OVATE and SlOFP20 interact with members of a protein complex that regulates the formation of preprophase band and organization of cortical microtubule array. Transient co-expression in N. benthamiana resulted in relocalization of OVATE and SlOFP20. Our findings are starting to shed light on the role of OFPs in proximal-distal patterning of fruit and provide insights into fundamental aspects of plant organ growth. This project is supported by the Agriculture and Food Research Initiative competitive grant 2017-67013-26199 of the USDA National Institute of Food and Agriculture.

W794: Plant Reproductive Genomics
Genetics and Developmental Evolution of Fruit Morphological Variation in Physalis
Chaoying He, Institute of Botany, The Chinese Academy of Sciences, Beijing, China
Morphological variations of fruits, such as shape and size, are a result of adaptive evolution, and play an essential role in facilitating seed dispersal. Most Solanaceous species produce berry, which size and color show a dramatic diversity in *Physalis* that has distinguished fruit morphology with a papery husk as the accessory trait of fruits. The distinct trait of *Physalis* species is termed Chinese lantern or inflated calyx syndrome since it is a derivative of the calyx. It is a post-floral morphological novelty. The tremendous trait diversity should largely result from genome diversity. Thus, we presumed that certain genomic variations might have occurred in *Physalis* relative to *Solanum* and be responsible for the origin of the lantern trait. Here, I will report the main progress on the evolutionary developmental genetics of berry size variation and Chinese lantern in *Physalis*. We revealed the role of MADS-box genes *MPF2* and *MPF3* in the origin of Chinese lantern, and characterized two genes *POS1* and *POS2* accounting for berry size control. The genetic analyses of Physalis fruit development and evolution are going on at both genomic and transcriptomic levels. Our studies reveal that the genome diversity underlies morphological variation, and particularly suggest that the recruitment of a pre-existing gene and subsequent modification of its interaction and regulatory networks are frequently involved in the evolution of morphological diversity. Our work also provides insights into how to improve the productivity of crops. Moreover, *Physalis* has been established as an emerging model plant for development, evolution and ecology.

W795: Plant Transgene Genetics
Regulatory Elements Diversity and Coding Sequence Composition Impacts on Transgene Expression Stability
Lyudmila Sidorenko[1], Tzuu-fen Lee[2], Aaron Woosley[1], William Moskal[1], Scott Bevan[1], P. Ann Owens Merlo[1], Terence Walsh[1], Xiujuan Wang[1], Staci Weaver[1], Todd Glancy[1], PoHao Wang[1], Xiaozeng Yang[1], Shreedharan Sriram[1] and Blake Meyers[2], (1)DOW AgroSciences LLC, Indianapolis, IN, (2)Donald Danforth Plant Science Center, St. Louis, MO
Transgene silencing is a significant challenge for plant genetic engineering as it leads to phenotypic variability and eventual loss of transgene expression. Improved transgene design could help to achieve the desired levels of transgene expression and to ensure that expression is stable over multiple generations. Our studies evaluated several transgene design parameters as means to improve transgene expression. Results of these studies showed that diversification of regulatory elements and increasing %GC of coding sequences improve performance of diverse transgenes in *Arabidopsis* and maize. Implications of these results for improving reliability of transgene expression strategies will be discussed.

W796: Plant Transgene Genetics
Plant Synthetic Promoters and Transcription Factors
Wusheng Liu, North Carolina State University, Raleigh, NC

W797: Plant Transgene Genetics
Development of Reduced-Gluten Wheat Genotypes for Gluten-Insensitive Individuals using Transgenic Approach
Sachin Rustgi, Clemson University, Florence, SC

W798: Plant Transgene Genetics
Development of "Purple Endosperm Rice" by Engineering Anthocyanin Biosynthesis in the Endosperm with a High-Efficiency Transgene Stacking System
Qinlong Zhu, South China Agricultural University, Guangzhou, China

W799: Plant Transgene Genetics
Dissecting the Mechanism of Salicylic Acid-Mediated Defense and Growth Tradeoff in Arabidopsis
Maria A. Ortega, Brian Kvitko, Scott A. Harding and Chung-Jui Tsai, University of Georgia, Athens, GA
The phytohormone salicylic acid (SA) regulates plant growth, development, and defense responses to abiotic and biotic stresses. Mutant *Arabidopsis* lines that constitutively overproduce SA show enhanced tolerance to abiotic and biotic stresses, however their growth is compromised as a tradeoff. Studying SA inhibition on plant growth has been challenging, in part because many SA-elevated *Arabidopsis*

mutants are perturbed in immune signaling pathways due to pleiotropic effects. We developed a set of transgenic *Arabidopsis* that hyperaccumulates SA by expressing the bifunctional *Yersinia enterocolitica* SA synthase gene with a plastidic presequence (*Fd-Irp9*). This novel set of high-SA transgenic lines facilitates dissecting the effect of SA on growth and defense tradeoffs without other confounding factors. The high-SA transgenic lines show enhanced resistance to *Pseudomonas syringae* pv. Tomato (Pst) strain DC3000, and tolerance to salinity and osmotic stresses. We also found that transgenic plant growth was reduced in an SA-dependent manner and it was exacerbated at cool temperatures. RNA-Seq analysis revealed an inhibitory effect of SA on a group of cold-regulated (*COR*) genes. *COR* genes are induced during cold acclimation, and they are associated with freezing tolerance, presumably by stabilizing cell membranes and/or proteins. We hypothesized that high-SA plants are unable to mount a protective *COR* response to cool temperature due to SA suppression of *COR* expression, leading to membrane damage and reduced growth. We selected two *COR* genes for over-expression in the high-SA lines to bypass SA suppression. Constitutive expression of *COR* genes rescued the growth phenotype in a high-SA line without affecting disease resistance. The set of high-SA, *COR*-expressing transgenic *Arabidopsis* are useful to elucidate the molecular basis of the SA-*COR* antagonism, which would led to the development of strategies to genetically decouple *COR*-mediated growth protection from SA-mediated defense for crop improvement.

W800: Plant Transgene Genetics
A CRISPR–Cpf1 System for Efficient Genome Editing and Transcriptional Repression in Plants
Yiping Qi, University of Maryland, College Park, MD

W801: Polyploidy
Deciphering the Allotetraploid Genome of *Coffea arabica* L.
Alexandre de Kochko, IRD UMR DIADE, Montpellier, France and Arabica Coffee Genome Consortium
Coffea arabica, one of the two coffee cultivated species, is the sole tetraploid species of the *Coffea* genus, which contains to date 125 accepted species. *C. arabica* resulted from a recent and spontaneous cross between *C. eugenioides*, a wild species growing in East Africa and *C. canephora*, the other cultivated species. The two parental genomes are highly homeologous, up to 98% in their conserved regions, which constitutes a difficult challenge for reconstructing the two parental genomes anchored on 22 pseudomolecules. The international Arabica Coffee Genome Consortium (ACGC) has undertaken the sequencing of this genome as well as that of the two parental species.
The most advanced technologies were used to reach this goal: long and short read sequencing (genomic and transcriptomic), optical mapping and chromosomal conformation capture (Hi-C). Finally the results were anchored to genetic maps.
Genome assemblies indicate that, as expected, no important chromosomal rearrangements intervene between the two subgenomes, which present a high colinearity. A peculiar situation characterizes the *Coffea* genomes, indeed, while centromeric retrotransposons are present on all chromosome pairs but one, no centromeric repeats (satellite sequences) have been identified.
A comparative analysis, based on the gene annotation of the three genomes, indicates minor differences between them, at the opposite, a fast and efficient mechanism of transposable elements control and elimination took place in *C. arabica*. The uniqueness of the event leading to the *C. arabica* species formation will be also discussed.

W802: Polyploidy
Ploidy Dynamics in an Opportunistic Fungal Pathogen
Meleah Hickman, Emory University, Atlanta, GA

W803: Polyploidy
Regulatory Changes Underlying Differential Expression in High-Yielding Triploid Willow (*Salix spp.*) Bioenergy Crops
Craig H. Carlson, Cornell University, Geneva, NY
Many studies have highlighted the complex, multigenic basis for heterosis (hybrid vigor) in inbred crops. Despite the lack a consensus model, it is vital that we turn our attention to understanding heterosis in undomesticated, outcrossing, heterozygous, and often polyploid species, such as willow (*Salix*). Shrub willow is a dedicated energy crop and is bred to be fast-growing and high-yielding on marginal land without competing with food crops. A trend in willow breeding is the consistent pattern of heterosis in triploid progeny produced from crosses between diploid and tetraploid species. Critical in understanding heterosis, the heritability of gene expression is dependent on allele-specific expression by local and remote factors in the genome. Here, we test whether differentially expressed genes are responsible for heterosis in triploid crosses made between diploid *S. purpurea*, diploid *S. viminalis*, and tetraploid *S. miyabeana* parents. Three biological replicates of progeny and parent shoot tips were collected after 11 weeks in the greenhouse and individually sequenced via RNA-Seq (2×101). Our results highlight regulatory factors influencing differential expression and top modules of co-expressed genes correlated with heterosis for phenotypes collected in the greenhouse and in the field. We have previously shown in diploid F_1 and F_2 *S. purpurea* that expression-level dominance and sex dimorphic expression is pervasive; hence, the effects of ploidy, sex, and pedigree in triploids will be discussed. Altogether, these data will be used to develop predictive models of heterosis and complement the growing genomic resources available for the improvement of shrub willow bioenergy crops.

W804: Polyploidy
Unstable Allotetraploid Tobacco Genome Due to Frequent Homeologous Recombination, Segmental Deletion and Chromosome Loss
Shumin Chen, Feihong Ren and **Hanhui Kuang**, Huazhong Agricultural University, Wuhan, China
How genes in polyploid have evolved remains poorly understood. In this study, a high throughput method was developed to identify spontaneous loss-of-function alleles for the resistance gene *N*, which provides resistance against the tobacco mosaic virus (TMV) in allotetraploid tobacco. A total of 2,134 loss-of-function alleles of the *N* gene were identified after screening 14 million F_1 hybrids. Analysis of these mutants showed striking contrast of evolutionary patterns for genes in polyploidy compared with diploid. Only 14 of these 2,134 mutants were caused by spontaneous point mutations or indels, while the others were caused by homeologous recombination (with a frequency of ~1/12,000) or loss of chromosome (~1/15,000). A similar frequency (~1/13,000) of chromosome loss was found for another chromosome,

which also had a frequency of ~1/16,000 for spontaneous segmental deletion. Both homeologous recombination and chromosome loss considerably decreased the viability of the mutants. Our data suggested that mutation rate in polyploids is dramatically higher than diploid, mainly attributed to homeologous recombination and tolerance of chromosome loss. Frequent mutations tend to drive polyploids to extinction unless a novel mutation helps the polyploid to effectively compete with diploids or find a new ecological niche.

W805: Polyploidy

New Traits in *Spartina* Species: A Consequence of Hybridizations and/or Polyploidy?

Armel Salmon[1], Armand Cavé-Radet[1], Delphine Giraud[1], Julien Boutte[1], Hélène Rousseau[1], Oscar Lima[2], Mathieu Rousseau-Gueutin[3], Abdelhak El-Amrani[1] and Malika Lily Ainouche[1], (1)University of Rennes 1 - UMR CNRS 6553 ECOBIO, Rennes, France, (2)UMR CNRS 6553 ECOBIO - University of Rennes 1, Rennes, France, (3)INRA, Le Rheu, France

Hybridization and genome duplication (polyploidy) are particularly well-illustrated in genus *Spartina*, resulting in complex genomes which exhibit several duplication events, including those dating back to the Poaceae family. The most complex genome (namely the invasive allo-dodecaploid *Spartina anglica*) arose recently in Europe *c.a*. 150 years ago, by genome duplication of the homoploid hybrid *S. x townsendii* resulting from an interspecifc cross between the introduced *S. alterniflora* (2n=6x=62) as maternal parent and the European native *S. maritima* (2n=6x=60). This well described system represents one of the rare known examples of recent allopolyploid speciation where the actual parents and F1 parents may be compared to the new allododecaploid species in natural populations. *Spartina* species play an important ecological role in the sedimentary dynamics of coastal saltmarshes in several continents and some of them evolved new traits, such as the ability to produce DMSP (DiMethylSulfionoPropionate) or increased tolerance to organic compounds such as hydrocarbons or to heavy metals.

In the perspective of understanding the role of reticulate evolution and whole genome duplications in the phenotypic plasticity and adaptation of the recently formed *Spartina anglica*, the high ploidy levels of this (dodecaploid) species and its (hexaploid) parents makes genetic and genomic studies particularly challenging. We are using massive parallel sequencing technologies to explore the genomes and transcriptomes of these species, and we have developed bioinformatic approaches and tools for detecting the different putative orthologous copies originating from the parents (duplicated homoeologs) in *S. anglica*.

W806: Polyploidy

A Mechanism for Genome Size Reduction Following Genomic Rearrangements

Longhui Ren[1], Wei Huang[1], Ethalinda Cannon[1], David Bertioli[2] and Steven Cannon[3], (1)Iowa State University, Ames, IA, (2)University of Georgia, Athens, GA, (3)USDA–Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA

The factors behind genome size evolution have been of great interest, considering that eukaryotic genomes vary in size by more than three orders of magnitude. Using a model of two wild peanut relatives, *Arachis duranensis* and *Arachis ipaensis*, in which one genome experienced large rearrangements, we find that the main determinant in genome size reduction is a set of inversions that occurred in *A. duranensis*, and subsequent net sequence removal in the inverted regions. We observe a general pattern in which sequence is lost more rapidly at newly distal (telomeric) regions than it is gained at newly proximal (pericentromeric) regions – resulting in net sequence loss in the inverted regions. The major driver of this process is recombination, determined by the chromosomal location. Any type of genomic rearrangement that exposes proximal regions to higher recombination rates can cause genome size reduction by this mechanism. In comparisons between *A. duranensis* and *A. ipaensis*, we find that the inversions all occurred in *A. duranensis*. Sequence loss in those regions was primarily due to removal of transposable elements. Illegitimate recombination is likely the major mechanism responsible for the sequence removal, rather than unequal intrastrand recombination. We also measure the relative rate of genome size reduction in these two *Arachis* diploids. We also test our model in other plant species and find that it applies in all cases examined, suggesting our model is widely applicable.

W807: Polyploidy

Dynamics of Duplicated Networks in Polyploids

Guanjing Hu, Iowa State University, Ames, IA

Polyploidy is a widespread phenomenon throughout eukaryotes. The duplication of genetic materials immediately reshapes the architecture of biological networks, and sets in motion the evolutionary changes of network components and their interacting relationships, which ultimately affects how molecules generate phenotypes with important ecologicaland evolutionary consequences. Although co-expression network analysis has been fruitful to underlie the genotype to phenotype equation for many systems, the analysis in polyploid species poses unique challenges due to the technical difficulty in estimating duplicated gene expression levels. To address these challenges and pitfalls in the use of RNA-seq data for duplicated gene expression and co-expression network analyses, we exemplify the analytic workflow and define reasonable practice for revealing the dynamics of duplicated gene networks in allopolyploid cotton. By examining network topological changes within and between sub-genomes by allopolyploidization, a significant increase of inter-connection between sub-genomes was observed. While no global bias was found within each sub-genome in terms of gain or loss of network connections, asymmetrical sub-network topology was identified implicating the functional and regulatory divergence of duplicated genes. These results demonstrate the potential for duplicated network analysis for understanding the regulatory architecture of complex genomes and phenotypes.

W808: Polyploidy

From the Freezer to the World: Genomics of Allopolyploidy-Facilitated Niche Expansion and Adaptation in White Clover.

Andrew G. Griffiths[1], Roger Moraga[1], Stig Uggerhøj Andersen[2], Vikas Gupta[2], Timothy Bilton[3], Marni Tausen[4], Matthew A. Campbell[5], Rachael Ashby[3], Istvan Nagy[6], Anar Khan[7], Craig Anderson[1], Benjamin K. Franzmayr[1], Kerry Hancock[1], Alicia Scott[1], Nick W. Ellison[1], Murray P. Cox[8], Torben Asp[9], Thomas Mailund[10] and Mikkel Heide Schierup[2], (1)AgResearch, Palmerston North, New Zealand, (2)Aarhus University, Aarhus, Denmark, (3)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (4)Department of Molecular Biology and Genetics, Aarhus, Denmark, (5)Department of Ecology and Evolutionary Biology, Santa Cruz, CA, (6)Aarhus University, Slagelse, Denmark, (7)AgResearch, Mosgiel, New Zealand,

(8)Massey University, Palmerston North, New Zealand, (9)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, (10)Bioinformatics Research Centre, Aarhus, Denmark
The merging of distinct genomes, 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=16) forage crop 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. We have confirmed the progenitors, and shown that the progenitor subgenomes within white clover have largely retained their integrity and gene expression activity following allopolyploidisation. Furthermore, we show that this hybridisation event occurred during the depths of the last glaciation at a time when the European progenitor ranges (coastal and montane) likely overlapped.
Born of climate change, white clover represents a clear example of allopolyploidy-facilitated niche expansion, where the two progenitor genomes reunited and expanded from disparate and highly specialised 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 homoeologous copies of genes involved in flavonoid biosynthesis, a key pathway involved in adaptive traits such as plant/microbial interactions.

W809: Polyploidy
Subgenome Analyses of Polyploid Blueberry
Marivi Colle, Department of Horticulture, Michigan State University, East Lansing, MI and Patrick Edger, Michigan State University, EAST LANSING, MI
Genomic analyses of blueberry (Ericaceae) has been greatly hampered due to the lack of adequate genomic resources, largely due to the complexity of the tetraploid genome structure. Here we present a haplotype-phased, chromosome-scale assembly of the tetraploid northern highbush blueberry (2n = 4x = 48), which will facilitate the discovery and analysis of genes encoding economically important traits. Our near complete reference genome consists of 48 pseudomolecules with ~1.92Gb of assembled sequence, only ~1.29% gaps, and a total of 128,559 protein coding genes. We will present results from analyses investigating differences among subgenomes in blueberry.

W810: Polyploidy
Revisiting Pivotal-Differential Genome Relationships
Ghader Mirzaghaderi, Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran (Islamic Republic of) and **Annaliese S Mason**, Justus Liebig University, Giessen, Germany
Pivotal-differential genome relationships in the wheat (*Triticum* and *Aegilops* genera) group refer to the pattern whereby one genome in an allopolyploid is relatively conserved (pivotal) between diploid and allopolyploid wheat group species, whereas the other genome(s) is altered (differential) relative to other diploid and allopolyploid species containing this genome. This pattern was first identified based on comparative morphology (flowering spikes), whereby species could be grouped into A, D and U genome clusters. However, substantial cytogenetic evidence also supports this genome relationship, and with recent genomic advancements in wheat we suggest that it is time to interrogate this relationship further, and to extend these concepts to other plant taxa where it may be relevant. In particular, we propose that pivotal-differential genome patterning within taxa may have three possible explanations that should be tested. Firstly, variation between species sharing a differential genome may be directly inherited from variation present within the ancestral diploid species (e.g. different progenitor cytotypes or subspecies). Secondly, variation between species may be induced as a result of the allopolyploid formation event, perhaps as a result of dominance relationships between subgenomes. Thirdly, hybridization between two allopolyploid species that share a (pivotal) genome in common but differ in their second genome may give rise to a new, rearranged (differential) genome after hybridization and genome stabilization (e.g. AABB x AACC -> AADD). Interrogation of future pan-genome data coupled with synthetic recreation of historical hybridization events is predicted to reveal the mechanisms underlying pivotal-differential genome patterns.

W811: Polyploidy
Structural Equation Models Based on Multivariate Diversity Assessment of Diploid and Tetraploid Hulled Wheat Species
Abdullah A. Jaradat[1], Linda Dykes[2] and Steven S. Xu[2], (1)USDA-ARS, Morris, MN, (2)USDA-ARS, Fargo, ND
Hulled wheats are largely untapped genetic resources with >10,000 years of genetic memory and diversity that can be used for wheat quality improvement, development of healthy products, and adaptation to climate change. Multivariate diversity was assessed in the diploid *Triticum monococcum* L. var *monococcum* (AA genome) and tetraploid *Triticum turgidum* L. var. *dicoccum* (BBAA genome) populations representing landraces and improved cultivars from the primary and secondary centers of diversity of each species. A relational database (273 traits/repeated measurements classified into architectural, agronomic, eco-physiological, phytochemical, ionome, and rheological and textural interrelated modules, was constructed using phenotypic and high throughput data of kernels, flour, dough, bread, seedlings, and mature plants. Most (65%) of twenty principal components derived from these traits discriminated between, and explained larger variance in the BBAA (67.0%) than in the AA (43.7%) genome. Increased ploidy did not cause significant changes in 27% of all traits; while most architectural, agronomic and reproductive traits, except plant height and tillering, increased in magnitude with increased ploidy level. Both ploidy levels had comparable biochemical and ionome profiles; however, polyploidy resulted in smaller total phenolics, β-glucan, yellow pigment index, carotenoids, and micronutrient (Fe and Zn) content; while it resulted in better tolerance to salinity at the seedling stage, larger kernel weight, grain and protein yield, gluten index, ash content; and slightly denser and larger loaf volume. We used structural equation modeling to hypothesize a 'quality' latent variable and identified causal relationships, tradeoffs, intra- and interrelationships due to direct and indirect, positive or negative correlations between traits within and among the six modules at each ploidy level. A minimum set of species-specific traits was identified and their effects on the 'quality' of each species was quantified using a prediction profiling procedure. Adjustments in these traits will be used to optimize 'quality' and to design an 'ideotype' at each ploidy level. The procedure can be adapted to design hulled wheat species *in silico* with enhanced resource use efficiency and yield potential; higher tolerance to abiotic and biotic stresses; and improved nutritional profiles.

W812: Population and Conservation Genomics 1

Parallel Epigenetic Modifications Induced by Hatchery Rearing in a Pacific Salmon

Louis Bernatchez[1], Jérémy Le Luyer[2], Martin Laporte[3], Terry Beacham[4], Karia Kaukinen[5], Ruth E. Withler[6], Jong S. Leong[7], Eric Rondeau[8] and Ben F. Koop[7], (1)Université Laval, Québec, QC, Canada, (2)Université Laval, Canada, (3)University Laval, Quebec, QC, Canada, (4)Molecular Genetics Lab, Pacific Biological Station, Nanaimo, BC, Canada, (5)3Department of Fisheries and Ocean, Pacific Biological Station, Canada, (6)Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada, (7)University of Victoria, Victoria, BC, Canada, (8)University of Victoria, Canada

Wild stocks of Pacific salmonids have experienced sharp declines in abundance over the past century. Consequently, billions of fish are released each year for enhancing abundance and sustain fisheries. Yet the beneficial role of this widely used management practice is highly debated since decreased of fitness of hatchery-origin fish in the wild has been documented. Artificial selection in hatcheries has often been invoked as the most likely explanation for reduced fitness, and most studies to date focused on finding signatures of hatchery-induced selection at the DNA level. We tested an alternative hypothesis that captive rearing induces epigenetic reprogramming by comparing genome-wide patterns of methylation and variation at the DNA level in hatchery-reared Coho Salmon (*Oncorhynchus kisutch*) with their wild counterparts in two geographically distant rivers. We found a highly significant proportion of epigenetic variation explained by the rearing environment that was as high as the one explained by the river of origin. Of 100 Differentially Methylated Regions (DMRs) between hatchery and wild origin salmon, 89 were hypermethylated in hatchery fish. The differentially methylated regions show enrichment for biological functions that may affect the capacity of hatchery-born smolts to migrate successfully in the ocean. Shared epigenetic variation between hatchery-reared salmon provides evidence for parallel epigenetic modifications induced by hatchery rearing in absence of genetic differentiation between hatchery and natural-origin fish for each river. This is the first study to highlight epigenetic modifications induced by captive rearing as a potential explanatory mechanism for reduced fitness in hatchery reared salmon.

W813: Population and Conservation Genomics 1

BEAN ADAPT: The Genomics of Adaptation during Crop Expansion of Common Bean

Roberto Papa, Università Politecnica delle Marche, Ancona, Italy, Alisdair R. Fernie, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, Paul Gepts, University of California, Davis, CA, Andreas Graner, Leibniz Institute of Plant Genetica and Crop Plant Reserch, Gatersleben, Germany and Scott A. Jackson, University of Georgia, Athens, GA

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 8X coverage per accessions. We identified ~1.5 million of SNPs with less than 5% of missing data and analysed them to characterize genome-wide variation. Moreover, 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.

W814: Population and Conservation Genomics 1

Exploring the Mechanisms Maintaining Species Integrity in Hybridizing Eurasian *Populus* spp: Coupling Population Genomics with Experiments

Christian Lexer, University of Vienna, Vienna, Austria

Current developments in evolutionary and population genomics hold great promise for research on reproductive isolation and speciation, including the mechanisms maintaining species integrity when previously diverged taxa are challenged by hybridization. Population genomic studies of the ´model forest tree´ genus *Populus* have already contributed greatly to our current understanding of these topics. In this talk, I will present recently published and ongoing research by our group and collaborators to address the mechanisms maintaining reproductive isolation and trait differences in a complex of Eurasian poplars and aspens, including the wide-spread, ecologically divergent, hybridizing taxa *Populus alba* (White poplar) and *P. tremula* (European aspen). By coupling genomics with experiments, we are beginning to understand the nature of the genomic interactions responsible for both, fitness differences among genotypes in common garden trials and patterns of genomic diversity in natural populations. I will present results from recent whole genome sequencing (WGS) and reduced complexity library sequencing (RAD and GBS) studies of these species, their hybrid zones, and of common garden trials derived from them. Both model-free and model-based analytical approaches reveal the genomic footprint of a complex joint demographic history with recurrent gene flow episodes. Nevertheless, species integrity appears to be maintained by strong post-mating selection episodes, broadly consistent with the 'genomic coadaptation' model of barriers to gene flow in secondary contact. I will discuss the potential roles of this "genomic clash" and of heterosis in facilitating both, F1 hybrid persistence and the episodic breakdown of barriers between two species that apparently diverged for millions of years.

W815: Population and Conservation Genomics 1

Genome-Wide Evidence for Subspecies Recognition and Adaptive Divergence in the Tiger (*Panthera tigris*)

Yue-Chen Liu, Peking University, Beijing, China

Lack of consensus over the number of subspecies or conservation management unit in the tiger (*Panthera tigris*) has partially hindered the global effort to recover it from the brink of extinction. Today fewer than 4,000 free-ranging tigers survive in only 7% of their historical range and debates persist whether they should be considered six or two subspecies. The recent coalescence of all modern tigers to a Late Pleistocene

bottleneck poses challenges to detecting subspecies-diagnostic morphological traits in the tiger and the ultimate resolution to elucidating its intraspecific evolution and taxonomy lies at the genome scale. Here based on whole genome sequencing of 32 voucher specimens we present the first genome-wide evidence that supports five statistically robust monophyletic clusters corresponding to extant subspecies, which are Sumatran (*P. t. sumatrae*), Amur (*P. t. altaica*), Indochinese (*P. t. corbetti*), Bengal (*P. t. tigris*), and Malayan (*P. t. jacksoni*) tigers, as well as one unique but tentative lineage of South China tiger (*P. t. amoyensis*) due to limited sampling. Demographic reconstructions validated a severe population decline across the whole species around 110 kya followed by serial divergent events driven by global climate fluctuations. Overall, inter-subspecies gene flow is low corroborating these recently isolated but distinct phylogeographic units. In addition, we identified multiple genomic regions that may have formed the basis of adaptive evolution in various subspecies. These genome-wide signatures provide an explicit basis for subspecies recognition in the tiger and will help facilitate global conservation strategic planning for this charismatic flagship species.

W816: Population and Conservation Genomics 1

Extrinsic Forcing of Genomic Evolution during Speciation: A Geo-Genomic Study of Gopherus Desert Tortoises

Greer A. Dolby, Timothy H. Webster, Dale F. DeNardo, Melissa A. Wilson Sayres and Kenro Kusumi, Arizona State University, Tempe, AZ

There are often multiple extrinsic forces shaping patterns of genomic evolution among populations or species, and these forces operate on varying timescales. To understand how threatened *Gopherus* tortoises have evolved in response to the changing environments of the southwestern US requires accounting for the suite of extrinsic (geo-climatic) forces to which they have responded. Yet rarely are primary geologic and climatic data rigorously integrated into genomic studies in order to disentangle these relative drivers of organismal evolution. Because genomic divergence and speciation can occur by neutral drift, differential adaptation, or both, utilizing statistical frameworks in which multiple, non-mutually-exclusive hypotheses can be tested is necessary. Here we present ongoing work using low-coverage (~5x) whole-genome sequencing of individuals from sister *Gopherus* species and their hybrid populations. We aim to better understand the historic speciation and ongoing/recent hybridization of these non-model species through reconstructing demographic histories of both lineages, determining regions of elevated interspecific divergence, testing for genic regions under selection, and assessing signatures of genomic introgression. In addition, we are using primary sediment, tectonic, and climatic data over the past ~8 million years to establish geo-climatic hypotheses regarding what extrinsic factors might have shaped these patterns of genome evolution. Specifically, geologic development of the lower Colorado River region and paleo-monsoon system would likely have mediated gene flow and differential adaptation, respectively. Using a geo-genomic approach to disentangle the relative contributions of these extrinsic processes to genomic divergence in these species would advance our understanding of speciation and drivers of genomic evolution broadly.

W817: Population and Conservation Genomics 1

The Redwood Genomics Project for Conservation and Restoration of an Iconic California Endemic

Alison Dawn Scott[1], David B. Neale[1], Zane Moore[1] and Emily Burns[2], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Save the Redwoods League, San Francisco, CA

The California redwoods *Sequoia sempervirens* and *Sequoiadendron giganteum* are well-known for their size, long lifespan, economic importance, and most importantly for their role in the American conservation movement. In addition to these superlatives, the coast redwood (*Sequoia sempervirens*) boasts a genome over 31GB, made even more complex by its hexaploid status. Giant sequoia, as a diploid, rests at a modest 9GB. Using a combination of Illumina and Nanopore MinION sequencing, our team is assembling the genome of these majestic trees for the first time.

These genome assemblies will provide the foundation for the development of genomic tools to aid in the conservation and restoration of California's remaining redwood forests. We will use these tools to identify groves of particular conservation concern, identify adaptive variation in redwood populations, and improve management strategies. Incorporating genomic data with phenotypic observations and ecological variables will allow fine-scale management of redwoods across a heterogeneous landscape, helping preserve these iconic California endemics for generations to come.

W818: Population and Conservation Genomics 2

Conservation Genomics Applied to Federally Listed Rockfish Species under the US Endangered Species Act (ESA)

Krista M. Nichols, NOAA Fisheries, Seattle, WA

In 2010, the National Marine Fisheries Service listed Yelloweye (*Sebastes ruberrimus)* and Canary Rockfish (*S. pinniger*) as threatened and Bocaccio (*S. paucispinis*) as endangered in Puget Sound, WA, USA under the federal Endangered Species Act (ESA). However, this decision was made despite a lack of genetic data to directly address the first criterion of an ESA listing – Is the population segment "discrete" and "significant" from the remainder of the taxon? Indirect genetic evidence from other *Sebastes* spp. and other taxa was the primary basis of the listing decision. To address the first criterion, we collaborated with recreational fishing communities to collect tissue samples from these rarely encountered species in Puget Sound. With thousands of restriction-site associated DNA sequencing (RAD-seq) loci for each species, population genetics analyses to determine whether samples from Puget Sound were genetically "discrete" from samples collected from the outer coasts of the U.S. and Canada. Multiple analyses showed that Yelloweye Rockfish collected in inland waters of Puget Sound and British Columbia, Canada were genetically different from coastal populations. However, we found no evidence of population structure for Canary Rockfish. The sample size for bocaccio was insufficient to test the hypothesis. These data support the ESA designation status for Yelloweye Rockfish, with further evidence of finer level population structure evident within the Puget Sound. However, Canary Rockfish in Puget Sound are not a "discrete" population and do not meet the first criterion of the ESA. Collaboration among agencies and fishing communities, and the cost-efficient genetic analysis of thousands of genetic loci provided the framework for the first de-listing of a marine fish species under the ESA. Oceanographic parameters contributing to the extant genetic diversity in Yelloweye Rockfish are currently being explored.

W819: Population and Conservation Genomics 2

The Genetic Basis of Tiger Pelage Variations and Conservation Implications

Shu-Jin Luo, Peking University, Beijing, China
The white tiger, an elusive Bengal tiger variant with white fur and dark stripes, has fascinated human society for centuries ever since its discovery in the jungles of India. However, a heated debate centers on whether the white tiger is after all a genetic defect that is worthless for conservation. In fact, tigers exhibit an amazing array of phenotypic diversity with at least four pelage color morphs whose genetic bases have eluded us until now: wild-type orange, white, golden and stripeless snow tigers. Here we conducted RADSeq-based GWAS in a pedigree of 16 captive tigers segregating at the putative *white* locus, followed by WGS of the three parents, to identify a missense mutation p.A477V in *SLC45A2* as causal for the white tiger. Subsequently, using whole-genome resequencing we determined the genetic basis of the golden tiger to be a recessive p.H587Y in *CORIN*. Genotyping of SLC45A2 and CORIN in over 200 unrelated tigers validated that the snow tiger is caused by a combination of recessive mutations responsible for white and golden tigers respectively. These findings have advanced our understanding of mammalian coat color genetics by identifying *CORIN* as a mammalian *wideband* gene and also resolved the longstanding white tiger mystery. Our results highlighted the significance of the *SLC45A2* substitution as part of the species' natural polymorphism that primarily affects only pigmentation. Despite the low frequency, this polymorphism has persisted for at least several hundreds of years and should be considered a part of the genetic diversity of tigers that is worth conserving.

W820: Population and Conservation Genomics 2

Selection at a Locus of Major Effect Is Responsible for Evolution of Sexual Maturation in Columbia River Anadromous Steelhead

Steven Micheletti, Columbia River Inter-Tribal Fish Commission, Portland, OR
Evolution can lead to variation in phenology in natural organisms and previous studies have attributed sexual maturation and run-timing variation in anadromous steelhead to GREB-1L, an oestrogen target gene. However, previous GWAS techniques used to identify this gene only accounted for about 0.5-2.0% of the *O. mykiss* genome, leaving uncertainty on the genetic basis of this trait. Here we used pooled whole-genome resequencing to acquire extremely dense genome-wide data (6 million SNPs covering ~60% of the genome) for steelhead in multiple spawning tributaries to map loci associated with maturation and run-timing phenotypes. Dense mapping confirmed GREB-1L to be the locus of major effect, and a diagnostic marker from this gene showed that association with maturation and run-timing variation differed between the two major lineages (coastal and inland) in the Columbia River Basin. Individual migration patterns from PIT-tag data demonstrated that GREB-1L is directly and consistently associated with timing of sexual maturation, which is not always related to migratory timing of entry into freshwater. Steelhead that arrived to spawning sites early and premature versus those that arrived later as mature fish were generally fixed for alternative homozygous genotypes. Fish with heterozygous genotypes for this diagnostic locus typically had intermediate phenotypes. This study not only further demonstrates the genetic basis for sexual maturation in this species, but also illustrates how dense genome mapping and detailed phenotypes can clarify genotype to phenotype associations across broad geographic ranges of natural species.

W821: Population and Conservation Genomics 2

A Century of Conservation Genetics: Reconstructing Historical Population Size & Genetic Diversity of African Lions (*Panthera leo*) to Ensure their Future

Caitlin J. Curry, Texas A&M University, College Station, TX
Current scientific estimates of lion population sizes range from as low as 16,500 to as high as 200,000 individuals estimated using various methods. With the use of modern biotechnology, we determined the genetic architecture of both historical (>100 years ago) and contemporary (2000 to present) African lion populations across the traditional range states in Africa. Both datasets were analyzed using the same methods allowing for a more direct comparison over time than has previously been employed. The historical lion dataset is DNA isolated from high quality and well-documented museum specimens while the contemporary lion dataset is from modern material and data from several recently published studies. Genetic diversity was examined for both datasets by sequencing of the mitochondrial genome and genotyping of fourteen microsatellite molecular markers in the nuclear genome. Combining the datasets of historical as well as contemporary populations provides quantitative measures on the extent of change in the genetic diversity of lions over the past 100 years and a basis for assessing the genetic health of the African lion. Knowledge about the genetic health of the African lion will help us to identify existing wild lion populations that are most at risk and make recommendations to guide management actions to safeguard the future genetic health of the African lion.

W822: Population and Conservation Genomics 2

The Human Animal Bond: Understanding How the Dog's Brain Was Selected for Social Interaction with Humans during Thousands of Years of Domestication

Kristopher Irizarry, The Applied Genomics Center, College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA
The dog is called our "best friend," yet until very recently, the specific changes that led to the strong bond between dogs and humans has remained enigmatic. Genomics methods have offered an unprecedented opportunity to unravel the mysteries underlying dog domestication. These powerful and data-dense genetic approaches have refined our understanding how dogs transformed from wolves into the hundreds of breeds that live among us today. Through a combination of genetic association studies, whole genome sequencing and gene expression studies, the veil covering our evolutionary history with dogs has finally been lifted and the initial discoveries consist of many surprises, that when viewed in the context of 'our best friend,' make a lot of sense. A longstanding question pertaining to domestication of dogs is whether there was any selection for cognitive, behavioral or communication phenotypes that may have contributed to a strong interspecific bond between humans and their companion dogs. Through the integration of many diverse genetic studies, spread across numerous laboratories and countries, a picture emerges of how canine brains changed during domestication and how artificial selection modulated oxytocin mediated phenotypes resulting in the strongest human-animal bond that exists on the planet. Through a review of these studies and findings, we can paint a detailed picture of the how the dog's brain was sculpted over thousands of years by domestication at the hands of our ancestors.

W823: Population and Conservation Genomics 2

Diagnostic Population Genomics for Economical and Efficient Source Determination of Tephritid Fruit Fly Pests

Julian R. Dupuis[1,2], 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

Each year, thousands of exotic and invasive pest species are intercepted at ports of entry or detected in the environment; rapid species and source identification is crucial to management and eradication of these species, and prevention of their establishment. Multiple species of tephritid fruit flies in the genera *Anastrepha*, *Bactrocera*, *Ceratitis*, and *Zeugodacus* represent major threats to United States agriculture, and some are among the most economically damaging agricultural pests world-wide. These species often attack a wide range of commercial produce, and active monitoring programs in several regions of the United States aim to detect introductions. Our main goal is to develop cost-effective and efficient diagnostic SNP panels for identifying the geographic source of intercepted material, and I will discuss results from several species of tephritids. Given that globalized trade is constantly moving these flies and their habitat (i.e. fruits and vegetables) around the world, these systems present unique challenges for population genetic inference, which I will highlight. We have also developed a user-friendly, web-based visualization tool for contextualizing population genetic analyses in geographic space. This tool will allow regional pest managers to compare genotype results of intercepted material (generated using our diagnostic SNP panels) with the world-wide datasets created in our lab and determine the most likely source population. Taken together, these research efforts provide a highly-applied use for population genomics, valuable context for understanding tephritid fruit fly invasion biology, and effective tools for managers in the current era of global trade and travel.

W824: Poultry 1

The Genomic Architecture of Marek's Disease Virus-Induced Tumors

Alec Steep, Michigan State University, East Lansing, MI, Hongen Xu, Technical University of Munich, Freising, Germany, Alexis Black Pyrkosz, USDA, ARS, ADOL, East Lansing, MI, William Muir, Purdue University, West Lafayette, IN, Mary E. Delany, Animal Science, University of California, Davis, CA, Dmitrij Frishman, Technical U. of Munich, Freising, Germany and **Hans H Cheng**, USDA, Agricultural Research Service, East Lansing, MI

Marek's disease (MD) is characterized by the rapid onset of CD4 T cell lymphomas in chickens and is caused by the highly oncogenic Marek's disease virus (MDV). Despite widespread use of MD vaccines that prevent tumor formation, more virulent MDV strains have repeatedly arisen. Enhancing genetic resistance is an attractive complementary control strategy and, therefore, there is a critical need to understand how MDV induces neoplastic transformation. MDV Meq, a bZIP transcription factor and the viral oncogene, is required but not sufficient for tumor formation as all birds are infected. Thus, additional mutations in the host genome are required. To identify these somatic mutations, whole genome and transcriptome sequence analyses of MD tumors and matched control tissues were conducted. In ~85% of the tumors, mutations or low gene expression were found in IKAROS family zinc finger 1 (IKZF1), a zinc-finger transcription factor associated with T-cell development and chromatin remodeling. Similar to human Acute Lymphocytic Leukemia (ALL), MD tumors contain dominant negative somatic mutations in zinc-finger binding domains suggesting its role as a tumor suppressor gene. Furthermore, pathways enriched for differential gene expression between tumors and matched controls mimic proposed Ikaros targets, suggesting that deregulated Ikaros may reprogram T cells for hallmarks of cancer. Collectively, these data suggest MDV Meq and host IKZF1 regulate the decision between viral replication and latency, which is perturbed by somatic mutations in IKZF1 leading to tumorigenesis. This knowledge should aid genomic selection for MD resistance as well as targets for improved MD vaccines.

W825: Poultry 1

Lung Transcriptome Response to Newcastle Disease Virus in Differentially Resistant Inbred Chicken Lines

Susan J. Lamont, Department of Animal Science, Iowa State University, Ames, IA

W826: Poultry 1

Mechanism of *Salmonella* Colonization in Chicks

Huaijun Zhou, Animal Science, University of California, Davis, CA and Khin Khine Zar Mon, University of California, Davis, davis, CA

Salmonella Enteritidis (SE) is commonly isolated from the hatcheries and also a main source of food-borne illnesses after consumption of contaminated poultry products. SE in chicken can colonize persistently for a long period time within the host and serve as reservoir for transmission of the Salmonella to the naïve host. Mutual balanced metabolic relationship established between the host and resident gut microbiota could be also disrupted by the pathogen invasion. The aim of the current study was to examine three-way interaction that occurred between the host, resident gut microbiota and pathogen during SE infection in two-week old layer chicks. Microbiome profile was carried out through 16S rRNA gene sequencing and metabolite profile was analyzed utilizing the Gas Chromatography-Mass Spectrometry (GC-MS). The results revealed that gut homeostasis was disrupted with SE infection. Specifically, enriched colonization by non-specific minority members of the community following SE infection highlighted the possible pathogen-driven microbial community restoration effect. Arginase-mediated host immune modulation effect was detected through up-regulation of arginine and proline metabolism as part of SE associated host metabolic manipulation strategy to secure a niche within the hostile intestinal environment. Down-regulation of the TCA cycle activity on the other hand point towards host innate immunity metabolic reprogramming strategies to limit nutrient availability to pathogen that could further promote its survival and growth. Taken together, current study provided important insight into the metabolic adjustment strategies exercised by host and resident commensal population when dealing with pathogen associated disruption in the intestinal environment.

W827: Poultry 1

An Update of Chicken Gene Expression Resources

Fiona McCarthy[1], Amanda M. Cooksey[1], Cathy R. Gresham[2], Ken Pendarvis[1], Michael Rice[1], Jinhui Zhang[1], Eric Lyons[3], Bindu Nanduri[4] and Shane Burgess[1], (1)University of Arizona, Tucson, AZ, (2)Institute for Genomics, Biocomputing & Biotechnology, Mississippi State, MS, (3)The University of Arizona, Tucson, AZ, (4)Mississippi State University, Mississippi State, MS
New genomic technologies are accumulating information about functional elements within genomic sequence at a faster rate than ever before, and understanding how these elements contribute to traits and phenotypes requires a concomitant effort to understand the function of these genes and regulatory regions. We describe several resources that will help researchers to better utilize new and existing information and apply it to their own research projects. The first resource is Chickspress (http://geneatlas.arl.arizona.edu), a gene expression resource for chicken tissues. Chickspress incorporates both NCBI and Ensembl gene models, and links these gene sets with experimental gene expression data from mRNAs, ncRNA and proteins, and displays QTL information. Researchers may use this resource to identify gene expression sets based upon tissue types or to compare expression of a gene product across multiple tissues. Via the AgBase database we are also using data provided by Chickspress to develop workflows for predicting ncRNA function. Providing functional prediction for ncRNAs will enable information about these important regulatory molecules to be included in functional genomics analyses, and guide experiments to confirm ncRNA function. As part of the AgBase biocuration effort we continue to provide updated Gene Ontology (GO) annotations for functional analysis. We also provide standardized gene nomenclature for chicken genes via the Chicken Gene Nomenclature Committee (CGNC), and this data is distributed to NCBI Entrez. The CGNC is working closely with the Vertebrate Gene Nomenclature Committee (VGNC) to support the expansion of gene nomenclature efforts to include other FAANG species.

W828: Poultry 1
Avian ORFeome Initiative: A Community-Led Biological Resource for Avian Functional Genomics
Muhammad Munir, The Pirbright Institute, Surrey, United Kingdom

W829: Poultry 1
Effect of Heat Stress on Nutrient and Amino Acid Transporters in Meat-Type Chickens
Samuel E. Aggrey, University of Georgia, Athens, GA

W830: Poultry 1
Tissue Responses to Heat Stress in the Modern Broiler
Carl J. Schmidt, Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

W831: Poultry 1
Effect of Embryonic Thermal Manipulation and Posthatch Thermal Challenge on Turkey Growth and Skeletal Muscle Development
Gale M. Strasburg, Michigan State University, East Lansing, MI

W832: Poultry 1
Differential Response in the Cecal Tonsil to Aflatoxin B1 in Domesticated and Wild Turkeys
Kent M. Reed, University of Minnesota, St. Paul, MN

W833: Poultry 1
Gene Expression Differences Associated with Egg Production Rates in Turkey Hens
Kristen Brady[1], Julie Hicks[2], Julie Long[3], Hsiao-Ching Liu[2] and **Tom E. Porter**[1], (1)University of Maryland, College Park, MD, (2)North Carolina State University, Raleigh, NC, (3)USDA BARC, Beltsville, MD
Substantial variation in egg production exists among commercial turkey hens. Total egg production is correlated with ovulation frequency, which is governed by the hypothalamic-pituitary-gonadal (HPG) axis. Each ovulation is stimulated by a preovulatory surge of progesterone and luteinizing hormone (LH), caused by release of gonadotropin releasing hormone (GnRH) and repressed by gonadotropin inhibitory hormone (GnIH). Differences in the HPG axis between low and high egg producing turkey hens were explored by determining mRNA levels by RT-qPCR for key genes regulating ovulation and ovarian steroidogenesis. From the top and bottom 15% in egg production of 200 commercial line turkey hens, six high egg producing (HEP) hens and six low egg producing (LEP) hens were sampled, with half of the hens sampled outside of the preovulatory surge and half sampled during the preovulatory surge (n=3 per group). The hypothalamus, pituitary gland, and granulosa layers of the largest follicle (F1G) and the fifth largest follicle (F5G) were collected. LEP hens exhibited increased mRNA levels for genes associated with ovulation inhibition, such as hypothalamic GnIH (*NPVF*) and pituitary GnIH receptor (*NPFFR1*), as well as decreased mRNA levels for genes associated with ovarian stimulation, such as pituitary follicle stimulating hormone (*FSHB*) and LH (*LHB*). Interestingly, LEP hens demonstrated higher mRNA levels for the FSH receptor (*FSHR*) and the LH receptor (*LHCGR*) in the F5G, as well as the progesterone receptor (*PGR*) in the hypothalamus. Genes associated with progesterone production, steroidogenic acute regulatory protein (*STAR*), cholesterol side-chain cleavage enzyme (*CYP11A1*), and 3β-hydroxysteroid dehydrogenase (*HSD3B1*), were up-regulated in the F1G of HEP hens, while in LEP hens these genes were up-regulated in the F5G. Different degrees of stimulation and inhibition within the HPG axis at the mRNA level were noted in LEP and HEP hens in each of the tissues examined. The results support altered functioning of the HPG axis as an underlying cause for differences in egg production among turkey hens. The same RNA samples are being analyzed by RNAseq to identify additional gene networks in the HPG axis influencing turkey egg production.

W834: Poultry 1
Inversion Variants in the Genus *Gallus* Identified by Comparative Genome Analyses
Giridhar Athrey, Texas A&M University, College Station, TX

W835: Poultry 1
Genomic Analysis of the Wingless-2 Mutation
Mary E. Delany, Animal Science, University of California, Davis, CA

W836: Poultry 1
Extensive Chicken MHC-*Y* Haplotype Diversity
Ronald M. Goto[1], Jibin Zhang[1], Janet E. Fulton[2] and **Marcia M. Miller**[1], (1)Beckman Research Institute, City of Hope, Duarte, CA, (2)Hy-Line International, Dallas Center, IA
MHC-*Y* and MHC-*B* are two clusters of major histocompatibility complex genes located on chicken chromosome 16. MHC-*Y* and MHC-*B* haplotypes assort independently as the result of a highly recombinogenic sequence separating the two regions. MHC-*B* is considered to be the classical (peptide antigen presenting) MHC of the chicken. The MHC-*Y* region is enigmatic. While initially revealed by the presence of aberrant restriction fragments in MHC-*B* Southern hybridizations, MHC-*Y* is now known to contain specialized MHC class I-like, MHC class II beta-chain and C-type lectin-like genes. The MHC-*Y* class I-like genes are polymorphic and clearly specialized. Progress in defining how these genes contribute to immunity has been hampered, in part, by lack of an efficient means for identifying the MHC-*Y* genotypes of individual birds. We have developed a PCR-based fingerprinting method for distinguishing MHC-*Y* haplotypes. PCR products are generated using primer pairs to conserved sequences that span a simple-sequence repeat region of variable length found upstream of each MHC-*Y* class I gene present within MHC-*Y*. Sequence lengths vary between haplotypes. Haplotype-specific fingerprints revealed in agarose gels allow MHC-*Y* genotypes to be distinguished. These fingerprints provide evidence for extensive MHC-*Y* polymorphism present within different heritage breeds and within a number of broiler and layer lines. Many MHC-*Y* haplotypes segregate in chicken populations. This new MHC-*Y* typing method makes it possible to type large numbers of birds and therefore much easier to investigate the contributions of MHC-*Y* genetics to desirable traits in chickens raised for meat and eggs.

W837: Poultry 1
Steps to Efficient Gene Editing in the Chicken using Cultured Primordial Germ Cells (PGCs).
Helen M. Sang, University of Edinburgh, Easter Bush, United Kingdom

W838: Poultry 1
OmniChickens™ – an Elegant Model of Gene Editing in Poultry
Robert J Etches, Ligand Pharmaceuticals, Emeryville, CA
The genome of chickens can now be modified using a variety of tools to insert DNA into random locations within the genome (van de Lavoir et al, 2006; Macdonald et al, 2012), knock out specific genes using homologous recombination vectors (Schusser et al, 2013; Schussser et al, 2016) or CRISPr (Dimitrov et al 2016), and excise large segments of the genome using Cre recombinase (Leighton et al, 2016). These tools have been combined to create OmniChickens™, a platform to enable discovery of fully human sequence antibodies with therapeutic potential. OmniChickens™ were created by inactivating the endogenous heavy and light chain genes using homologous recombination vectors designed to remove small segments of the JH region of the heavy chain locus or the VJC region of the light chain locus. A larger region of the heavy chain locus was deleted by inserting loxP sites in the JH region and 5' of the functional heavy chain V region. When birds carrying these loxP sites were mated to birds expressing Cre recombinase, the offspring carry an approximately 20Kb deletion spanning the region flanked by the loxP sites. The homologous recombination and CRISPr vectors described above also inserted an attP site into the locus. Sequences encoding human immunoglobulin heavy and light chain V regions were inserted into the attP sites using phi-31 integrase to create chickens expressing chimeric antibodies with fully human sequence V regions and chicken constant regions. These birds are now used by several pharmaceutical companies as a tool for the discovery of therapeutic antibodies.
These genetic modifications demonstrate that the tools of cell and molecular biology can be used to create complex genetic modifications in the genome of chickens. By combining tools for gene targeting (eg. Homologous recombination, TALENs, CRISPr), for DNA insertion (eg. phi-31 integrase) and for targeted deletion (eg. using Cre recombinase), the potential for complex genetic modifications to the genome of birds is limited only by imagination.
References
Dimitrov L, Pedersen D, Ching KH, et al (2016). PLoS ONE 11:e0154303. doi: 10.1371/journal.pone.0154303
Leighton PA, Pedersen D, Ching K, et al (2016). Transgenic Res. doi: 10.1007/s11248-016-9952-6
Macdonald J, Taylor L, Sherman A, Kawakami K, Takahashi Y, Sang HM, Mcgrew MJ (2012). Proc Natl Acad Sci USA 109:E1466-E1472. doi: 10.1073/pnas.1118715109
Schusser B*, et al*. (2013). *Proc Natl Acad Sci U S A* 110(50):20170-20175.
Schusser B, Collarini EJ, Pedersen D, et al (2016). Eur J Immunol. doi: 10.1002/eji.201546171
van de Lavoir M-C, Diamond JH, Leighton PA, et al (2006) Germline transmission of genetically modified primordial germ cells. Nature 441:766–769. doi: 10.1038/nature04831

W839: Poultry 1
Generating a Vocal Learning Songbird Resource for Generation of PGCs, Genetic Vectors and Transgenic Animals
Erich Jarvis, Howard Hughes Medical Institute, New York, NY

W840: Poultry 1
An Industry Perspective: Producing Transgenic/Gene Edited Agricultural Animals to Increase Health and Efficiency and Improve Animal Welfare.

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

W841: Poultry 1
An Industry Perspective on Application of Genetic Modification in Poultry
Janet E. Fulton, Hy-Line International, Dallas Center, IA

W842: Poultry 1
Advancing Germ Line Transmission for Avian Species to Enable Avian Genetic Rescue
Thomas Maloney, Revive & Restore, Sausalito, CA
Wild birds helped to initiate the modern conservation movement and continue to be at the forefront of conservation concerns and interventions. However, they have lagged considerably behind other organisms in biotechnology applications. New advances in biotechnologies make it possible to address intractable conservation concerns in the face of growing human influences (e.g., habitat loss & fragmentation, novel disease exposure and climate change). Biotechnology, in particular genome editing, offers a diverse array of potential applications for avian conservation from combatting disease, facilitating adaptation, controlling invasive species, aiding population recovery, to de-extinction. Revive & Restore's avian work is focussed on developing effective and transferable germ line transmission strategies for wild birds. Our work has been initiated as a de-extinction endeavor but, if successful will present a host of new tools to abate threats and enhance condition of wild birds. With diverse partners our projects are outlining the steps to implement genetic rescue for birds from the lab to the environment.

W843: Poultry 1
Current Applications of Gene Editing: Opportunities and Obligations
Kevin Wells, University of Missouri, Columba, MO

W844: Poultry 2
Assessing Feed Efficiency in Group-Housed Broilers in Commercial Conditions using a Real-Time Automated System
Guilherme J. M. Rosa, University of Wisconsin-Madison, Madison, WI

W845: Poultry 2
Identification of Stem and Absorptive Cells in the Chicken Yolk Sac and Intestine using *in situ* Hybridization
Eric A. Wong and Haihan Zhang, Virginia Tech, Blacksburg, VA
The chick derives its nutrients from the yolk during embryogenesis and feed in the intestine post-hatch. Both the yolk sac and the intestinal villi contain absorptive cells that mediate the uptake of nutrients, as well as other differentiated cells that arise from stem cells. The objective of this project was to identify absorptive and stem cells in the yolk sac and intestinal villi using the RNAscope in situ hybridization method with cell-type specific probes. Yolk sac was collected at embryonic day 11 to 19 (e19) and day of hatch (doh). Small intestine was collected at e19, doh and days 1, 4 and 7 post-hatch. Absorptive cells were identified with probes for peptide transporter 1 (PepT1) mRNA and putative stem cells were identified with probes for olfactomedin 4 (Olfm4) or Leucine-rich repeat G protein-coupled receptor 5 (Lgr5) mRNA. In the small intestine, putative stem cells expressing Olfm4 and Lgr5 mRNA were localized to the intestinal crypts. At e19, when the intestinal crypts were still rudimentary, cells staining for Olfm4 mRNA were already present. Unlike mammals, the crypt did not contain a population of Paneth cells that did not express Olfm4/Lgr5 mRNA. At doh, there was a population of cells that expressed neither PepT1 nor Olfm4, which may be progenitor transit amplifying cells. In the yolk sac, endodermal cells expressed PepT1 mRNA. Cells that expressed Lgr5 but not Olfm4 were localized to the vascular endothelial cells lining the blood vessels and may be hematopoietic stem cells.

W846: Poultry 2
Genetics of Ascites: Energy Metabolism, and Ties to Woody Breast?
Douglas D. Rhoads[1], Khaloud Alzahrani[1], Alia Parveen[1], Shatovisha Dey[1,2] and Nicholas Anthony[3], (1)University of Arkansas, Cell and Molecular Biology Program, Fayetteville, AR, (2)Indiana University School of Medicine, (3)University of Arkansas, Cell and Molecular Biology Program
Our collaborative consortium has been pursuing the underlying genetics of ascites in broilers. Recently, we used whole genome resequencing in our ascites experimental research line to identify 31 chromosomal regions with potential association with ascites phenotype. Two of those regions have now been validated for containing ascites QTLs. One region contains the CPQ gene and the other contains the LRRTM4 gene. The exact role of these genes in affecting ascites phenotype is not known. We had recently published findings that there are differences in breast mitochondrial biogenesis in breast muscle of males in the RES and SUS ascites research lines. We have now expanded this analysis to additional skeletal muscles, and to both genders. Our qPCR analyses show a marked difference in the ratio of mitochondrial to genomic sequences in breast, thigh and lung, and extreme variations between the sexes. For males, there is a significantly higher level of mitochondrial DNA in the susceptible line than in the resistant line. The difference is most dramatic in lung>thigh>breast. Others have identified differences in central metabolic protein levels when comparing normal to Woody Breast samples. Woody Breast is associated with tissue hypoxia similar to ascites. We therefore analyzed mitochondrial DNA levels in normal and Woody Breast broiler samples. Our current findings are that there is a significant difference. Surprisingly, the level of mitochondrial DNA in normal breast is more than twice that of Woody Breast ($P=0.0000048$). We propose that Woody Breast is a manifestation of the interaction between mild tissue hypoxia and differences in metabolic demands associated with mitochondrial biogenesis.

W847: Poultry 2
Microscopic and Transcriptomic Analyses of Wooden Breast Disease in Commercial Broiler Chickens: A Time-Series Study

Behnam Abasht, University of Delaware, Newark, DE

W848: Poultry 2
Epigenetic Effects of a Diet High in Methyl-Donors on Performance, Gene Expression, and DNA Methylation
Christopher M. Ashwell, North Carolina State University, Raleigh, NC

W849: Poultry 2
Use of Chemical Proteomics to Identify Novel Chicken Kinases and Deubiquitinases
Mariola J. Edelmann, University Of Florida, University of Florida, FL

W850: Poultry 2
HPIdb: A Curated Database for Host-Pathogen Interactions
Bindu Nanduri, Mississippi State University, Mississippi State, MS

W851: Proteomics
Soybean Proteomics to Reveal the Flooding-Tolerant Mechanism
Setsuko Komatsu, University of Tsukuba, Tsukuba, Japan and Katsumi Sakata, Maebashi Institute of Technology, Maebashi, Japan
Soybean is sensitive to flooding stress, which markedly reduces its growth. To identify the mechanism of flooding tolerance at initial stage in early-stage soybean, proteomic technique was used. Flooding tolerant mutant line and abscisic acid (ABA)-treated soybean, which exhibited flooding tolerant phenotype, were used as materials. Two-day-old soybeans were flooded for 3 h as initial flooding stress and roots were collected for proteomic as well as metabolomic and transcriptomic analyses. Data were analyzed using functional categorization, cluster separation, and *in silico* protein-protein interaction analyses. Furthermore, commonly changed metabolites, proteins, and genes between mutant and ABA-treated soybeans were considered as flooding-tolerance related candidate factors. Finally, omics results were integrated to analyze the flooding tolerant mechanism in soybean. These results suggest that flooding tolerance at initial stage in early-stage soybean might be through protecting newly synthesized proteins and enhancing activities of antioxidative enzymes to remove reactive oxygen species. Furthermore, regulation of energy metabolism is crucial step for flooding tolerance and inhibition of cell wall loosening might contribute to flooding tolerance in soybean.

W852: Proteomics
Biotic Stress Modulation of Protein Acetylation Revealed by Acetylome Profiling
Justin W. Walley, Iowa State University, Ames, IA
Lysine acetylation is a key post-translational modification that regulates diverse proteins involved in a range of biological processes. The role of histone acetylation in plant defense is well established and it is known that pathogen effector proteins encoding acetyltransferses can directly acetylate host proteins to alter immunity. However, it is unclear whether endogenous plant enzymes can modulate protein acetylation during an immune response. Here we investigate how the effector molecule HC-toxin, a histone deacetylase inhibitor, produced by *Cochliobolus carbonum* race 1, promotes pathogen virulence in maize through altering protein acetylation. Using mass spectrometry we globally quantified the abundance of 3,636 proteins and the levels of acetylation at 2,791 sites in maize plants treated with HC-toxin as well as HC-toxin deficient (Tox-) or producing strains (Tox+) of *C. carbonum*. Analyses of these data demonstrate that acetylation is a widespread post-translational modification impacting proteins encoded by many intensively studied maize genes. Additionally, these data establish that the activity of plant-encoded enzymes can be modulated to alter acetylation of non-histone proteins during an immune response. Furthermore, proteins that are hyperacetylated in response to HC-toxin and Tox+ treatment are enriched in transcriptional regulators. Among the hyperacetylated proteins is a bHLH transcription factor that when mutated to mimic constitutive acetylation results in enhanced ability to trigger target gene expression in protoplast assays. Finally, we used virus induced gene silencing (VIGS) to knockdown expression of this bHLH, in a normally resistant corn line carrying the resistance gene *Hm1*, which resulted in susceptibility towards *C. carbonum* (TOX+). Together these findings suggest that HC-toxin transcriptionally triggers an inappropriate defense response to promote *C. carbonum* virulence.

W853: Proteomics
Elucidating the Role of MORC1 Protein Interactors during Plant Immunity against *Phytophthora spp.*
Natasha Jackson, University of California, Riverside, Riverside, CA
Microrchidia (MORC) proteins are a subset of the GHKL ATPase superfamily, containing GHKL and S5 domains that form a catalytically active ATPase module. Proteins containing this GHKL ATPase motif play roles in chromatin remodeling, heat shock responses, signal transduction, and DNA mismatch repair. MORC proteins have been described as components involved in the RNA-directed DNA methylation pathway and chromatin remodeling. Recently, we have found that MORC1 is required for plant immunity against the root rot pathogen, *Phytophthora cinnamomi*. Previously, we reported that MORC1 regulates cell death and plant immunity against *P. infestans* in a species-specific manner behaving as a positive regulator in Arabidopsis and potato and as a negative regulator in tomato and *Nicotiana benthamiana*. We mapped this antagonistic phenotype to the C-terminal region of these MORC1 proteins suggesting that the species-specific effects on resistance are due to how and to whom these MORC1 proteins interact with (positive and negative regulators) at their C-terminal regions. We have identified two proteins that differentially interact with the C-terminal region of the potato and tomato MORC1 proteins. Our results have shown that silencing the *N. benthamiana* homolog of the potato MORC1-Interacting Protein (a transcription factor), compromised the cell death induced by INF1, the major secreted elicitin of *P. infestans*. Furthermore, silencing this transcription factor also increased susceptibility to *P. infestans* and *P. cinnamomi* in *N. benthamiana*. Altogether, our results suggest that this transcription factor also acts as positive regulator of cell death and plant immunity against these devastating oomycete pathogens.

W854: Proteomics

Proteomic Analysis of Liver Tissue from Heat Stressed and Newcastle Disease Virus Infected Inbred Chicken Lines.

Perot Saelao[1], Ying Wang[1], Rodrigo A. Gallardo[2], David A. Bunn[1], Terra R. Kelly[2], Susan J. Lamont[3] and Huaijun Zhou[1], (1)Animal Science, University of California, Davis, CA, (2)University of California, Davis, School of Veterinary Medicine, Population Health and Reproduction, Davis, CA, (3)Department of Animal Science, Iowa State University, Ames, IA

Newcastle disease virus (NDV) and heat stress are two major factors impacting rural poultry production in developing nations. A holistic understanding of the effects of these two combined stress factors can assist in informing and developing improved treatments through novel genetic approaches. The objective of this study was to identify specific protein and signal pathways associated with NDV infection and heat stress in two highly distinct inbred chicken lines. Two inbred chicken lines, Fayoumi and Leghorn, were treated with NDV at 21 days of age while under the effects of constant heat exposure (35C, 65% humidity) starting from 2 weeks of age. Liver samples were collected at 2 and 6 days post-infection and flash frozen for protein analysis. Overall, Fayoumi birds had more proteins up regulated compared to Leghorn birds at both time points. Proteins encoded by genes MYL4, UFC1, COLA1, and CATH1 were inversely regulated between the two lines at both time points, with Fayoumi having an increase these proteins, while Leghorn decreased its production of these proteins. Furthermore, key pathways in cell proliferation and AGE-RAGE pathways were highly differentiated between the two lines. Proteomic analysis of chicken liver samples under the effects of heat stress and Newcastle disease virus infection enables the identification of retrospective biomarkers to further investigate and characterize the response to these stressors. These results continue to highlight the unique and highly novel response difference between the two genetic lines and will continue to help inform on the significant proteins and response pathways in resolving NDV infection during heat stress.

W855: Proteomics

Proteome Profiling of Equine Myofibrillar Myopathy Identifies Diminished Peroxiredoxin 6

Sudeep Perumbakkam[1], Stephanie J. Valberg[1], Erica C. McKenzie[2] and Carrie Finno[3], (1)Michigan State University, East Lansing, MI, (2)Oregon State University, Corvallis, OR, (3)University of California Davis, Davis, CA

Arabian horses are adept at endurance races, routinely competing in distances of 50 to 100 miles. Exertional rhabdomyolysis (ER), literally the dissolution of skeletal muscle with exercise, is a common occurrence in endurance horses affecting 4 to 18% of competitors. To date, however, genetic causes of ER have not been identified in the Arabian breed. Recently, our lab has identified a muscle disease with a suspected genetic predisposition in Arabian horses called myofibrillar myopathy (MFM). MFM causes muscle stiffness, exertional muscle pain and is characterized by myofibrillar disarray and ectopic protein aggregates of unknown origin. To investigate the pathophysiology of equine MFM, we compared the skeletal muscle proteome and 3 h post-exercise transcriptome of gluteal muscle in MFM and control Arabian horses using isobaric tags for relative and absolute quantitation (iTRAQ). Proteome analysis revealed significantly lower content of antioxidant peroxiredoxin 6 (PRDX6, ↓4.14 log2 fold change [FC]), sarcomere protein tropomyosin (TPM2, ↓3.24x) and higher fatty acid enzyme carnitine palmitoyl transferase (CPT1B, ↑3.49x) in MFM vs. control muscle at rest. Results indicate that, in MFM horses, protein aggregation may arise from oxidative damage to sarcomeric and cytoskeletal proteins as a result of diminished cysteine rich antioxidants such as peroxiredoxin 6 and a limited capacity to reduce free radicals generated through fatty acid oxidation during exercise.

W856: Proteomics

High-Quality Assembly of European House Dust Mite Genome, Transcriptome and Proteome Reveal a Wide Range of Novel Allergens

Ming-Qiang WANG[1], Kevin Yi Yang[1], Carol Po-Ying Lau[1], Zeng Xi[1], Elaine Lai-Han Leung[2], Zhi-Gang Liu[3] and Stephen Kwok-Wing Tsui[1], (1)School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong, (2)State Key Laboratory of Quality Research in Chinese Medicine, Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Macao, Macao, (3)State Key Laboratory of Respiratory Disease for Allergy at Shenzhen University, School of Medicine, Shenzhen University, Shenzhen, China

European house dust mite (HDM), scientifically known as *Dermatophagoides pteronyssinus*, is an especially important allergenic animal species. In this study, we aimed to construct a high-quality reference genome with the transcriptome and proteome methods to characterize novel allergen genes. A total of 62 Gb genomic data deposited with the BioProject number PRJNA388362 was generated from PacBio SEQUEL, Illumina HiSeq 2000 and Ion Torrent sequencing platforms. We performed *de novo* assembly, scaffolding, gap filling and polishing processes to obtain the 66.8 Mb assembly contains 1,390 contigs and 634 scaffolds, with scaffold and contig N50 being 194 kb and 80 kb, respectively. The assembly contains 8.9% repetitive sequences and 15,339 annotated protein-coding genes. The genome completeness as determined by BUSCOs analysis using 1,066 core genes from arthropoda_odb9 dataset was 89.1%. Our high-quality genome represents an 10-fold improvement in contig N50 compared to a previously published American HDM genome. In addition, gene structures of 21 canonical and 11 non-canonical allergens gene from *Der p 1* to *Der p 34* have been characterized with transcriptome data supported. We also identified 53 potential allergen protein sequences which were highly similar with allergens sequences of other species. Immunoblotting results showed that 50 HDM proteins that were bound by specific IgE in the sera of patients with HDM allergy. Those proteins were identified by MALDI-ToF MS and analyzed with the proteome predicted by the assembled genome. In summary, this study provides important genetic resources for further development of diagnostics and immunotherapeutic vaccines for HMD allergic patients.

W857: QTL Cloning

Natural Variation in QTL Controlling Root Growth Angle Enhances Salt Avoidance in Rice

Yusaku Uga, National Agriculture and Food Research Organization, Tsukuba, Japan

Crop production is greatly affected by environmental stresses, such as drought, submergence, and high salinity. Global climate change that has occurred in recent years has exacerbated the effects of these stress factors on crop growth. The root system is an essential organ for taking up water and nutrients from the ground. Adequate root system architecture (RSA) is especially important for crop growth in soil conditions where

water and nutrients are deficient. The root growth angle (RGA), which is one of the important components of RSA, determines the direction of root elongation in the soil and affects the area within which roots may capture water and nutrients. Previously, our group isolated *DRO1*, a quantitative trait locus (QTL) controlling RGA, on rice chromosome 9. Field tests for *DRO1* demonstrated that deep rooting improved yield performance in rice under drought conditions. Recently, we cloned another QTL for RGA on rice chromosome 7, *qSOR1* (*quantitative trait locus for SOIL SURFACE ROOTING 1*). A non-functional allele of this QTL leads to soil-surface rooting in paddy and upland fields. Only Bulu-type Indonesian lowland rice with soil-surface roots have this haplotype, suggesting that soil-surface root occurs due to the presence of this rare allele of *qSOR1*. Using a near-isogenic line having a non-functional allele of *qSOR1*, we demonstrated that QTL associated with soil-surface rooting as opposed to deep rooting could have the potential for genetic improvement of RGA to enhance salt avoidance.

W858: QTL Cloning

Machine Learning Approaches to Target Functional Non-Coding Variation

Maria Katherine Mejia Guerra[1], Karl A. Kremling[1] and Ed Buckler[2], (1)Cornell University, Ithaca, NY, (2)USDA-ARS, Cornell University, Ithaca, NY

In Maize and others organisms, the non-genic portion of the genome hosts the vast majority of sequence polymorphism that are genetic determinants of complex phenotypes (as determined from GWAS). Non-genic variants identified from GWAS frequently overlap with (1) variants statistically associated to changes in gene expression (*i.e.*, eQTLs), and (2) genomic regions biochemically characterized as "functional" with a likely role in regulation of gene expression. However, it is an open challenge to determine/predict which non-coding variants are causative mutations from sequence inspection. Here we present a computational framework to model the importance of sequence features (*i.e.*, *k*-mers) that are characteristic of functional biochemical regions and eQTLs clusters - regulatory regions. Our framework consist in two complementary models adapted from the field of natural language processing (1) "bag-of-*k*-mers" which learn differential weighting of *k*-mers, and (2) "vector-*k*-mers", that captures relationships between *k*-mers. The resulting models are successful to distinguish between regulatory and non-regulatory regions with an accuracy above 90%. The versatility of the models make then suitable for tasks such as *de novo* annotation of genomic regions, prediction of regulatory vocabularies, and prediction of the non-coding-variant effects *de novo* from sequence with base pair resolution.

W859: QTL Cloning

Leveraging the Tetraploid Wheat Genomes for Cloning *Cdu-B1*, a Major Gene for Cd Accumulation in Durum Wheat Grain

Sean Walkowiak[1], Neil Harris[2], Gregory Taylor[2], Krystalee Wiebe[1], P.R. MacLachlan[1], Luigi Cattivelli[3], Marco Maccaferri[4], Assaf Distelfeld[5], Andrew G. Sharpe[6], Ron E. Knox[7] and Curtis J Pozniak[1], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)University of Alberta, Edmonton, AB, Canada, (3)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (4)University of Bologna, Bologna, Italy, (5)Tel Aviv University, Tel Aviv, Israel, (6)Global Institute of Food Security, Saskatoon, SK, Canada, (7)Agriculture and Agri-Food Canada, Swift Current, SK, Canada

Cadmium (Cd) accumulation in the grain of durum wheat presents a serious concern for human health. As a result, durum wheat breeding programs select for low grain Cd. Differences in Cd accumulation among cultivars of durum wheat are attributed to the major-effect gene *Cdu-B1* located on chromosome 5B. The objective of this study was to identify the functional determinant of *Cdu-B1*. The fine mapped interval for *Cdu-B1* was anchored to the complete genome sequences of the durum cultivar ‘Svevo’ (a high Cd accumulator) and the wild emmer wheat accession ‘Zavitan’ (a low Cd accumulator). A sequence comparison of *Cdu-B1* between Svevo and Zavitan revealed a gene candidate, *HMA3-B1*. This gene encodes a P_{1B}-ATPase transition metal transporter and contains a 17 bp duplication in the first exon in Svevo relative to the wild-type allele in Zavitan. A molecular marker for the 17 bp duplication was used to evaluate a diverse set of breeding lines from global breeding programs and was able to identify low and high Cd accumulators with perfect precision. Furthermore, functional assays using yeast expression systems confirm a role for the wild-type *HMA3-B1* gene in regulating Cd accumulation in grain by mediating vacuolar Cd sequestration. In addition, the 17 bp duplication allele present in high Cd genotypes was non-functional. The molecular marker developed from this work is currently deployed in global breeding programs to develop wheat lines with low grain Cd.

W860: QTL Cloning

Finding Invisible QTL Using Missing Data: Examples from a Strongly Restructured Crop Genome

Iulian Gabur, Department of Plant Breeding, Justus Liebig University Giessen, Germany, Giessen, Germany, Regine Delourme, UMR APBV, Le Rheu, France, Andreas von Tiedemann, Department of Crop Sciences, Plant Pathology and Crop Protection Division, Georg August University, Goettingen, Germany, Sebastien Faure, BIOGEMMA, Mondonville, France, Christophe Jestin, Terres Inovia, Thiverval-Grignon, France, Frank Breuer, KWS Saat AG, Einbeck, Germany, Emmanuelle Dyrszka, Syngenta, Saint-Sauveur, France, Rod Snowdon, Justus Liebig University, Giessen, Germany and Christian Obermeier, Plant Breeding, Giessen, Germany

Oilseed rape (*Brassica napus*) is a recent allopolyploid crop species arising from the hybridization between two diploid progenitor species *B. oleracea* and *B. rapa*. Comparative analysis of structural organization and allelic diversity associated with resistance factors to important fungal oilseed rape diseases was performed within the GeWiDis consortium (“Exploiting genome wide diversity for disease resistance improvement in oilseed rape”).

Presence/absence variation (PAV) exists in diploid and polyploid plant species and commonly applied filtering approaches for SNP array data often remove a large proportion of markers from analysis. We describe a strategy for quality-filtering of SNP probes showing increased frequencies of failed calls in multi-parental mapping populations which allows to recover valid InDel SNPs. These InDel SNPs were localized within small to large-scale deleted chromosomal regions and confirmed by whole genome resequencing and optical mapping of the parents. GWAS analysis which includes the recovered InDel markers allowed to detect so far invisible QTL and indicate that structural variation strongly influences quantitative disease resistance in the allopolyploid *B. napus* genome.

W861: QTL Cloning

Mutations in the Branched Head Homoeo-Allele *Bht-B1* Modify Inflorescence Architecture in Tetraploid Wheat

Gizaw M Wolde, Leib-Inst Plant Gen & Crop Plant Res, OT Gatersleben, Stadt Seeland, Germany

Inflorescence morphology directly affects the reproductive success and yield of crops. The wheat inflorescence, also known as spike, forms an unbranched inflorescence where individual spikelets are arranged distichously on the central axis of the spike, the rachis. Previously, we reported the causative mutation in the *branched head*^{t} (*bh*^{t}) gene of tetraploid wheat (*TtBH-A1*) being responsible for the loss of spikelet meristem identity, converting the non-branching wheat spike into a branched spike. Since spike-branching in wheat is a quantitatively inherited trait, we further performed whole-genome quantitative trait loci (QTL) analysis and Genome Wide Association Scans (GWAS) based on 146 recombinant inbred lines (RILs) and a collection of 302 tetraploid wheat accessions, respectively. Results showed that besides the previously found gene at the *bh*^{t}*-A1* locus on the short arm of chromosome 2A, mutations in the homoeologous gene, *TtBH-B1*, was linked to the increased penetrance and expressivity of the supernumerary spikelet (SS) and /or mini-spike formation during spike-branching thereby increasing spikelet and grain number per plant. Furthermore, we developed *bh*^{t}*-A1* Near Isogenic Lines (*bh*^{t}*-A1*-NILs) using an elite durum wheat cultivar, Floradur, for the molecular genetic dissection of the wheat spike morphogenesis and the agronomic implications of the homoeo-allele(s) for increasing grain yield production in wheat.

W862: QTL Cloning

TaZIM1, a Major QTL and Novel Negative Regulator of Heading Date and Kernel Weight, Experienced Strong Selection during Wheat Breeding

Xueyong Zhang, Hong Liu, Tian Li, Yamei Wang, Jun Zheng, Huifang Li and Chenyang Hao, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

Heading date is a critical determinant of regional adaptation for crops and it has a significant impact on crop yield. We identified an atypical GATA-like transcription factor, *TaZIM1* as a negative regulator of wheat heading. We showed that TaZIM1 possesses a weak transcription repression activity and its CCT domain functions as the major inhibitory region. Expression of *TaZIM1* demonstrated a typical circadian clock oscillation pattern under different illumination conditions. Overexpression of *TaZIM1* in wheat causes a delay in the heading date and a decrease of thousand kernel weight (TKW) under long-day conditions. Moreover, TaZIM1 can directly bind to the promoters of *TaCO1* and *TaFT* and down-regulate their expression. Sequence analysis of common wheat cultivar collection identified three and two haplotypes for *TaZIM1-7A* and *TaZIM1-7B*, respectively. Association analysis revealed that *TaZIM1-7A-HapI*/*Hap-III* and *TaZIM1-7B-HapI* have undergone strong positive selection during modern breeding probably due to their association with earlier heading and higher TKW. We developed diagnostic markers for these haplotypes, which can be utilized in further improvement of wheat cultivars via marker-assisted breeding.

Key words: *TaZIM1*, negative regulator, heading date, kernel weight, diagnostic markers

W863: Quinoa and close relatives

An Updated Chromosome-Scale Assembly of Quinoa using Hi-C

David E Jarvis, Brigham Young University, Provo, UT

W864: Quinoa and close relatives

Towards Understanding Salt Tolerance in Quinoa

Heng Zhang, Shanghai Center for Plant Stress Biology, CAS, Shanghai, China

W865: Quinoa and close relatives

QTL-Mapping of Grain Yield and Biomass Production under Low and High Salt Stress under Field Conditions

Eibertus N. van Loo, Plant Breeding, Wageningen University & Research, Wageningen, Netherlands

W866: Quinoa and close relatives

Peroxisome Proliferation in Response to Heat and Drought Stress in Quinoa

Leonardo A Hinojosa[1], Marwa N.M.E Sanad[2], Andrei Smertenko[3] and Kevin Murphy[1], (1)Washington State University, Pullman, WA, (2)National Research Center, Giza, Egypt, (3)Institute of Biological Chemistry, Washington State University, Pullman, WA

The role of peroxisomes in adaptation to abiotic stresses remains poorly understood. Peroxisome proliferation are known to correlate with tolerance to drought, salinity, and heavy metals. In this work, we measured the dynamic of peroxisome proliferation in response to heat, drought and a combination of both stresses in quinoa. Quinoa is an Andean crop with known adaptations to marginal agro-environments, such as soil salinity and drought. Hence, quinoa represents an excellent model system for understanding the role of peroxisomes in regulating ROS-homeostasis under abiotic stress. Eight genotypes of quinoa were grown in greenhouse conditions 25/19 °C (day/night). At the beginning of flowering, irrigation was stopped and the temperature was increased to 35/30°C (day/night) for 5 days. Chlorophyll fluorescence, stomatal conductance, hydrogen peroxide content in leaves, and yield were evaluated as indicators of stress. We observed reduction of stomatal conductance and yield in all genotypes under all types of stress. However, in drought and drought × heat combined stress, chlorophyll fluorescence was reduced and the hydrogen peroxide content was higher. Peroxisome abundance increased in response to all types of stress. The greatest peroxisome proliferation under heat stress were observed in the genotype Pison, whereas in drought or drought × heat stress combination caused greatest peroxisome proliferation in the genotype BGQ 352. Principal component analysis (PCA) demonstrated that hydrogen peroxide and peroxisome abundance were negatively correlated with yield. Our results demonstrate the suitability of peroxisome abundance phenotyping as a parameter for breeding quinoa varieties with higher yield under abiotic stress.

W867: Quinoa and close relatives

Chenopodium Germplasm Resources in Northern New England (NNE), and the Potential to Re-Domesticate Locally Adapted *C. berlandieri*
Thomas Davis, University of New Hampshire, na, NH

W868: Quinoa and close relatives
A PacBio and Hi-C Based Proximity-Guided Assembly of *Chenopodium pallidicaule*
Hayley Hansen, Brigham Young University, Provo, UT
Chenopodium pallidicaule, known commonly as kaniwa or cañahua, is a crop grown in the Altiplano of Peru and Bolivia. It is an A genome diploid (2n = 2x = 18) relative of the allotetraploid (AABB) *Chenopodium quinoa,* and shares its nutritional benefits. Both species contain a complete protein, have a low glycemic index, and offer a wide variety of vitamins and minerals. An initial assembly was created using Illumina mate-pair and paired-end sequences with the ALLPATHS assembler. This assembly contained 3,015 scaffolds with an N50 of 356,353bp. This sequence was improved by *in vivo,* Hi-C based, proximity-guided assembly. The Hi-C data produced an assembly with 623 scaffolds and an N50 of 35,641,356bp. Over 95% of the assembly was found in the 9 longest scaffolds, which represent the haploid chromosomes of *C. pallidicaule*. PacBio long reads at 17x coverage were used to gapfill. 90% of N's from the proximity-guided assembly were filled. Finally, the assembly was polished with Quiver and Pilon. The final assembly was compared to assemblies of *C. quinoa, A. hypochonriacus,* and *B. vulgaris* and the expected relationships were observed. This work has established a genetic framework that will allow for future development of a *C. pallidicaule* as a highly nutritious grain crop in addition to further development of *C. quinoa*.

W869: Quinoa and close relatives
Comparative Genome of Repeat Elements Among Selected Species under Caryophyllales
Dr. Shubham Dixit, Institute of Bioinformatics and applied biotechnology, Bangalore, India

W870: Resources and Programs for Undergraduate Education in Genomics
Session and Speaker Introductions
Scott Woody, UW-Madison, Madison, WI
Brief introductory remarks to welcome attendees and to introduce speakers featured in the 2018 Genomics Education session

W871: Resources and Programs for Undergraduate Education in Genomics
STEMM: Sequencing Technology Education using Microbial Metagenomes
Elizabeth Dinsdale, San Diego State University, San Diego, CA
Since the initial publication of the human genome in 2001, sequence acquisition platforms have rapidly advanced, enabling higher throughput at dramatically lower costs. These developments have allowed sophisticated and powerful genomic research programs and the use of genomic data in personalized medicine. While the application of genomic approaches is broad and pervasive in academic and commercial pursuits, training of the next generation of practitioners has lagged. As one step to remedy that circumstance, we have developed a hands-on, inquiry-based course in metagenomics. Students sample a microbial community of interest, filter and extract genomic DNAs of bacterial, archaeal, and viral inhabitants of that niche, construct libraries and sequence the metagenomes to survey the range and nature of microorganisms present. Assembly and annotation of gDNA scaffolds are at the core of the exercise. Our approach is amenable to all "Next-Gen" sequencing platforms, and we have exploited the capacity of the Nanopore Minion technology since it enables sample collection and sequence acquisition in the field. The experiments are student designed, thereby providing a meaningful biological and personal context that motivates students. To date, students in our metagenomics course have surveyed and characterized microbial communities affiliated with the marine environment including, crabs, snails, shrimps, anemones, fish, sharks, and kelp. Pre- and post-activity attitudinal surveys were used to assess student learning outcomes and indicate increases in both content knowledge and confidence in their ability to conduct legitimate scientific research. Students comment that this is the best course they at taken and have subsequently changed their career goals.

W872: Resources and Programs for Undergraduate Education in Genomics
Educational Workflows in Metagenomics: Microbiomes and Environmental (e)DNA
Dave Micklos, DNA Learning Center, Cold Spring Harbor, NY

W873: Resources and Programs for Undergraduate Education in Genomics
NCBI-Style Hackathons: A Collaborative, Goal-Oriented Approach to Training in Bioinformatics and Software Prototyping
Ben Busby, NCBI, Bethesda, MD

W874: Resources and Programs for Undergraduate Education in Genomics
Transcripity Split: Course-Based RNA-Seq Analysis using the Ultrafast Kallisto-Sleuth Pipeline
Ray A. Enke, James Madison University, Harrisonburg, VA

W875: Resources and Programs for Undergraduate Education in Genomics
Introducing Genomics through Genome Annotation: Broadening Access to Course-Based Undergraduate Research Experiences
Sarah C R Elgin, Washington University in St Louis, St. Louis, MO and Genomics Education Partnership, G-OnRamp, and Genomics Education Alliance

The Genomics Education Partnership (GEP; http://gep.wustl.edu) introduces undergraduates to genomics/bioinformatics by engaging them in a research project. Currently students are working to improve the genome sequence quality and annotate the small, heterochromatic "dot" chromosome (F element) and a comparable euchromatic portion of the D element in a group of *Drosophila* species. Our goal is to use phylogenetic footprinting to identify organizational features and regulatory elements that promote gene expression in the heterochromatic (F element) environment. Assessments show that GEP students from very diverse schools learn about genes and genomes, and gain an appreciation of the research process. To broaden the scope of scientific projects undertaken, GEP has partnered with Galaxy (http://galaxyproject.org) to produce G-OnRamp (http://gonramp.org -see for workshops) — a Galaxy server enabling biologists with little/no informatics expertise to generate a UCSC Assembly Hub or a JBrowse/Apollo genome browser for a newly sequenced genome, with sequence similarity, *ab initio* gene predictions, genomic repeats, and RNA-seq evidence tracks. GEP faculty have begun to use the browsers produced by G-OnRamp to study the evolution of biochemical pathways, initially triglyceride production in parasitoid wasps. Given the growing importance of genomics in the life sciences, members of several programs that provide course-based undergraduate research experiences (GEP, GCAT-SEEK, ComGen, Ciliate Genomics Consortium, Genome Solver, CSHL DNA Learning Center, etc.) have proposed the formation of a Genomics Education Alliance https://figshare.com/articles/A_Genomics_Education_Alliance/5197228). We invite other groups interested in sharing resources and participating in outreach to join us. Supported by NSF IUSE 1431407 (GEP) and NIH R25GM119157-02 (G-OnRamp).

W876: Resources and Programs for Undergraduate Education in Genomics
Assessing the Impact of a Cure on Student Success using Quantitative and Mixed Methods Approaches
James Burnette, University of California, Riverside, CA
Biology 20: The Dynamic Genome (DG) is a CURE for first-year undergraduates at UCR. The first half of the course focuses on key concepts in genetic information transfer and genome organization. Students complete three experiments where they learn techniques of PCR, gel electrophoresis, and pipetting to generate data that demonstrate the results of splicing, genome polymorphism, and regulation of transcription and bioinformatics such as BLAST, multiple sequence alignment, and genome browsers. During the second half of the course students complete a guided project from a professor's research. All students who take DG are part of the learning community program at UCR as are the students who take the traditional Biology 5LA lab course. Thus, we have a randomized controlled experimental design to assess the impact of the two lab courses on first-year undergraduates in the life sciences.
The strictly qualitative approach uses data collected by the institution such as grades, math readiness, ethnicity, and other socioeconomic factors. Mining these data strikingly shows that 93% of DG students earn a B (n = 648) or better in the introductory lecture course Biology 5A while only 43% of the control group earn an A or B (n= 809) regardless of math readiness or socioeconomic factors. Secondly, we are using a mixed-methods approach of interviews and surveys to identify which specific components of DG are causing the effect. Results from the pilot survey will be presented.

W877: Rhinoceros Genomics: Assessing genome-wide variation in rhinoceros species
Low-Coverage Genome Sequencing of Black Rhinoceros to Inform Conservation Management
Sergio A Redondo, Stanford University, Stanford, CA

W878: Rhinoceros Genomics: Assessing genome-wide variation in rhinoceros species
The Last Survivors - Genomics of the Critically Endangered Sumatran Rhinoceros
Johanna Elisabet Nyström von Seth, Swedish Museum of Natural History and Stockholm University, Stockholm, Sweden

W879: Rhinoceros Genomics: Assessing genome-wide variation in rhinoceros species
Comparative Genomics of Ecologically Relevant Genes Families in the Sumatran, Black, and White Rhinoceros
Tate Tunstall, San Diego Zoo Institute for Conservation Research, Escondido, CA

W880: Rhinoceros Genomics: Assessing genome-wide variation in rhinoceros species
Toll-like Receptor Diversity in African Rhinoceros Taxa
Alfred L. Roca, University of Illinois at Urbana-Champaign, Urbana, IL

W881: Rhinoceros Genomics: Assessing genome-wide variation in rhinoceros species
A Progress Update on the Rhinoceros Phylogenomics Project
Love Dalén, Swedish Museum of Natural History, Stockholm, Sweden
The rhinoceros phylogenomics consortium is a large-scale international effort that aims to generate high-quality genome sequences from all extant rhinoceros species, as well as several species that became extinct during the Late Pleistocene. This will enable comparative analyses of the biogeographic history and evolution in the Rhinocerotidae, as well as to resolve the systematics of this family. Moreover, high quality reference genomes will be of considerable value for conservation genomics projects that aim to develop tools and datasets to, for example, assess levels of genetic diversity and inbreeding within species, conduct molecular tracking and estimate levels of gene flow among populations. The aim of this presentation is to give a status update on recent efforts to generate denovo genome assemblies from as many species as possible, and high coverage resequencing data from the remaining taxa.

W882: Rice Functional Genomics
Rice Is Not *Arabidopsis*: Rice Chromatin Has Prominent Topologically Associated Domains (TADs), *Arabidopsis* Chromatin Has Not

Chang Liu, ZMBP, Allgemeine Genetik , Universität Tübingen, Tübingen, Germany, Jia-Wei Wang, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology (SIPPE), Shanghai, China and **Dr. Detlef Weigel**, Max Planck Institute for Developmental Biology, Tübingen, Germany
The non-random three-dimensional organization of genomes is critical for many cellular processes. Recently, analyses of genome-wide chromatin packing in the model dicot plant *Arabidopsis thaliana* have been reported. At a kilobase scale, the *A. thaliana* chromatin interaction network is highly correlated with a range of genomic and epigenomic features. Surprisingly, Topologically Associated Domains (TADs), which appear to be a prevalent structural feature of genome packing in many animal species, are not prominent in the *A. thaliana* genome. Using a genome-wide chromatin conformation capture approach, Hi-C, we have determined high-resolution chromatin packing patterns of rice. We found new structural features of chromatin organization at both chromosomal and local levels compared to *A. thaliana*, with thousands of distinct TADs that cover about a quarter of the rice genome. The rice TAD boundaries are associated with euchromatic epigenetic marks and active gene expression, and enriched with a sequence motif that can be recognized by plant-specific TCP proteins. In addition, we report chromosome decondensation in rice seedlings undergoing cold stress, despite local chromatin packing patterns remaining largely unchanged. The substantial variation found already in a comparison of two plant species suggests that chromatin organization in plants might be more diverse than in multicellular animals.

W883: Rice Functional Genomics
Improvement of Panicle Branching and Grain Yield in Rice
Shuansuo Wang, Qian Liu, Shan Li, Jianqing Zhang, Jianfeng Chen, Kun Wu and **Xiang dong FU**, State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China, Beijing, China
In the past 50 years, the Green Revolution based on the adoption of semi-dwarf cereals which had an increased lodging-resistance under high levels of nitrogen fertilization, was able to achieve crop-yield increases. However, introduction of the semidwarf genes into rice and wheat caused the reduction of panicle (or ear) branching and nitrogen use efficiency. Here we show that a quantitative grain yield trait locus *DEP1* is associated with the increases in panicle branching and nitrogen use efficiency (NUE) in rice. The *DEP1* gene encodes a non-canonical Gγ subunits, loss-of-function mutation at the *DEP1* locus exhibits a decreased number of grains per panicle, whereas gain-of-function mutation confers an increased number of grains per panicle and improved NUE. We also found that the DEP1 protein physically interacts with both Gα (RGA1), and Gβ (RGB1) subunits, and reduced RGA1 or enhanced RGB1 activity represses nitrogen-mediated growth responses. To uncover the signaling mechanisms downstream of the G protein-mediated regulation of panicle branching, we performed both genetic screening to identify the *sod* (suppressor of *dep1*) mutants and yeast two-hybrid screening to identify DIPs (DEP1-interacting proteins). Further experiments showed that Gβγ dimer physically interacts with transcription factor DIP1, the DEP1-DIP1 interactions could enhance its transcriptional activity, and consequently promote co-operative transactivation of common target genes. Our findings reveal a framework for transcription factor DIP1-dependent control of seed number and size in response to G protein-coupled extracellular signals. The manipulation of G-proteins represents new strategies to increase yield potential in crops.

W884: Rice Functional Genomics
Genomic Dissection and Prediction of Transcriptome Dynamics under Fluctuating Field Conditions
Atsushi J. Nagano[1], Makoto Kashima[1], Ayumi Tezuka[1], Ayumi Deguchi[1], Koji Iwayama[2] and Hiroki Saito[3], (1)Faculty of Agriculture, Ryukoku University, Otsu, Japan, (2)Center for Data Science Education and Research, Shiga University, Hikone, Japan, (3)Graduate School of Agriculture, Kyoto University, Kyoto, Japan
Detailed molecular mechanisms of plant environmental responses have been revealed by laboratory experiments. However, it is not enough to understand plant envrionmental responses under field condtions. To bridge between laboratory and field, we developed statistical models using extensive transcriptome data of rice leaves in the field and the corresponding meteorological data. We showed that the transcriptome dynamics of rice leaves in a paddy field were mainly governed by ambient temperature and circadian clock. The statistical model successfully predicted field transcriptome dynamics in Nipponbare, a standard cultivar. However, the effect of genomic buckground on transcriptome dynamics is not known. To establish a method of predicting the transcriptome dynamics in various genomic backgrounds, we developed several novel technologies such as automated parallel preparation of RNA-Seq library, a fast and easy-to-use R library for statistical modeling. Using these technologies, we obtained RNA-Seq data of 1,300 field samples of chromosomal substitution lines between Koshihikari and Takanari, and analyzed the transcriptome data with the corresponding meteorological and genotype data. Two statistical models describing transcriptome dynamics in Koshihikari and Takanari, respecively, were developed. We detected 2,911 genes with different expression dynamics between two cultivars by comparing with predicted transcriptome dynamics of Koshihikari and that of Takanari. Genomic regions controlling the expressional differences were successfully identified in approxmately half of the genes with different expression dynamics. Our models enable the prediction of field transcriptome dynamics not only in two parental cultivars but also in their progeny lines.

W885: Rice Functional Genomics
Role of WRKY Factors in Rice Response to Abiotic Stresses
Antonio Costa De Oliveira, Universidade Federal de Pelotas, Pelotas-RS, RS, Brazil

W886: Rice Functional Genomics
A Novel Transcription Factor in Rice Modulates Broad-Spectrum Blast Resistance through Regulation of H2O2 Accumulation
Xuewei Chen, Sichuan Agricultural University, Chengdu, China

W887: Rice Functional Genomics

Genome Sequencing and Comparative Analysis of the Early Flowering Rice Variety Kitaake

Rashmi Jain[1], Guotian Li[2], Jerry Jenkins[3], Shengqiang Shu[4], Dario Copetti[5], David Kudrna[6], Mawsheng Chern[7], Nhan T. Pham[8], Tong Wei[9], Joel Martin[4], Phat Duong[10], Wendy Schackwitz[4], Anna Lipzen[4], Eda Karaagac[1], Randy Ruan[11], Kerrie W. Barry[4], Rod Wing[5], Jeremy Schmutz[3] and Pamela Ronald[12], (1)University of California, Davis, Davis, CA, (2)oint BioEnergy Institute, Emeryville, CA, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)DOE Joint Genome Institute, Walnut Creek, CA, (5)Arizona Genomics Institute, University of Arizona, Tucson, AZ, (6)Arizona Genomics Institute, Tucson, AZ, (7)UC Davis/JBEI, Davis, CA, (8)UC-Davis, Davis, CA, (9)University of California Daivs, davis, CA, (10)UC-Davis, DAVIS, CA, (11)University of California Davis, Davis, CA, (12)Joint BioEnergy Institute (JBEI), Emeryville, CA

Kitaake, an early flowering rice cultivar, is emerging as a preferred model for rice genetic analysis. Compared with the *Japonica* rice variety, Nipponbare, and the *Indica* rice variety, 93-11, the Kitaake rice variety has a smaller stature, much shorter life cycle (ca. 9 weeks vs. 6 months) and is easier to transform. To facilitate the use of Kitaake, we conducted a whole genome assembly of Kitaake using PacBio sequencing supplemented with Illumina sequencing. The estimated genome size of Kitaake is 377 Mb consisting of 33 scaffolds. We annotated high quality 35,596 non-transposable element protein coding genes in Kitaake and identified genetic variations between Kitaake and Nipponbare. There are 265,377 SNPs and 66,329 INDELs (<30 bp) comparing with Nipponbare. The genomic variations identified in this study are important for future functional genomics studies of important rice traits, such as early flowering of Kitaake. The Kitaake genome is available online at the JGI Phytozome. The Kitaake genome and the previously established whole-genome sequenced mutant population provide powerful resources in rice functional genomics studies.

W888: Root Genomics

Linking Growth Dynamics and Architecture in Rice Root Development

Kevin Lehner, Benfey Lab; Duke University Program in Genetics and Genomics, Durham, NC

W889: Root Genomics

Genomic and Phenomic Approaches to Understand Root Growth and Development for Alfalfa Improvement

Silvas J. Prince[1], Tim Hernandez[1], Nadim Tayeh[1], Rokebul Anower[1], Deborah Samac[2], Elison Blancaflor[1], Christy Motes[1] and Maria J. Monteros[1], (1)Noble Research Institute, Ardmore, OK, (2)United States Department of Agriculture, St.Paul, MN

The plant root system architecture (RSA) affects the capacity for nutrient and water uptake. Certain root traits are favorable under specific abiotic stress conditions and can enhance the performance, productivity and persistence of perennial legume forage species including alfalfa. The objectives of this study were to identify key root traits that result in increased biomass production in alfalfa and to discover the underlying genes for these traits. Field-based evaluations were used to select three alfalfa populations contrasting for branched vs. tapped roots that were part of a genome-wide association study. The RSA of alfalfa seedlings from two additional segregating populations were used to identify specific root traits associated with increased biomass production. The individuals from all three populations were genotyped using genotyping-by-sequencing (GBS) to identify quantitative trait loci (QTL) for root growth and to determine shifts in allele frequencies in key genes. Differences in the root anatomical traits, cortex cell file number and size were also identified using microscopy-based analyses. Further field-based evaluations of alfalfa populations with contrasting root traits planted at different planting densities and grown under irrigated vs. rain-fed conditions provided additional insights on the beneficial root ideotypes to increase alfalfa productivity under different management strategies. The identification of the genetic determinants underlying beneficial root traits can facilitate the implementation of genomics-based approaches in alfalfa breeding programs to increase productivity and persistence.

W890: Root Genomics

Small RNA Profiling of Cavendish Banana Roots in Response to the Inoculation of *Fusarium oxysporum f.sp. cubense*

Shulang Fei, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia

W891: Root Genomics

Root Nodule Transporters and their role in Nitrogen Fixation and Legume Productivity

Mechthild Tegeder, Washington State University, Pullman, WA

W892: Root Genomics

***VERNALIZATION1* Modulates Root System Architecture in Wheat and Barley**

Lee Hickey, Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

As the primary interface for resource acquisition, plant roots play a key role in growth regulation. Evidence from rice, maize and sorghum demonstrates that the below-ground plant architecture significantly impacts plant performance under abiotic constraints. Roots assume critical functions in water uptake, nutrient acquisition and anchorage, an essential characteristic to maintain plant stability under increased grain load. Despite their fundamental importance, knowledge about genetic control of root growth in major grain crops is limited and very little is known about interactions between below-ground and above-ground plant development. Here we demonstrate that *VERNALIZATION1* (*VRN1*), a key regulator of flowering behavior in cereals, also modulates root architecture in wheat and barley. Associations of *VRN1* haplotypes to root growth habit were discovered in wheat by genome-wide association studies, and confirmed by allelic analyses in wheat and barley populations. Functional characterization in transgenic barley confirmed that *VRN1* influences root growth angle directly, via gravitropism. These discoveries provide unexpected insight into underground functions of a major player in the well-characterized flowering pathway, revealing the intersection of above-ground gene regulation with the largely unexplored genetic architecture of plant root development. Understanding the pleiotropic involvement of this key developmental gene in overall plant architecture will help to breed cereal cultivars adapted to specific environmental scenarios.

The full paper is available open access in Molecular Plant: http://www.cell.com/molecular-plant/fulltext/S1674-2052(17)30304-0

W893: Root Genomics
Genetic Analysis of Root Fungal Disease Resistance in Soybean
Mariola Klepadio, University of Missouri, Columbia, MO

W894: Sequencing Complex Genomes
A Novel and Efficient Approach for Phasing Highly Heterozygous Plant Genomes
Tingting Zhu[1], Le Wang[1], Frank M. You[2], Juan C. Rodriguez[1], Karin R. Deal[1], Limin Chen[1], Jie Li[3], Sandeep Chakraborty[1], Bipin Balan[1], Cai-Zhong Jiang[1,4], Patrick J. Brown[1], Charles A. Leslie[5], Mallikarjuna Aradhya[6], Abhaya M. Dandekar[1], Daniel A. Kluepfel[4], Jan Dvorak[1] and Ming-Cheng Luo[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Agriculture and Agri-Food Canada, Morden, MB, Canada, (3)Genome Center, University of California, Davis, Davis, CA, (4)USDA-ARS Crops Pathology and Genetic Research Unit, Davis, CA, (5)Department of Plant Science, University of California, Davis, Davis, CA, (6)USDA National Clonal Germplasm Repository, Davis, CA
Assembling highly heterozygous plant genomes from short sequence reads is challenging due to difficulty in recovering the different haplotypes. Standard assembly protocols tend to collapse homozygous regions and report heterozygous regions as alternative contigs; such multiple assemblies are hard to resolve leading to fragmented assemblies larger than the expected size. We devised a novel method that overcomes genome heterozygosity by assembling two haploid genomes of an interspecific hybrid. Here we report the *de novo* assembly of two haploid genomes in interspecific hybrid MS1-56 (*Juglans regia* cv. Serr × *Juglans microcarpa*). We used a combination of BioNano genome (BNG) mapping, PacBio single-molecule real-time (SMRT) and Illumina sequencing technologies along with standard and custom designed assembly protocols to achieve complete assembly of two haploid genomes (*J. regia* and *J. microcarpa*) comprising the genome of hybrid MS1-56. By coupling SMRT sequencing and BNG mapping technologies, we were able to generate a 1.07 Gb highly contiguous assembly, with a contig N50 size of 8.0 Mb and a scaffold N50 size of 34.8 Mb. We also constructed BNG maps for both parental species of MS1-56 and successfully partitioned the two haplotypes from the sequence assembly of MS1-56, i.e. 529 Mb for *J. regia* 'Serr' and 538 Mb for *J. microcarpa*, respectively. We then applied the genetic map of *J. regia* cv. Chandler onto each assembled genome, resulting in 532 Mb scaffolds in *J. regia* 'Serr' and 524 Mb scaffolds in *J. microcarpa* anchored onto 16 chromosomes in each genome, of which 12 and 14 chromosomes in *J. regia* 'Serr' and *J. microcarpa*, respectively, were able to be resolved into single scaffolds. After gap closing, the total number of N's dropped to 0.76% in *J. regia* 'Serr' and 0.82% in *J. microcarpa*. Characterization of the repetitive portion of the two genomes revealed over 350,000 transposable elements in both genomes. In addition, approximately 31,000 and 29,000 evidence-supported genes were predicted in the *J. regia* 'Serr' and *J. microcarpa* genomes, respectively. To date, this work presents the most contiguous and complete genome assembly of a highly heterozygous plant species. It should also be noted that high-quality haplotype genomes for both parental species were generated from a single sequencing of one hybrid offspring.

W895: Sequencing Complex Genomes
Genome and Evolution of the Shade-Requiring Medicinal Herb *Panax ginseng*
Tae-Jin Yang[1], J. Murukarthick[2] and Nam-Hoon Kim[2], (1)Seoul National University, Seoul, South Korea, (2)Seoul National University, Seoul, Korea, Republic of (South)
Panax ginseng C. A. Meyer, reputed as the king of medicinal herbs, has slow growth, long generation time, low seed production, and complicated genome structure that hamper its study. Here, we unveil the genomic architecture of tetraploid *P. ginseng* by *de novo* genome assembly, representing 2.98 Gbp with 59,352 annotated genes. Resequencing data indicated that diploid *Panax* species diverged in association with global warming in Southern Asia, and two North American species evolved via two inter-continental migrations. Two whole genome duplications (WGD) occurred in the family Araliaceae (including *Panax*) after divergence with the Apiaceae, the more recent one contributing to the ability of *P. ginseng* to overwinter, enabling it to spread broadly through the Northern hemisphere. Functional and evolutionary analyses suggest that production of pharmacologically important dammarane type ginsenosides originated in *Panax* and are produced largely in shoot tissues and transported to roots; that newly evolved *P. ginseng* fatty acid desaturases increase freezing tolerance; and that unprecedented retention of chlorophyll a/b binding protein genes enables efficient photosynthesis under low light. A genome-scale metabolic network provides a holistic view of *Panax* ginsenoside biosynthesis. This study provides valuable resources for improving medicinal values of ginseng either through genomics-assisted breeding or metabolic engineering. **Acknowledgement** - This work was supported by the grant funded by Next-Generation BioGreen21 Program (No. **PJ01103001, PJ01100801**), Rural Development Administration, Republic of Korea.

W896: Sequencing Complex Genomes
The Coconut Genome: Providing a Reference Sequence Towards Coconut Varietal Improvement
Darlon V. Lantican[1], Susan R. Strickler[2], Alma O. Canama[1], Roanne R. Gardoce[1], Lukas Mueller[2] and Hayde F. Galvez[1], (1)University of the Philippines Los Baños, College, Laguna, Philippines, (2)Boyce Thompson Institute, Ithaca, NY
We present the whole genome sequencing (WGS) of Catigan Dwarf (CATD) coconut variety, chosen for its genome simplicity and low heterozygozity. PacBio SMRT sequence data was generated at 15X coverage and corrected with assembled 50X Illumina paired-end MiSeq reads. Through the hybrid assembly approach, the draft assembly of the dwarf coconut genome has N50 of 120 kb. The genome was further improved through Dovetail Chicago sequencing. The input de novo hybrid assembly, Miseq PE reads and Chicago library reads were used as input data in the HiRise pipeline in order to scaffold the genome assembly. As a result, the final assembly has now a total sequence length of 2.1 Gb consisting of 8,062 scaffolds with N50 value of 569 kb. This covers around 97.6% of the estimated CATD genome of 2.15 Gb based on the homozygous k-mer peak. Around 1.556 Gb is identified as interspersed repeat sequences or 73.93% of the total assembled genome. A total of 35,231 high-confidence gene models are identified using the MAKER annotation pipeline. Result of the BUSCO analysis has revealed a 83.1% completeness of the current genome annotation, and 7.4% fragmented single copy orthologs (USCO) based on 1440 genes in the plant specific OrthoDB database. The assembly statistics and quality evaluation results demonstrate that our current draft scaffold assembly has

covered most of the dwarf coconut and gene units. Such provides a good reference coconut genome for various applications such as re-sequencing projects, gene mining and DNA marker development, and functional annotation/genomics approaches.

W897: Sequencing Complex Genomes
The Sunflower Genome
Nicolas Langlade, Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France

W898: Sequencing Complex Genomes
The Genome Sequences of Peanut (*Arachis hypogaea*), a Segmental Allotetraploid
David Bertioli, University of Georgia, Athens, GA

W899: Sequencing Complex Genomes
A Whole Genome Assembly of Rye (*Secale cereale*)
Timothy Rabanus-Wallace, IPK, Gatersleben, Germany

W900: Sex Chromosomes and sex determination
Improved Genome Assembly of American Alligator Genome Reveals Conserved Architecture of Estrogen Signaling
Edward Stallknecht Rice, UC Santa Cruz, Santa Cruz, CA and **Richard E. Green**, UC Santa Cruz, USA, Santa Cruz, CA; University of California Santa Cruz, Santa Cruz, CA
The American alligator, *Alligator mississippiensis*, like all crocodilians, has temperature-dependent sex determination, in which the sex of an embryo is determined by the incubation temperature of the egg during a critical period of development. The lack of genetic differences between male and female alligators leaves open the question of how the genes responsible for sex determination and differentiation are regulated. One insight into this question comes from the fact that exposing an embryo incubated at male-producing temperature to estrogen causes it to develop ovaries. Because estrogen response elements are known to regulate genes over long distances, a contiguous genome assembly is crucial for predicting and understanding its impact. We present an improved assembly of the American alligator genome, scaffolded with *in vitro* proximity ligation (Chicago) and Hi-C data. We perform RNA sequencing of tissues from American alligator embryos to find genes that are differentially expressed between embryos incubated at male- versus female-producing temperature. Finally, we use the improved contiguity of our assembly along with the current model of CTCF-mediated chromatin looping to predict regions of the genome likely to contain estrogen-responsive genes. We find that these regions are significantly enriched for genes with female-biased expression in developing gonads after the critical period during which sex is determined by incubation temperature. We thus conclude that estrogen signaling is a major driver of female-biased gene expression in the post-temperature sensitive period gonads.

W901: Sex Chromosomes and sex determination
Identification of the Y-Encoded Suppressor of Feminization in Kiwifruit
Takashi Akagi[1,2], Isabelle M. Henry[3], Haruka Ohtani[1], Takuya Morimoto[1], Kenji Beppu[4], Ikuo Kataoka[4] and Ryutaro Tao[1], (1)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (2)JST-PRESTO, Saitama, Japan, (3)Plant Biology and Genome Center, UC Davis, Davis, CA, (4)Faculty of Agriculture, Kagawa University, Kagawa, Japan
Dioecy has evolved independently in multiple plant lineages and the genes involved in this differential development are just starting to be uncovered in a few species. We have used genomic approaches to investigate this pathway in kiwifruits (the genus *Actinidia*). Genome-wide cataloguing of male-specific subsequences, combined with transcriptomic analyses, led to the identification of a type-C cytokinin response regulator as a potential sex determinant gene in this species. Functional transgenic analyses in two model systems, *A. thaliana* and *N. tabacum* indicated that this gene acts as a dominant suppressor of carpel development, prompting us to name it '*Shy Girl* (*SyGI*)'. Evolutionary analyses in a panel of *Actinidia* species revealed that *SyGI* is located in the Y-specific region of the genome and probably arose from a lineage-specific gene duplication event. Comparison with its duplicated autosomal counterpart, and with orthologs from other angiosperms, suggests that the *SyGI*-specific duplication and subsequent evolution of *cis*-elements, resulting in stronger expression in carpels pattern, may have played a key role in the acquisition of separate sexes in this genus.

W902: Sex Chromosomes and sex determination
Initiation and Spread of the Y Chromosome in the Genus *Phoenix*
Maria Fernanda Torres, WCMC-Qatar, Doha, Qatar and **Joel Malek**, Genomics Core, Weill Cornell Medical College in Qatar, Doha, Qatar
The genus *Phoenix* includes the commercially important date palm tree among 13 other species, all of which are dioecious. Members of the genus grow in a tropical and subtropical area spanning from Asia to North Africa.
In our previous work to understand the genetic control of sex-determination in the date palm we identified regions of the genome linked to sex and showed that date palm employs an XY system. However, the sex-linked region spans multiple megabases and identification of the original mutations is challenging given recombination suppression. Others have shown that sex-determination likely arose prior to speciation in the genus. Therefore, to identify potential mutations original to sex-determination in the genus we have sequenced male and female individuals from all species in the genus. We then searched for short sequencers (kmers) linked to sex in all males. This was followed by phased sequencing of regions surrounding the identified kmers. These data identified a focal point containing a small number of genes that appear original to dioecy in the genus. Spread of non-recombination from this focal point was analyzed further in various clades of the genus.
Here we show that phylogenetic analysis of the Y-linked genes combined with genus-wide sequencing allowed us to better understand the spread of recombination suppression in the genus. While all clades in the genus radiate from the same sex-determination locus, each clade

appears to spread in a different manner. Variation in this process and its implications for the development of dioecy in general will be discussed.
This study was made possible by grant NPRP-EP X-014-4-001 from the Qatar National Research Fund (a member of Qatar Foundation).

W903: Sex Chromosomes and sex determination
Identification of RAN1 Orthologue Associated with Sex Determination through Whole Genome Sequencing Analysis in Fig (*Ficus carica* L.)
Kazuki Mori[1], Kenta Shirasawa[2], Hitoshi Nogata[3], Chiharu Hirata[4], Kosuke Tashiro[1], Tsuyoshi Habu[5], Sangwan Kim[1], Shuichi Nogata[6], Satoru Kuhara[1] and **Hidetoshi Ikegami**[6], (1)Kyushu University, Fukuoka, Japan, (2)Kazusa DNA Research Institute, Chiba, Japan, (3)Fukuoka Keichiku Agricultural Extension Center, Yukuhashi, Japan, (4)Fukuoka Agriculture and Forestry Research Center, Chikushino, Japan, (5)Ehime University, Ehime, Japan, (6)Fukuoka Agriculture and Forestry Research Center, Yukuhashi, Japan
As the sexuality of fig affects the syconium edible properties, it is an important factor in fig breeding programs. With the aim of identifying sex determining genes of fig, we generated the first draft genome sequence of fig and applied map based cloning approaches. Linkage analysis with a high-density genetic map established by a restriction-site associated sequencing technique, and genome-wide association study followed by whole-genome resequencing analysis identified two missense mutations in RESPONSIVE-TO-ANTAGONIST1 (RAN1) orthologue encoding copper-transporting ATPase suggesting a link to ethylene perception mechanism. This result suggests that RAN1 is a possible sex determinant candidate in the fig genome. In addition, we have confirmed the effect of gibberellin on flower sex and the adjustability of a RAN1 mutation based marker to another ficus species. Here, we propose a model of the mechanism and the evolution process of sex determination in the genus ficus.

W904: Sex Chromosomes and sex determination
The Evolution from Autosome to X and Y Chromosomes in Asparagus
Alex Harkess, Donald Danforth Plant Science Center, St. Louis, MO
An elegant model for the conversion of an autosomal pair in a hermaphroditic species to a sex chromosome in a dioecious species was formalized by Deborah and Brian Charlesworth in 1978. Briefly, the conversion from autosome to sex chromosome could require just two loci linked perfectly in non-recombination on a young Y chromosome: one locus must dominantly suppress female (pistil) organogenesis, while another locus must promote the formation of male (anther) organogenesis, but this model has never had strong genic support in any dioecious species. Garden asparagus (*Asparagus officinalis*) is an ideal system to test this question, given that the X and Y sex chromosomes are evolutionarily young and cytologically homomorphic. Leveraging a doubled haploid YY individual and a doubled haploid mapping population, we generated a chromosome-level genome assembly using Illumina, PacBio, and Bionano optical maps and identified a 1Mb region of non-recombination on the Y chromosome that is largely missing from the X. Several independent male-to-hermaphrodite mutants (gamma irradiation, spontaneous SNP) implicate a single gene *SOFF* (Suppressor of Female Function) on the Y chromosome as being responsible for female suppression. While only 12 gene annotations are in this non-recombining region, there exists *Tapetum Development and Function 1*, a gene with an *Arabidopsis* knockout phenotype very similar to Asparagus females. EMS mutagenesis confirms that *tdf1* knockouts are neuters. With an additional PacBio and optical map genome assembly for a sibling XX female, we have compared the structure of the X and Y chromosomes and identified ~130kb of X-specific sequence, including several X-specific genes. Additionally, an in-progress Oxford Nanopore genome for a hermaphroditic individual is being generated to determine the ancestral autosome sequence. This is the strongest evidence to date for the Charlesworth's "two gene" model.

W905: Sex Chromosomes and sex determination
Investigation of Sex Determining Region in *S. latifolia*
Radim Cegan, Institute of Biophysics ASCR vvi, Brno, Czech Republic
Silene latifolia, dioecious species possessing heteromorphic sex chromosomes, is a model plant to study sex determination, evolution of sex chromosomes and dosage compensation. Three different regions of the Y chromosome containing gynoecium suppressing factor, stamen-promoting factor and male fertility factor play major role in sex determination. Although *S. latifolia* sex chromosomes are being studied extensively, the sex-determining genes have not been identified yet. By combination of genomic (BAC library screening, flow-sorting, microdissection, fluorescence *in situ* hybridization and existing genomic data) and transcriptomic approaches (transcriptome profiling of epigenetically treated plants with hermaphroditic flowers in combination with analysis of existing RNA-Seq databases) we have investigated the structure, gene composition and function of *S. latifolia* gynoecium suppression factor region. In parallel we have developed an efficient and reproducible protocol for genetic transformation and *in vitro* regeneration. Furthermore, we adopted the CRISPR/Cas9 nuclease system to edit the *S. latifolia* genome. Our results show that this system is highly efficient and feasible and to our knowledge, this is the first report of genome editing on plant sex chromosomes. This work was supported by the Czech Science Foundation project No. 17-00567S.

W906: SGN and RTB Databases: Genomics and Breeder Tools.
Breeding Databases and Tools
Lukas Mueller, Boyce Thompson Institute, Ithaca, NY
The Mueller lab develops databases and web tools for breeding data management and analyses. Currently, we have databases for the Root, Tuber and Banana crops (http://cassavabase.org, http://yambase.org, http://sweetpotatobase.org, http://musabase.org). This session will highlight the breeder tools that seamlessly integrate from designing experiments to the data analyses.

W907: SGN and RTB Databases: Genomics and Breeder Tools.
solGS: A Web Tool for Genomic Selection

Isaak Y. Tecle[1], Naama Menda[1], Bryan Ellerbrock[1], Guillaume J. Bauchet[1], Alex C. Ogbonna[2], Nicolas Morales[1], Lukas A. Mueller[1] and Cassavabase Team, (1)Boyce Thompson Institute, Ithaca, NY, (2)National Root Crops Research Institute (NRCRI), Ithaca, NY

Genomic selection, due to its reliance on high-density markers and statistical complexity, presents significant challenges in data management, analysis and sharing results. solGS, a web-based tool, alleviates these challenges; it has a database to store phenotype and genotype data and an intuitive web-interface for statistical analyses. It uses RR-BLUP for the statistical modeling and GBLUP method for breeding values estimation. It performs exploratory analysis, such as descriptive statistics, population structure, phenotypic and genetic correlations, and selection index calculation. It visualizes data in interactive plots. solGS is, currently, used by the NextGen Cassava Breeding Project (http://nextgencassava.org) and implemented on http://cassavabase.org/solgs and other databases (http://yambase.org, http://sweetpotatobase.org, http://musabase.org). GS breeders can adapt the tool for any organism.

W908: SGN and RTB Databases: Genomics and Breeder Tools.

New Interactive Tools on SGN Websites

David Lyon, Boyce Thompson Institute, Cornell University, Ithaca, NY

In this workshop section, a demonstration and tutorial will be presented on a range of new interactive tools available at SGN websites. Participants will be introduced to the the Trial Comparison tool, which can be used to visually compare the performance of different accessions across a set of trials. They will then be guided through using the Graphical Filtering tool, which allows for a list of plots to be filtered with selected ranges from trait measurement distributions. Finally, participants will learn to use the Pedigree Viewer tool, which allows for easy navigation of pedigree information in the database.

W909: SGN and RTB Databases: Genomics and Breeder Tools.

Updates to the Tomato Reference Genome and ITAG Annotation

Prashant S Hosmani[1], Mirella Flores-Gonzalez[1], Henri van de Geest[2], Stephane Rombauts[3], Gabino F. Sanchez Perez[2], Lukas Mueller[1] and **Surya Saha**[1], (1)Boyce Thompson Institute, Ithaca, NY, (2)Plant Research International, Wageningen University & Research, Wageningen, Netherlands, (3)VIB Department of Plant Systems Biology, Ghent University, Gent, Belgium

We describe here our efforts to improve the tomato reference genome and the corresponding annotation. We have integrated 1,069 full-length phase htgs3 BACs into the tomato genome to cover gap regions and replace shorter whole genome shotgun contigs which removed 11.7Mb of contig gaps. BioNano optical maps generated for heinz 1706 largely confirmed the reference genome assembly and excluding some minor structural and homopolymer corrections.

We also present an improved annotation (ITAG 3.2) for the new tomato reference genome. The annotation pipeline involved updating ITAG2.40 gene models, preserving locus identifiers (30,868) from the previous ITAG2.40 annotation and identifying new gene models using newly trained *ab-initio* gene predictors. We have incorporated expression data from various sources in our pipeline including tissue and treatment specific RNAseq data, 5' and 3' UTR enriched RNAseq data, RENseq for NBS-LRR genes and Pacbio RNA Iso-seq data which were all kindly provided by Solanaceae community. The *de novo* prediction of genes in the maker pipeline has identified 4,900 novel genes well-supported by expression evidence. Annotation of all the genes has associated quality metrics for structure and functional characterization. The increased continuity of the SL3.0 build and availability of diverse and comprehensive expression data has resulted in a significantly improved tomato annotation (ITAG3.2).

All data are available through the SOL Genomics Network website (SGN, https://solgenomics.net) and FTP site (ftp://ftp.solgenomics.net/tomato_genome/).

W910: SGN and RTB Databases: Genomics and Breeder Tools.

High-Quality *de novo* Genome Assembly of the Tomato Genome using Combination of PacBio, BioNano and Chromium 10x Long Molecules Sequencing Technologies

Mohamed Zouine, INPT-ENSAT, Castanet-Tolosan, France

Long sequencing technology offer the possibility of dramatically improving the contiguity of genome assemblies and able to extend paths into problematic or repetitive regions. In addition to Long sequencing reads approaches, recent technologies like optical mapping and linked reads from 10X Genomics are capable to bring additional scaffolding to achieve chromosome-level assemblies.

In order to improve the actual tomato reference genome sequence, we generated 70X coverage of Pacific Biosciences (PacBio RSII) long-reads, 100X of optical mapping using two different enzymes and 100X illumina HiSeq3000 2x150b paired-reads sequencing from Chromium linked 10X Genomics libraries.

The integration of these three approaches allowed to reach a genome size of **~830 Mb** with an **N50 of 45 Mb**. The assembly contiguity reached chromosome-arm-levels. Interestingly, one full chromosome (Ch12) has been fully assembled in one scaffold. Some of the remaining scaffolds revealed large parts of some centromeric regions, even including some of the heterochromatic regions. We assessed the quality of contigs and scaffolds using Illumina mate-pair libraries and genetic map information.

The integration of the genetic map allowed to generate the 12 pseudomolecules corresponding to the 12 tomato chromosomes. Several regions corresponding to chromosome zero in the SL3.0 reference genome were included in the current assembly.

We took advantage from the large RNA-Seq data of the TomExpress platform (http://gbf.toulouse.inra.fr/tomexpress/) and use them to annotate this new genome assembly. To assess the completeness of the genome and the transcriptome assembly, busco v3.0 software has been used and shown a high percent of gene coverage (> 96%).

W911: SGN and RTB Databases: Genomics and Breeder Tools.

The Tomato Expression Atlas and Beyond: Visualizing and Interpreting Transcriptomic Data

Adrian Powell and Lukas A. Mueller, Boyce Thompson Institute, Ithaca, NY

With decreasing sequencing costs, transcriptomic gene expression data sets are currently being generated in large numbers. However, the synthesis of gene expression patterns in large data sets, as well as their use in generating further experimental hypotheses, remains challenging. Expression atlas tools assist in addressing such challenges. Using tomato fruit as a model, the Sol Genomics Network (SGN) has developed the Tomato Expression Atlas (TEA) as a web-based tool for visualizing gene expression data. The TEA is available at http://tea.solgenomics.net/; the source code is available at https://github.com/solgenomics/Tea. This implementation of the expression atlas is being extended to a variety of other systems, giving rise to the Cassava Expression Atlas, the Sweet Potato Expression Atlas, and the Psyllid Expression Network. These expression atlases enable exploration of expression in a variety of specific tissues and developmental stages by means of expression cubes, bar graphs, and heatmaps, as well as new visualizations such as scatterplots and networks. With these tools, researchers are able to easily access and interpret gene expression data from different samples in a compact, intuitive form; these tools can thereby provide clues about genes and processes that are important for phenotypes of interest. The SGN expression atlases also allow for data sets to be made publicly available to researchers around the world.

W912: Small RNA

Analysis of Small Non-Coding RNA Profiles Resulting from a Pasteurella multocida Challenge in Cattle

Tasia M. Taxis[1], Fred M. Tatum[2], Robert E. Briggs[2] and **Eduardo Casas**[2], (1)Michigan State University, East Lansing, MI, (2)USDA, ARS, National Animal Disease Center, Ames, IA

Small non-coding RNAs (sncRNAs), such as microRNAs and tRNA-derived fragments (tRFs) have been linked with immune response. The objective was to assess changes in microRNAs and tRFs associated with virulent *Pasteurella multocida* (a bovine respiratory pathogen) challenge in calves previously exposed to isogenic *P. multocida* strains. Three modified *P. multocida* strains were produced by removing one of three genes encoding putative virulence factors. Cattle were intranasal-inoculated with $2x10^9$ cfu/mL as follows: Control (sham inoculation; n=3), Wt (inoculated with wild-type *P. multocida*; n= 4), FhaB2 (n= 4), HyaE (n=4), and NanP (n= 4). Fifty six days later all animals were intratracheally challenged with $2.2x10^9$ cfu/mL of wild type *P. multocida*. ELISAs for immunoglobulin levels and bacterial shedding were assessed weekly and showed significant differences between exposed and sham groups. Small noncoding RNAs were sequenced from sera and white blood cells from all animals before and after intratracheal challenge (days 49 and 61, respectively). A significant interaction of treatment and challenge was detected for bta-miR-150 (P= 0.0025) and $tRF5^{MetCAT}$ (P= 0.0009). For bta-miR-150, Control and FhaB2 counts decreased after challenge, Wt and NanP increased, and no difference was observed for HyaE. For $tRF5^{MetCAT}$, Control, FhaB2, and HyaE counts increased, and Wt and NanP decreased after challenge. While Control and FhaB2 counts decreased for bta-miR-150, they increased for $tRF5^{MetCAT}$. An inverse relationship between bta-miR-150 and $tRF5^{MetCAT}$ was observed. Results indicate that variation in calf response to *P. multocida* with specific genomic modifications leads to differences in host's counts of sncRNAs.

W913: Small RNA

Combinatorial mRNA Regulation by miRNAs and Pumilio

Ted Karginov, University of California - Riverside, Riverside, CA

Regulation of gene expression by ~1500 RNA-binding proteins (RBPs), and their resulting effects on cellular function, are gaining increasing appreciation. Often using sequence-specific binding to 3' UTRs to target coordinate sets of transcripts, RBPs control mRNAs at the level of splicing/maturation, cytoplasmic export, stability and translation. Furthermore, higher-order combinatorial interactions between RBPs on individual mRNAs have been proposed to underpin the regulatory network. We took a global experimental approach followed by targeted validation to assess the extent of interaction between two well characterized and highly conserved RBPs, Argonaute (Ago) and Pumilio (Pum). CLIP-seq identified significant overlap of bound mRNAs, and Ago2 binding site behavior depended on co-occupancy with Pum. In addition, Pum binding sites that overlap with Ago2 showed differential behavior upon Dicer loss. Candidate 3' UTR interaction sites where Ago and Pum are colocalized were selected for functional validation using luciferase reporter assays. We identified 3 sites to host antagonistic interactions between Pum and Ago, where Pum acts to decrease Ago binding. Our evidence indicates that Pum, a protein characterized as mostly repressive, can change its behavior in a context dependent manner. Interestingly, the binding sites for the two proteins are too far for potential antagonism due to steric hindrance, and suggests some alternate mechanism. The importance of this research lies in its elucidation of complex interactions between RBPs, and how extensive these interactions are transcriptome-wide.

W914: Small RNA

RNA Recruitment to Cytoplasmic Germ Granules Triggers Its Processing into Pirnas

Katalin Fejes Toth, California Institute of Technology, Pasadena, CA

The piRNA pathway represses transposable elements in the gonads and thereby plays a vital role in protecting the integrity of germline genomes of animals. Mature piRNAs are processed from longer transcripts, piRNA precursors (pre-piRNA). In Drosophila, processing of pre-piRNA is initiated by piRNA-guided Slicer cleavage, or by the endonuclease Zucchini (Zuc). As Zuc does not have any sequence or structure preferences *in vitro*, it is not known how piRNA precursors are selected and channeled into the Zuc-dependent processing pathway. We show that a heterologous RNA that lacks complementary piRNA is processed into piRNA upon recruitment of several piRNA pathway factors. This processing requires Zuc and the helicase Armitage (Armi). Aubergine (Aub), Argonaute 3 (Ago3), as well as components of the nuclear RDC complex, which are required for piRNA biogenesis in germ cells, are dispensable. Our approach allows discrimination of proteins involved in transcription and export of piRNA precursors from components required for the cytoplasmic processing steps. piRNA processing correlates with localization of the substrate RNA to nuage, a distinct membraneless cytoplasmic compartment, which surrounds the nucleus of germ cells, suggesting that sequestration of RNA to this subcellular compartment is both necessary and sufficient for selecting piRNA biogenesis substrates.

W915: Small RNA

Duplication and Neofunctionalization of Dicer-like Proteins in the Monocots

Alex Harkess, Donald Danforth Plant Science Center, St. Louis, MO

Dicer-like proteins (DCLs) are involved in variety of small RNA-related functions across the angiosperms, spread across five major clades in the gene family (DCL1-DCL5). Whereas 24nt heterochromatic siRNAs are processed by DCL3, a unique class of 24nt meiotic phased siRNAs are instead processed by DCL5. Given that DCL5 is only found in some monocot species, including the grasses, an open question is understanding when and how DCL5 evolved and specialized to process reproductively-enriched phased siRNAs. Using transcriptome-based phylogenomics in collaboration with the Monocot Tree of Life project (MonAToL), we first place a long sought-after polyploidy event "Tau" on the monocot phylogeny. Intriguingly this major polyploidy event likely coincides with the duplication of the ancestral DCL3 to produce DCL5. A pan-monocot analysis of molecular evolution shows distinct signatures of constraint and positive selection on certain residues of the DCL3 and DCL5 proteins, providing targets for understanding how polyploidy, rapid diversification, and neofunctionalization of DCL5 shaped the evolution of reproductively enriched phased siRNAs and their biogenesis machinery in the monocots.

W916: Small RNA
Analysis of Small Non-Coding RNA Profiles Resulting from a *Pasteurella multocida* Challenge in Cattle
Eduardo Casas, USDA, ARS, National Animal Disease Center, Ames, IA

W917: Small RNA
Establishing DNA Methylation in Arabidopsis
Julie Law, Salk Institute, La Jolla, CA

W918: Small RNA
Small RNAs from Diverse Sources Target Reactivated Transposable Elements in *Arabidopsis thaliana*
Kaushik K. Panda, Josquin Daron and R. Keith Slotkin, The Ohio State University, Columbus, OH
Transposable elements (TEs) are mobile genetic units found in all eukaryotic genomes. TE insertions and/or rearrangements cause mutations and DNA damage. To defend their genome, organisms have evolved various silencing mechanisms to repress TE activity. Small RNA-directed DNA methylation (RdDM) is one such silencing mechanism that is well-studied in the reference plant Arabidopsis. Canonically, RdDM functions through RNA Polymerase IV (Pol IV) generation of 24 nucleotide small RNAs (24nt sRNAs) that are incorporated into ARGONAUTE 4 (AGO4) and AGO6 proteins. Recently, non-canonical RdDM pathways have been discovered wherein sRNAs generated from Pol II transcripts (which were thought to silence TEs only post-transcriptionally) also direct DNA methylation.
Multiple non-canonical RdDM mechanisms have been reported in single locus studies; however, only recently have we investigated the role of all known RdDM pathways to silence TEs on the genome-wide level. By analyzing the features of TE targets for both canonical and non-canonical RdDM, I found key differences that distinguish the TE targets: the canonical 24nt sRNA mediated RdDM mainly targets the edges of TEs that are near genes, whereas the non-canonical 21-22nt sRNA mediated RdDM pathway (which utilizes Pol II mRNAs) causes methylation throughout the length of its target TEs. I also determined that full-length TEs, capable of self-transposing and/or catalyzing non-autonomous TE transposition, are preferentially targeted by 21-22nt sRNAs. This preference is driven by the selective cleavage of full-length TE mRNAs, which subsequently generates the secondary 21-22nt sRNAs that drive DNA methylation. The finding demonstrates that chromatin silencing patterns can be reflections of RNA-based degradation specificities. In addition, I discovered a category of Pol IV-independent 24nt sRNAs that also initiate DNA methylation. This pathway has a minor role in TE silencing in Arabidopsis, where TEs are silenced, but a much more significant role in the TE-rich maize genome. Together, these recent discoveries demonstrate that there is a complex network of sRNAs that synergistically target TEs for post-transcriptional degradation with the end result of chromatin modification.

W919: Solanaceae
Cold Storage of Tomato: The Good, the Bad and the Ugly
Raquel Rosales[1], Yury Tikunov[1], Maike Hanenberg[2], Wouter Verkerke[2] and **Arnaud G. Bovy**[1], (1)Wageningen UR Plant Breeding, Wageningen, Netherlands, (2)Wageningen UR Greenhouse Horticulture, Bleiswijk, Netherlands
Cold storage at around 12 ºC is recommended to reduce tomato decay and maintain tomato fruit quality while avoiding a reduction of tomato flavour that can occur when fruit are stored below the recommended temperature. Numerous reports have demonstrated that tomato storage at 5 ºC (household refrigerator temperature) affects the production of important tomato volatiles and a few of them showed that low temperature also affected the sensory evaluation or the consumer liking. However, the number of genotypes tested was very limited. In this study, the individual effects of storage time and low temperature on tomato flavour were evaluated using several modern tomato genotypes that contrast in their organoleptic quality. Fruits stored during one week at chilling (5 ºC) and non-chilling (15 ºC) temperature were compared with fresh harvested fruits using a consumer and/or an expert panel. Changes in fruit quality parameters (texture, soluble solids content, and acidity), primary metabolites, and volatile organic compounds (VOCs) were also measured. Our results indicate that although cold storage did impact the metabolic profile in all genotypes, these metabolic changes were only translated in a small, but significant effect on consumer liking in genotypes with low flavor quality (low levels of sugars, acids and VOCs). Moreover, the sensory data showed that in some cases the effect of storage was larger than the effect of low temperature.

W920: Solanaceae
Morphological and Genetic Diversity in a Collection of Tomato Traditional Cultivars
Richard Finkers, Wageningen UR Plant Breeding, Wageningen, Netherlands
Efficient utilisation of Genetic diversity within Plant Breeding is one of the elements in the ambition to produce twice as much food with twice less input. Technological advances currently drive initiatives to discover genetic and phenotypic diversity in many of the *Solanaceae*. Coupled with whole genome re-sequencing approaches, allelic (haplotype) diversity underlying this phenotypic diversity is becoming more-and-more accessible, even for tetraploid potato. However, these activities also lead to a data management challenge. In the presentation we will explore part of the genotypic and phenotypic diversity in large germplasm collections and present some of the computational algorithms developed, mainly to summarize haplotype diversity of sequence based approaches.

W921: Solanaceae
Specialized Metabolism in Tomato: Understanding the Function of Flavonoid Metabolites in Roots, Guard Cells, and Pollen
Gloria K. Muday, Wake Forest University, Winston Salem, NC

W922: Solanaceae
Rpi-amr1*, a Novel Class of Solanum Resistance Genes against *Phytophthora infestans
Kamil Witek, Hari S Karki, Florian Jupe, Agnieszka I. Witek and Jonathan DG Jones, The Sainsbury Laboratory, Norwich, United Kingdom
R gene enrichment and long-read sequencing (SMRT-RenSeq) is a valid tool for identification and cloning of novel functional NLR-type disease resistance genes in plants. Here we utilize SMRT-RenSeq to clone a novel *Resistance to Phytophthora infestans (Rpi)* gene from the wild *Solanum americanum*. Our approach is reference free, and allele mining identified wide distribution of this new gene.
We identified a hotspot for *P. infestans* resistance in numerous *S. americanum* accessions at the distal end of the short arm of Chr 11. We combined bulked segregant analysis and SMRT-RenSeq to clone *Rpi-amr1e* and show that it confers strong resistance against multiple isolates of *P. infestans* in cultivated potato. The gene encodes a typical coiled-coil (CC) NLR protein; however, it belongs to a previously uncharacterized class of CNL genes. In collaboration with Chih-hang Wu in the Kamoun lab, we found that its function is dependent on the helper NLR NRC. We used SMRT RenSeq and association genomics to clone functional *Rpi-amr1e* alleles from several *S. americanum* accessions. Despite 80-90% amino acid identity between these paralogs, they still confer resistance to *P. infestans*. Moreover, employing targeted enrichment-based genotyping, we found that in one of the lines of *S. americanum, Rpi-amr1e* had been translocated to Chr 1.

W923: Solanaceae
Genome-Wide Prediction in Tetraploid Potato using Pedigree and Marker Information
Jeffrey Endelman, University of Wisconsin, Madison, WI

W924: Solanaceae
Integration of Environmental and Internal Cues for Triggering Vegetative Reproduction in Potato
Christian Bachem, Laboratory of Plant Breeding, Wageningen University & Research, Wageningen, Netherlands

W925: Sorghum/Millet
Advances in Developing Sorghum Hybrids for use as a Feedstock for Bioenergy Production in Brazil
Rafael Parrella, Embrapa National Maize and Sorghum Research Center, Sete Lagoas, Brazil and Robert Schaffert, Embrapa Maize and Sorghum, Sete Lagoas, Brazil
Bioenergy Sorghum is dived into four categories based on transformation vehicle and photoperiod sensitivity. Sweet sorghum is the traditional feedstock with juice extraction rich in sucrose for fermentation (G1-tecnology). Sorghum for ethanol production can be photo-insensitive or photosensitive. Sweet sorghum cytoplasmic male sterile lines (A-lines) have been developed for use in confecting both photo-insensitive and photosensitive sweet sorghum hybrids. These hybrids produce 120 – 150 t ha^{-1} fresh biomass during the long days of summer making sweet sorghum competitive with sugarcane. Biomass sorghum is a photosensitive hybrid with dry stems using G2 technology to transform cellulose and hemicellulose into fermentable sugars or burned to generate electricity (co-generation). Embrapa has developed biomass hybrids with reduced lignin for G2 technology and hybrids with increased lignin with greater calorific value for burning. Random mating sweet sorghum B and R populations using ms_3 have been developed to increase sucrose using recurrent selection. The harvest of sorghum for ethanol production, unlike grains where the product can be stored, is a 24/7 operation where the feedstock must be delivered to the processing mill continuously for several weeks. We developed a "Period of Industrialization – PIU" protocol for selecting sweet sorghum cultivars with high sucrose extraction for a minimum of 30 days, facilitating "Industrial Planning". Sucrose content is highly correlated with longer PIU. The major research challenge in biomass sorghum for co-generation is reducing the water content. Genes for both biotic and abiotic stress tolerance are introgressed into breeding lines and hybrids using both molecular and traditional breeding methods.

W926: Sorghum/Millet
Sorghum for the Great Plains - Integrated Strategies to Enhance Yield Potential and Abiotic Stress Resilience
S.V. Krishna Jagadish, Kansas State University, Manhattan, KS
Early season chilling and late season heat and drought stresses are persistent challenges to enhance sorghum productivity in the Great Plains of the US. Recently, using 400 different hybrids, grown over 30 years in 11 different locations of Kansas, we show that both pre- and post-flowering heat stress have equal and negative impact on sorghum yield. This analysis demonstrated narrow genetic base in US breeding program in terms of resilience to heat stress. For early season chilling, improving germination and emergence index, accompanied by superior seedling vigor will provide opportunities to shift planting dates towards cooler months, which can potentially minimize terminal heat and drought stress damage. Diverse genetic resources including sorghum association panel has been phenotyped under chambers and field conditions, employing both classical physiological and aerial high-throughput phenotyping approaches. Genetic regions inducing greater resilience to chilling stress at vegetative stage have been identified. For terminal heat stress, sorghum's inherent mechanism to minimize heat stress damage during flowering by employing early morning flowering was documented, which calls for revised strategies to address heat stress damage. In addition, grain filling pattern along the gradient of panicles was documented under different environmental conditions which provided important clues on rate and duration of grain filling, importance of stay green, source-sink relationships that can potentially impact the overall yield potential of sorghum. Progress achieved towards enhancing early season chilling tolerance, to minimize terminal heat stress damage and thoughts on enhancing sorghum's yield potential will be discussed.

W927: Sorghum/Millet

The Sorghum Leaf Metabolome and Proteome and its Role in Morpho-Physiology and Drought Tolerance

Courtney E. Jahn[1], Marie F. Turner[1], Sarah B. Miller[1], Adam L. Heuberger[1], Lisa M. Wolfe[1], Edward J. Wolfrum[2], Corey D. Broeckling[1] and Jessica E. Prenni[1], (1)Colorado State University, Fort Collins, CO, (2)National Renewable Energy Laboratory, Golden, CO

Sorghum bicolor is a globally important food and fuel crop with substantial genetic diversity and phenotypic variation in plant architecture, drought-tolerance, and biomass yield and composition. Recent studies have demonstrated that genotype can influence biochemical variation in plants, however, little is known about how variation in sorghum metabolism is associated with diversity in morpho-physiology. Further, variation in plant metabolism influence developmental processes, and thus underpin many of the ways plant respond to changes the environment. Here, the metabolic basis of morpho-physiology was characterized using a metabolomics and proteomics approach for 11 diverse sorghum lines. Metabolite data was collected from leaves using non-targeted and targeted UPLC- and GC-MS profiling, and the data was associated to variation in 21 morphological and physiological traits. Further, changes to the metabolome and proteome were evaluated in a drought stress assay that compared a tolerant and susceptible line, and the leaf metabolite and protein data were linked to above and below ground traits, including a metabolomics analysis of root exudation. Several primary metabolites (e.g. amino acids and purines), secondary metabolites (e.g. phenolics), and proteins (malate dehydrogenase) were found to be associated to traits including photosynthesis, biomass yield and drought tolerance. In addition, we report a preliminary model to predict biomass yields based on leaf metabolite data. Taken together, these results provide a systems-level overview of morpho-physiological traits, biomass accumulation and drought tolerance, and provide candidate phenotypic and metabolic markers to facilitate breeding for enhanced biomass production with improved tolerance to water stress.

W928: Sorghum/Millet

Exploiting Sorghum Genetic and Genomic Resources to Support Dissection of Complex Traits

Doreen Ware[1], Yinping Jiao[2], Young Koung Lee[3], Nicholas Gladman[4], Ratan Chopra[5], John Burke[6], Gloria Burow[6], Chad Hayes[6], Michael Regulski[4], Shawn A. Christensen[7] and Zhanguo Xin[8], (1)USDA/ARS - Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)USDA-ARS/Cold Spring Harbor Laboratory, Lubbock, TX, (3)Cold Spring Harbor Laboratory, COLD SPRING HARBOR, NY, (4)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (5)USDA- ARS, Lubbock, TX, (6)USDA-ARS, Lubbock, TX, (7)USDA-ARS, Gainesville, FL, (8)USDA ARS, Lubbock, TX

With the advance of next-generation sequencing, it is feasible to develop genomic knowledge for any crops and translate this information into genetic improvement. In this talk we will highlight work dissecting a gene network associated grain number per panicle (GNP) in sorghum, a major determinant of grain yield in cereals. Utilizing a sorghum EMS population, a series of sorghum [Sorghum bicolor (L.) Moench] multiseeded (msd) mutants, were isolated, which can double GNP by increasing panicle size and altering floral development so that all spikelets are fertile and set seed. Using a bulk segregant sequencing approach, we identified Msd1 as a TCP transcription factor. The six causal SNPs found in msd1 gene are highly conserved in grass species and not found in the currently genotyped natural population. The TCP gene was found to be specifically expressed in floral organs during inflorescense development. Whole-genome expression profiling revealed that jasmonic acid (JA) biosynthetic enzymes are transiently activated in pedicellate spikelets. Young msd1 panicles have 50% less JA than wild type (WT) panicles, and application of exogenous JA rescued the msd1 phenotype. Our results reveal a new biological mechanism for increasing GNP, with the potential to boost yield, and provide insight into the regulation of plant inflorescence architecture and development.

W929: Sorghum/Millet

Gene Discovery in *Setaria viridis*: A Gateway to Maize and Sorghum Crop Improvement

Pu Huang[1], Elizabeth A. Kellogg[1] and **Thomas P. Brutnell**[2], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)Enterprise Institute for Renewable Fuels Donald Danforth Plant Science Center, St. Louis, MO

Setaria viridis is an emerging model system for the study of panicoid grasses that include some of the world's most important food, feed and energy crops. The short stature, prolific seed production and ease of growth of this readily transformable diploid annual species facilitates genetic and genomic characterizations of gene function. Sequencing of the *S. viridis* reference genome A10.1 by the Department of Energy Joint Genome Institute has opened the door to large scale functional genomics applications including both forward and reverse genetics analysis. Deep sequencing of a collection of 600 diverse accessions has facilitated the use of GWAS studies to identify candidate genes underlying important agronomic traits and the development of large NMU-mutagenized populations will enable gene discovery through forward genetic screens and bulked segregant analysis. In this presentation I will provide specific examples of how we are using S. *viridis* to identify and characterize genes required for C_4 photosynthesis, detail a forward genetic screen that led to the identification of a novel component of inflorescence architecture and present the results of a genome wide association study that led to the identification of a novel gene that controls seed shattering, a key domestication trait. I will also demonstrate how it is possible to quickly translate discoveries in Setaria into strategies for crop improvement in maize, sorghum and the millets.

W930: Sorghum/Millet

Employed Forward Genetics and Map-Based Cloning Approaches in *Setaria* Functional Genomics Research

Xianmin Diao, Hui Zhi, Sha Tang, Guanqing Jia and Xiaotong Liu, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

Map-based cloning is a classical approach for gene isolation and function analysis. A great number of important genes controlling plant growth and development were identified through this approach in *Arabidopsis*, rice and other plant species. For years, our team has been making efforts to make map-based cloning works in foxtail millet, *Setaria italica*, which is an emerging model for Panicoideae grasses and C_4 photosynthesis, due to its small diploid genome size, short growth duration and prolific seed setting. We had developed an EMS induced mutation library with the *Setaria* reference genome variety “Yugu 1” which is composed of over 40 thousand lines, and more than 500 lines with obvious morphological alteration were identified. A platform of BSA-Seq, Mut-Map and classical linkage mapping of the mutated genes

was constructed which was approved being efficient. With the application of this platform, the putative causal genes of more than 30 morphologically mutated lines were fine mapped or cloned, and some of those genes were functionally characterized, such as *SiAGO1b*, *SiYGL1, SiSp-1* and *SiEGY1*. Using forward genetics we also cloned several naturally occurred mutations in foxtail millet including a semi-dominant dwarf gene used in foxtail millet breeding. So far our practice in foxtail millet forward genetics demonstrated that the cloning of mutated target genes can be very efficient in this species which provide substantial foundation for the initiation of Setaria as a model for functional genomics of Panicoideae crops and C_4 photosynthesis.

W931: Soybean Genomics

Genomic Resources for Soybean Research at NCBI

Anjana Raina Vatsan, Francoise Thibaud-Nissen and Terence D. Murphy, National Center for Biotechnology Information (NCBI), Bethesda, MD

The National Center for Biotechnology Information (NCBI) provides a range of resources and tools for storage and analysis of genomic data from a wide variety of organisms. A diverse subset of genome assemblies publicly available in the International Nucleotide Sequence Databases (GenBank/DDBJ/EMBL) are selected for annotation by NCBI's Eukaryotic Genome Annotation Pipeline and inclusion in NCBI's RefSeq dataset (https://www.ncbi.nlm.nih.gov/refseq/). To date, 73 plant genomes have been annotated by NCBI, including 8 species in the Fabaceae family. Three plant species, soybean, corn, and tomato, have been selected for further expert curation by the NCBI RefSeq group. 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/), and gene annotation and other 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/.

W932: Soybean Genomics

Survey for the Key Domestication Genes in Soybean

Min Wang, Wenzhen Li, Chao Fang, Fan Xu, Yucheng Liu, Chengcai Chu and **Zhixi Tian**, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Domestication has been considered as one of the most important technological innovations in the human history. Soybean is a crop with substantial economic value, accounting for more than half of global oilseed production. It has been suggested that cultivated soybean was domesticated from wild soybean in China 5,000 years ago. Identification of the genes contributing to domestication will facilitate super variety development. Preciously, we performed a large scale assessment of soybean domestication and improvement by resequencing of 302 wild, landrace and cultivated soybean lines. Bioinformatics analysis identified a total of 121 and 109 selective sweeps during soybean domestication and improvement, respectively. We further investigate the gene responsible for specific traits via genome-wide association study, which lead to the identification of several key genes related to important domestication traits.

W933: Soybean Genomics

Development of Genomic and Genetic Resources for Japanese Soybean Germplasm

Akito Kaga, Institute of Crop Science, NARO, Tsukuba, Japan

Soybean is an important source of traditional foods such as tofu, natto, miso, soy sauce and edamame in Japan. Japanese soybeans have been selected for the food usages from a long time ago and have different characteristics from overseas soybean. To provide a platform for breeding and diversity research of Japanese soybean, several experimental resources have been developed based on the genomic information.
Approx. 1,300 representative accessions that retain diversity in the NARO Genebank were selected as a core collection; 830 Japanese local landraces and 340 exotic landraces. By assessing SNP markers and agro-morphologic traits, two mini-core collections, each consisting of 96 accessions from Japanese and exotic germplasm, have also been developed. Recently, SNPs among representative Japanese soybeans were identified by NGS and genotype data of SNPs has been integrated to the two mini-core collections for GWA and GWP studies.
A soybean mutant library consisted of 1,536 plants was constructed for Japanese soybean Enrei. The mutation frequency, 1 mutation per 74 kbp, is almost saturated and sufficiently high to obtain a plant with a nonsense mutation in each gene from the mutant library. The proposed method for mutant screening using NGS will be useful to identify mutants for functional studies of soybean genes and have a potential to yield new alleles for soybean breeding.
In addition, chromosomal segment substitution lines (CSSLs) using progeny of a cross between the Japanese soybean Enrei and Chinese soybean Peking have been developed. CSSLs are a unique resource in soybean and are useful for evaluation of QTLs with small genetic effects as single genetic factors in a uniform genetic background. CSSLs might be a good starting material for breeding and for isolation of uncharacterized useful Peking genes conferring disease and stress resistance.

W934: Soybean Genomics

Genome-Wide Analysis of Alternative Splicing (AS) in *Glycine max* and its Related Species *Phaseolus vulgaris*: Conservation of as Events Among Homologous Genes

Luis P Iñiguez, Centro de Ciencias Genómicas - Universidad Nacional Autónoma de México, Ciudad de México, Mexico

The vast diversification of proteins in eukaryotic cells has been related to multiple transcript isoforms from a single gene that result from alternative splicing (AS) of the primary transcripts. Analysis of RNA sequence data derived from expressed sequence tags and next generation RNA sequencing has been crucial for AS documentation and genome-wide AS studies. We developed a pipeline for the identification of AS events in *Glycine max* that was used as our plant model. The evolutionary conservation of AS events has been related to potential biological function, thus we analyzed AS events in its related specie *Phaseolus vulgaris*. We identified 134,316 AS events from *G. max* in 70% of expressed genes and 85,570 AS events in 72% of expressed genes from *P. vulgaris*. These were categorized into seven different AS events types, being intron retention the most abundant event followed by alternative acceptor and alternative donor, these represented ~75% of all AS

events in both plants. The proportion of conservation of AS events in homologous genes from the two legumes was around 30% considering all types of AS events and variations of the proportion were observed according to the type of AS event. An overrepresentation of conserved AS events that affected 5'UTR regions was observed for certain types of events. The identification and analysis of AS events are first steps to understand their biological relevance. The results presented here from two related legume species revealed high conservation, over ~15-20 MY of divergence, thus pointing to the biological relevance of AS.

W935: Soybean Genomics
Environmental Modulation of Regulatory Gene Networks in Soybean
Yoshie Hanzawa, University of Illinois at Urbana-Champaign, Urbana, IL

W936: Soybean Genomics
Genomics of Soybean Cyst Nematode Resistance
Khalid Meksem, Southern Illinois University Carbondale, Carbondale, IL and Liu S, Kandoth PK, Lakhssassi N, Kang J, Colantonio V, Heinz R, Yeckel G, Zhou Z, Bekal S, Dapprich J, Rotter B, Cianzio S, Mitchum MG, Meksem K
Two types of resistant soybean (Glycine max (L.) Merr.) sources are widely used against soybean cyst nematode (SCN, Heterodera glycines Ichinohe). These include Peking-type soybean, whose resistance requires both the rhg1-a and Rhg4 alleles, and PI 88788-type soybean, whose resistance requires only the rhg1-b allele. Multiple copy number of PI 88788-type GmSNAP18 in one genomic segment simultaneously contribute to rhg1-b resistance. Using an integrated set of genetic and genomic approaches, we demonstrate that the rhg1-a Peking-type GmSNAP18 is sufficient for resistance to SCN in combination with Rhg4. The two SNAPs (soluble NSF attachment proteins) differ by only five amino acids. Our findings suggest that Peking-type GmSNAP18 is performing a different role in SCN resistance than PI 88788-type GmSNAP18. We previously showed that the *Rhg4* gene encodes a predicted cytosolic serine hydroxymethyltransferase (GmSHMT08); however, the novel gain of function of *GmSHMT08* in SCN resistance remains to be characterized. Using a forward genetic screen, we identified an allelic series of *GmSHMT08* mutants that shed new light on the mechanistic aspects of *GmSHMT08*-mediated resistance. The new mutants provide compelling genetic evidence that Peking-type *rhg1* resistance in cv Forrest is fully dependent on the *GmSHMT08* gene and demonstrates that this resistance is mechanistically different from the PI 88788-type of resistance. As such, this is an example of a pathogen resistance gene that has evolved to underlie two types of resistance, yet ensure the same function within a single plant species.

W937: Soybean Genomics
Soybean Cyst Nemtode Genome Assembly
Rick Masonbrink, Iowa State University, Ames, IA

W938: Statistical Genomics
Optimal Designs for Genetic Improvement
William D. Beavis, Iowa State University, Ames, IA
To realize the promises of plant genomics, discoveries need to be translated into genetically improved crop cultivars. If the objectives of a genetic improvement project are clear and measurable they can be translated into objective functions (mathematical models) with decision variables and constraints that can be shown to have feasible solutions. Among the feasible solutions, most require trade-offs among the objectives that are often best framed in terms of maximizing the probability of success while minimizing costs and time. Pareto optimality plots can be used to visualize predicted outcomes from competing designs and provide objective criteria for rational decisions. In particular, the systematic rigor associated with a cost, time and probability of success (CTP) framework is well suited to designing plant breeding projects that require dynamic decision making affected by the underlying stochastic processes of transmission genetics. These principles were applied to objectives for trait introgression projects involving one to many desirable alleles. The resulting Pareto optimality plots reveal that the CTP framework identified novel breeding strategies with twice the probability of success as previously identified 'best breeding strategies' without increasing time or expense. The framework also reveals when budgets are insufficient for reasonable probabilities of success.

W939: Statistical Genomics
A Rapid Epistatic Mixed-Model Association Analysis by Linear Retransformations of Genomic Estimated Values
Jian-Feng Liu, China Agricultural University, Beijing, China

W940: Statistical Genomics
Parallel Construction of Genome-Wide Gene Regulatory Networks
Min Zhang, Purdue University, West Lafayette, IN
Constructing gene regulatory networks is central to understanding the genetic architecture of complex traits. Taking advantage of both gene expression and genomic polymorphism data, we propose a two-stage penalized least squares method to build large systems of structural equations for network construction. The system can be constructed using large numbers of endogenous and exogenous variables at the first stage, followed by consistent selection of regulatory effects at the second stage. We have shown that the resultant estimates of regulatory effects have oracle properties. The method is computationally efficient and allows for parallel implementation. We demonstrate the superior performance of the method with computer-based simulation studies. In addition, the method was applied to real data.

W941: Statistical Genomics
Multiple Locus Genome-Wide Association Studies *via* Sparse Bayesian Learning
Meiyue Wang, University of California, Riverside, Riverside, CA

We developed a sparse Bayesian learning (SBL) technique for QTL mapping and genome-wide association studies (GWAS). The method deals with a multiple locus model that includes all markers in a single model. We adopted a coordinate descent approach to update parameters (marker effects) by estimating one parameter at a time conditional on values of all other parameters. The process requires multiple iterations and the final estimated parameters take the values after the iteration converges. The coordinate descent approach is much the same as the optimization method adopted by the least absolute shrinkage selection operator (LASSO) method. However, the new SBL method adopted a different type of penalty that allows the method to handle extremely large sample sizes that cannot be analyzed with the LASSO method. We conducted simulation studies to demonstrate the differences between the new SBL method and current GWAS methods, e.g., LASSO and EMMA (efficient mixed model association). Under the same level of Type I error, SBL often shows a higher statistical power than the other two methods. True loci are often detected with extremely small p values, indicating that SBL is insensitive to more stringent thresholds in test statistic chosen by investigators. We further applied the new method to detect QTLs in rice populations (including inbreds and hybrids). We detected several QTLs associated with yield and yield-component traits of rice.

W942: Statistical Genomics
A Tutorial on Predictability of Genomic Selection
Shizhong Xu, Department of Botany & Plant Sciences, University of California, Riverside, CA
Predictability is a measurement of effectiveness of a model for predicting future individuals (test sample) using data from current individuals (training sample). Predictability differs from model goodness of fit, which measures how good a model fits the data and pays no attention to data points beyond the training sample. In this tutorial, I introduce a few different ways of calculating predictability commonly seen in the literature and discuss the pros and cons of each method. The most commonly used method for evaluating the predictability of a model is the K-fold cross validation where one fold is predicted using parameters estimated from the other K – 1 folds. Eventually, all folds are predicted using the remaining folds excluding the fold being predicted. However, the fact that the HAT method used to diagnose leverage and influence in linear regression analyses can replace cross validation is rarely known in the genomic selection community. Furthermore, the HAT method has been extended to measure the predictability of a mixed model, which is the most popular method for genomic selection. In this tutorial, I also introduce the HAT method in mixed models and demonstrate its application to genomic selection for agronomic, metabolomic, transcriptomic and phenomic traits.

W943: Sugar Beet Workshop
DEG-SNP Associated with Resistance and Susceptibility to Beet Curly Top Virus
Imad Eujayl, USDA-ARS, Northwest Irrigation and Soils Res. Lab., Kimberly, ID and Carl Strausbaugh, USDA-ARS-NWISRL, Kimberly, ID
Resistance to *Beet curly top virus* (BCTV-transmitted by beet leafhopper) is an essential trait for sugar beet cultivars to be grown in arid and semi-arid areas worldwide. Resistant doubled haploid line (KDH13=PI663862) and susceptible inbred line (K19-19) were utilized to identify differentially expressed genes regulating reactions to BCTV. Transcriptomic data from seven treatments of KDH13; un-infested, infested but not-infected, and infected with one of the three BCTV strains (Cal/Log, Wor, and Sev) or a combination of the three strains were analyzed against transcriptomic data of the susceptible line that was infected with the three strains. Based on 28 pair-wise comparisons, differentially expressed transcripts/genes (DEG) were identified at threshold of False-Discovery-Rate (FDR) of <0.05 and a LogFC (fold change) >±2.0. BCTV resistance candidate genes included cysteine-rich receptor like proteins (CRKs), which accounted for up to 11% (171 members) of the significant DEG between resistant and susceptible lines were analyzed to discover SNP within the DEG. DEG-SNP…..

W944: Sugar Beet Workshop
Identification of Conserved and Novel Resistance Gene Signatures across Four Crop Types of *Beta vulgaris*
Andy Funk, Michigan State University, East Lansing, MI

W945: Sugar Beet Workshop
Heterosis Measurement Initiatives on Time Series Growth Pattern of Sugar Beet
Kazunori Taguchi[1], Wei Guo[2], Atsushi Itoh[1], Kazuyuki Okazaki[1], Yosuke Kuroda[1], Hiroaki Matsuhira[1], Shigenori Ueda[1], Seishi Ninomiya[2] and Masayuki Hirafuji[2], (1)NARO, Hokkaido Agricultural Research Center(HARC), Hokkaido, Japan, (2)The University of Tokyo, Tokyo, Japan
'Heterosis' is the phenomenon that progeny of diverse varieties of a species or crosses between species exhibit greater biomass, speed of development, and fertility than both parents. Sugar beet is one of the representative crop for getting high yield performance by using heterosis. But, it is still unclear of this genetic mechanism and why it occurs highly heterosis in a first crossed generation of 'F1'. Our ultimate goal is to solve about genetic background of heterosis, however, it is not easy work to trace for their growth patterns and speeds of numerical value in each plant.
One of the solution might be provided numerical value of digital captured time series data by UAV (Unmanned aerial vehicle) technology that could get as numerous images of plant growth period in a field level. In previous report of the PAGXXV, we tried to reconstruct 3D digital images captured from UAV and were converted to numerical data as sugar beet PVC (percentage of vegetational cover) and relative height on DSM (digital surface model) of each plot. Here, we tried to compare growth pattern of parents and F1s in half-diallele cross design.

W946: Sugar Beet Workshop
***Cercospora beticola* Comparative Genomics Sheds New Light on the Cercosporin Biosynthesis Pathway**
Melvin Bolton, USDA – Agricultural Research Service, Fargo, ND
Cercosporin is a light-activated secondary metabolite effector produced by many *Cercospora* species that contributes to fungal virulence. The metabolic pathway for cercosporin production has been well-characterized and was previously thought to consist of eight cercosporin toxin biosynthesis (*CTB*) genes. By comparing genome sequences of several ascomycetes, we found that the *CTB* cluster has experienced a number

of horizontal transfers across a spectrum of plant pathogenic fungi during evolution. Surprisingly, we noticed that these species also harbored an additional complement of genes on one flank of the established *CTB* cluster. Extensive microsynteny outside of the established cercosporin cluster prompted us to test whether the flanking genes in *C. beticola* are also required for cercosporin biosynthesis. Gene disruption of three genes led to the inability of the fungus to produce cercosporin. Taken together, our findings suggest that the *CTB* cluster includes more genes than previously known. A detailed characterization of these novel genes will be reported.

W947: Sugar Beet Workshop

Comparison of Three PCR-Based Methods for SNP Genotyping in Sugar Beet

Chiara Broccanello[1], Andy Funk[2], Claudia Chiodi[1] and **Piergiorgio Stevanato**[1], (1)University of Padova, Legnaro (Padova), Italy, (2)Michigan State University, East Lansing, MI

PCR allelic discrimination technologies have broad applications in the detection of single nucleotide polymorphisms (SNPs) in genetics and genomics. The use of fluorescence-tagged probes is the leading method for targeted SNP detection, but assay costs and error rates could be improved to increase genotyping efficiency. A new assay using RNase H2 PCR (rhPCR) attempts to reduce error rates from primer dimers while lowering costs compared to existing technologies. Before rhPCR can be widely adopted, it is important to validate its effectiveness versus established methods. The aim of this study was to compare the accuracy, sensitivity and costs of three PCR-based, high-throughput SNP genotyping approaches; TaqMan, KASP, and rhPCR. For each approach, assays were designed to genotype 33 SNPs in a set of 96 sugar beet individuals obtained from 12 genotypes. The sensitivity of each assay was tested using a series of 20 dilutions from 0.1 ng to 100 ng per reaction. Reactions were carried out on the QuantStudio 12K Flex Real-Time PCR System (Life Technologies, CA, USA). The call-rate, defined as the percentage of genotype calls relative to the possible number of calls, was 97%, 97.6%, and 98.1% for TaqMan, KASP, and rhPCR, respectively. Discordance among SNP calls was restricted to 9 of the 33 SNPs, with discordance between methods ranging from 0.11% to 0.66%. The sensitivity test demonstrated that the limit of detection (LOD) of rhPCR was the lowest of the three assays (0.2 ng of DNA per reaction). Costs per reaction were Euro 0.29 for TaqMan, Euro 0.11 for KASP, and Euro 0.13 for rhPCR. In conclusion, rhPCR produced slightly more calls than either TaqMan or KASP while remaining competitive in terms of cost per SNP.

W948: Sugar Beet Workshop

The Application of DArTseq Technology to Sugar Beet

Andrzej Kilian, Diversity Arrays Technology Pty Ltd, Canberra, Australia

W950: Sugar Cane (ICSB)

A Reference Sequence of the Monoploid Genome of Sugarcane

Olivier Garsmeur[1], Gaëtan Droc[1], Jane Grimwood[2], Bernard Potier[3], Karen S. Aitken[4], Jerry Jenkins[5], Carine Charron[1], Guillaume Martin[6], Edwin van der Vossen[7], Andrzej Kilian[8], Helene Berges[9], Blake Simmons[10], Jeremy Schmutz[11] and **Angélique D'Hont**[1], (1)CIRAD, UMR AGAP, Montpellier, France, (2)HudsonAlpha Institute for Biotechnology, Huntsvile, AL, (3)South African Sugarcane Research Institute, KwaZulu-Natal, South Africa, (4)CSIRO Agriculture and Food, St Lucia, Australia, (5)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (6)CIRAD, Montpellier, France, (7)Keygene, Wageningen, Netherlands, (8)Diversity Arrays Technology Pty Ltd, Canberra, Australia, (9)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (10)Joint Bioenergy Instutute, San Francisco, CA, (11)Hudson Alpha, Huntsville, AL

The sugarcane genome poses challenges that have not been addressed in any prior genome sequencing project. The main difficulties reside in the high polyploidy (2n ~ 12x ~ 120) and the high level of heterozygosity of cultivars, which make an assembly of the genome very challenging through classical whole genome shotgun sequencing approaches. We developed an alternative sequencing strategy that aims to produce the sequence of one monoploid genome. Our strategy is based on previous studies that demonstrated that sugarcane hom(e)ologous chromosomes share a very high level of micro-collinearity among themselves and show good micro-collinearity with sorghum. We used the Whole Genome Profiling technology (WGP™, KeyGene) to analyze a set of 20,736 BACs from cultivar R570. An average of 37.2 WGP sequence tags per BAC was generated that allowed the anchoring on the sorghum genome of 11,732 R570 BACs. A Minimum Tiling Path of 4,688 sugarcane BAC clones representing the minimum number of BACs that best cover the gene rich part of the sorghum genome was selected and sequenced using PacBio RSII technology through international collaborations. High quality sequences were obtained and almost all BACs could be assembled into single contigs. Overlapping BAC sequences were trimmed resulting in a Single Tiling Path of 382 Mb of high quality sequence. RNAseq resources from distinct tissues of R570 were produced using Illumina Hiseq 2500 and integrated in the gene annotation process that resulted in the prediction of 25,316 protein-coding gene models. A 12,468 SNP-based R570 genetic map revealed a few large chromosomal rearrangements in *S. spontaneum* as compared to *S. officinarum* and sorghum. A sugarcane web portal is currently being developed to make BAC sequences and gene annotations available to the sugarcane community. This high quality reference sequence that corresponds to the gene rich part of a sugarcane monoploid genome will represent a very important resource for genetic, structural and functional genomic studies in sugarcane and also an essential framework to help assemble a whole genome sugarcane sequence.

We acknowledge members of the International Consortium for Sugarcane Biotechnology (ICSB) for financial support.

W951: Sugar Cane (ICSB)

Genomic Analysis of *Saccharum spontaneum* Haploid AP85-441

Jisen Zhang[1], Xingtan Zhang[2], Qing Zhang[3], Tyler Jones[4], Chifumi Nagai[4] and **Ray Ming**[5], (1)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, (2)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, Fujian, China, (3)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, (4)Hawaii Agriculture Research Center, Kunia, HI, (5)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.

W952: Sugar Cane (ICSB)

Structural Basis of the Antifungal Activity of SUGARWINs Proteins and their Role in Plant Defense

Flávia P. Franco[1], Renata O. Dias[1], Danyelle Toyama[2], Adelita C. Santiago[2], Flávio Henrique-Silva[2], Daniel S Moura[1] and Marcio C Silva-Filho[1], (1)Universidade de São Paulo - USP, Piracicaba, Brazil, (2)Universidade Federal de São Carlos - UFSCar, São Carlos, Brazil

Plants respond to insect attack by inducing and accumulating a large set of defense proteins. We identified two homologues of a barley wound-inducible protein (BARWIN) in sugarcane, which were designated SUGARWIN1 and 2 (sugarcane wound-inducible proteins). Although BARWIN function has not been fully established, antifungal properties have been described for a number of homologues. SUGARWIN1 and 2 genes expression are induced in response to wound and *Diatraea saccharalis* damage. Although the recombinant SUGARWIN protein does not affect insect development, it promotes significant morphological and physiological changes in *Fusarium verticillioides* and *Colletotrichum falcatum*, which lead to fungal cell death via apoptosis. In this study, we deepen our understanding of the role of SUGARWINs in plant defense and the molecular mechanisms by which these proteins affect fungi by elucidating their molecular targets. We demonstrated that SUGARWINs are also induced by *C. falcatum* infection in sugarcane, and the induction of SUGARWINs can vary among sugarcane varieties. The sugarcane variety exhibiting the highest level of SUGARWIN induction exhibited a considerable reduction in *C. falcatum* infection. Furthermore, SUGARWIN1 exhibited ribonuclease and chitinase activity, whereas SUGARWIN2 exhibited only chitinase activity. This variable enzymatic specificity seems to be the result of divergent amino acid composition within the substrate-binding site that was demonstrated by protein modelling and docking studies. We confirmed this result producing mutants with altered active site and performing comparative analysis with the native protein. Our results show that SUGARWINs play an important role in plant defense against opportunistic pathogens and can be important to red rot disease control.

Financial Support: This work was supported by a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and by a Conselho Nacional de Desenvolvimento Cienti´fico e Tecnolo´gico (CNPq).

W953: Sugar Cane (ICSB)

Identification and Characterization of Genes Responsible for the Brown Rust Resistance (Bru1) Effect

Shailesh Vinay Joshi[1,2] and Dyfed Lloyd Evans[2,3], (1)South African Sugarcane Research Institute, Durban, South Africa, (2)School of Life Sciences, College of Agriculture, Engineering and Science,, Durban, South Africa, (3)South African Sugarcane Research Insitute, Durban, South Africa

Brown rust resistance is a major sugarcane innate defence against the fungal pathogen *Puccinia melanocephala*. Bru1 loci are reported to be responsible for brown rust resistance and are being used as a platform for marker assisted selection (MAS) in sugarcane. Sequences for Bru1 alleles have been reported in R570, LA Purple and RB867515. We have assembled the cognate genes in LCP85-834 and SP80-3280. There are 14 genes in the core region, which are present in different copies on different alleles. Every gene was characterised molecularly by domain analysis, structural analysis, homology mapping, phylogenetics, active site analysis and pathway mapping. All but five genes could be excluded, based on their functional annotation. Two of these genes were further excluded, as they were only involved in DNA repair. Genes *S6PDH, S6PDHa, CDKA;1* and *UCCP* were further analysed. CDKA;1 shows unusual convergent evolution between the grass and *Brassica* cognates. In Arabidopsis CDKA;1 is responsible for pathogen response including epigenetic protection. *S6PDH* has recently been identified as a key cell death response gene. Uniquely in grasses, sugarcane *S6PDHa* possesses a strong chloroplast transition peptide, indicating that it may function in sugarcane chloroplasts to downregulate inositol and induce cell death. The functional pathway has been modelled. Transcript mapping indicates that *UCCP, CDKA;1* and *S6PDH* are constitutively expressed but *UCCP* and *S6PDHa* are significantly upregulated, while *CDKA;1* is down-regulated in response to fungal pathogens. These two major genes (*UCCP* and *S6PDHa*), in concert, are responsible for the brown rust resistance phenomenon in sugarcane and could be used directly as gene/transcript markers for MAS. We report, for the first time, the functional mechanism for the Bru1 response in sugarcane.

W954: Sugar Cane (ICSB)

Genomic Breeding Value Estimation of Sugarcane Genotypes Based on High-Density SNP

Luciana Gonçalves Castellani[1], Itaraju Brum[2] and Mike Butterfield[1], (1)Centro de Tecnologia Canavieira, Piracicaba, Brazil, (2)CTC, Piracicaba, Brazil

In the sugarcane breeding program at CTC (Centro de Tecnologia Canavieira), breeding values (BV) estimated from progeny performance data are used to select the best parents and superior predicted cross combinations. The objective of this study was to evaluate the accuracy using genome-wide SNP markers to predict genomic estimated breeding values (GEBV) of parents, without using progeny data. In order to calibrate the model, six years of data (2008 to 2013) from 6 locations on cane yield (TCH – tons cane per hectare) and sucrose content (SC) were used. A total of 1,276 parents with progeny data during this period were genotyped at ~22,000 SNP markers distributed throughout the sugarcane genome. Phenotypic BVs were estimated using a Parental/Family model including the pedigree matrix. The SNP data was analysed with a Genomic BLUP (GBLUP) model with a leave-one-out cross validation strategy resampling the phenotypic BVs. Accuracies of the methods were assessed through the correlation between observed (BV) and predicted (GEBV) values. Overall, the estimated values from markers show

good agreement with expected from phenotype and pedigree; 0.45 for SC and 0.60 for TCH. The prediction accuracy for the six individual locations ranged from 0.22 to 0.65 for SC and 0.45 to 0.73 for TCH. Additional work is ongoing to improve the prediction models.

W955: Sugar Cane (ICSB)

The Development of a Genetic Analysis Pipeline for SNP Marker Delivery in Sugarcane Breeding

Meredith D McNeil[1], Shamsul Bhuiyan[2], Xianming Wei[3], Priya A Joyce[4] and Karen S. Aitken[1], (1)CSIRO Agriculture and Food, St Lucia, Australia, (2)Sugar Research Australia, Woodford, Australia, (3)Sugar Research Australia, Mackay, Australia, (4)Sugar Research Australia, Indooroopilly, Australia

Marker assisted selection (MAS) for individual QTLs linked to key traits has become routinely implemented in many major crop breeding programs. Due to its complex polyploid genetic structure and large genome, MAS has lagged behind in sugarcane breeding. Recently, an Affymetrix® Axiom® 45K SNP chip was developed for sugarcane which has dramatically improved the construction of high-density linkage maps and identification of target QTLs for agronomic traits. We have identified a number of significant marker-trait associations for important diseases, such as smut and pachymetra, through a combination of QTL mapping, association mapping and RNA-seq studies using this SNP chip. However, the use of such SNP chips for MAS is still price-prohibitive in sugarcane breeding. In this study, a comparison was made of SNPs selected from the Axiom® sugarcane SNP chip converted to two SNP marker technologies, the LGC® KASP™ assay and the Fluidigm® SNPType™ assay, for the development of a high-throughput, low-cost SNP marker panel targeted to disease resistance QTLs. This presentation details the genetic analysis pipeline developed for the generation of marker-trait associations for disease resistance, and then subsequent conversion of these SNP markers to a flexible, high-throughput marker platform to be used for MAS in sugarcane breeding. These results offer the opportunity to apply selection early in the breeding cycle in a more cost-effective way when traditional phenotypic screening is not possible.

W956: Sugar Cane (ICSB)

Genomic Selection in Sugarcane in Florida

Per Hilding McCord, USDA Agricultural Research Service, Canal Point, FL and Sushma Sood, USDA-ARS, CANAL POINT, FL

We are interested in testing the ability of genomic selection to predict the performance of new sugarcane clones with regard to important agronomic and disease resistance traits, as part of the breeding program at the USDA-ARS Sugarcane Field Station at Canal Point, Florida. For evaluation of the technique, we are phenotyping 416 individuals from an early stage (Stage 2) of the Canal Point program, and an additional 18 individuals from breeding programs in Louisiana. Yield, sucrose content, and disease data were measured on Stage 2 clones in an unreplicated trial in the 2015-2016 season. All 434 clones are being evaluated in a replicated trial for the 2017-2018 season, and have currently been phenotyped for stalk population, stalk diameter, and resistance to brown and orange rust via artificial inoculation. All individuals were genotyped via capture sequencing using 10,000 optimized probes, resulting in 21,277 markers. Genomic selection models have been developed using GBLUP. Prediction accuracy, or the correlation between predicted and measured values, varied from -0.009 for Brix in the Stage 2 trial, to 0.41 for stalk population in the full trial. Additional genomic selection models are being explored, such as the addition of large-effect single markers from genome-wide association analysis (GWAS) of the existing dataset, and the efficacy of a reduced set of markers is also under investigation.

W957: Sugar Cane (ICSB)

Sugarcane Biotechnology- What Does Tomorrow Look like

Harjeet Kaur Khanna[1], Karen S. Aitken[2], Priya A Joyce[3], Prakash Lakshmanan[3] and Anne L. Rae[4], (1)Sugar Research Australia, Brisbane, Australia, (2)CSIRO Agriculture and Food, St Lucia, Australia, (3)Sugar Research Australia, Indooroopilly, Australia, (4)CSIRO (Agriculture and Food), St Lucia, Australia

Biotechnology is by nature a multidisciplinary field and consequently, a very wide range of scientific disciplines have contributed to the fast development of this field. It has evolved into a game-changer technology over the last couple of decades. Benefits of crop biotechnology are so compelling and applications so wide that today most crops have added this discipline to their portfolio with the intent of either increasing productivity, or enhancing protection from diseases and pests, or adding nutritional value, or creating environmental benefits or tackling a whole range of other challenges. This versatility was engineered through techniques that not only helped us understand organisms better but also assisted us by providing smart genetic resource management tools. Sugarcane biotechnology today involves genetic engineering, molecular markers, tissue culture, and genomics. However, biotechnology is evolving so fast that we don't know when another fascinating breakthrough or application in another crop will leave us wondering. Advances in precision gene-editing techniques, Nano biotechnology and Data science are transforming the way crops are now being bred. The precision and control over changes made through genome editing is fast improving and who knows, tomorrow we may be using gene-editing, chromosome engineering and 3D bio-printers to seed novel plants harboring only those traits that we desire. New technologies and new ways of working will bring with them new challenges. Enabling scientists around the world to address these challenges will mean providing them with better tools and technologies and this will require significantly accelerated R&D investment. It will also create the need for increased collaboration between sugarcane researchers and industries around the world. When resources are limited, it makes logical sense to conduct at least the basic foundational research as a team and then individual industries can focus on using those results to create targets outputs for their own needs. There will also be an increased need to facilitate public discourse, policies and regulations that support the research and adoptions of newer biotechnologies and it makes sense to have a common voice. Let's think if we need to think differently.

W958: Sugar Cane Sequencing Initiative

Genomic Analysis of *S. officinarum* La Purple

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

W959: Sugar Cane Sequencing Initiative
Assembling a Hybrid Sugarcane Genome
Jeremy Shearman[1], Chutima Sonthirod[1], Chaiwat Naktang[1], Wirulda Pootakham[1], Thippawan Yoocha[1], Duangjai Sangsrakru[1], Nukoon Jomchai[1], Sissades Tongsima[1], Jittima Piriyapongsa[1], Chumpol Ngamphiw[1], Nanchaya Wanasen[1], Kittipat Ukoskit[2], Prapat Punpee[3], Peeraya Klomsa-ard[3], Klanarong Sriroth[3], Jisen Zhang[4], Xingtan Zhang[4], Ray Ming[5], Somvong Tragoonrung[1] and Sithichoke Tangphatsornruang[1], (1)National Center for Genetic Engineering and Biotechnology, Pathum Thani, Thailand, (2)Department of Biotechnology, Faculty of Science and Technology, Thammasart University, Pathum Thani, Thailand, (3)Crop Production, Mitr Phol Innovation and Research Center, Pathum Thani, Thailand, (4)Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fujian, China, (5)University of Illinois at Urbana-Champaign, Urbana, IL
We have used 700 PacBio RSII cells to sequence and assemble the genome of a South East Asian commercial cultivar, known as Khon Kaen 3. Modern commercial cultivars are auto- allopolyploids with complex chromosome pairing. The long read length of this sequencing technology allowed for the large complex polyploid genome of Khon Kaen 3 to be assembled into a relatively small number of contigs. We also generated a reduced representation assembly by cutting the corrected PacBio reads into 500 bp fragments followed by short read assembly. Comparison of the two assemblies shows that the large reads allowed for separation of orthologous, and perhaps even in some cases paralogous, sequence into separate contigs. This sequence separation poses a significant problem for mapping short reads to the assembly. We also compare our assembly to the genome assemblies of S. spontaneum and S. officinarum to identify which parts of each make up the Khon Kaen 3 genome and at what copy number. This reveals a make up consistent with what has previously been found using markers such as microsatellites and RFLPs.

W960: Sugar Cane Sequencing Initiative
Identifying Regulatory Sequences of Co-Expressed Genes in Sugarcane
Augusto L Diniz, Institute of Chemistry/University of São Paulo, São Paulo, Brazil

W961: Sugar Cane Sequencing Initiative
Unraveling the Genome of a High Yielding Colombian Sugarcane Hybrid
Jhon Henry Trujillo[1], Jorge Duitama[2], Cristian Dario Loaiza[1], Manuel Quintero[1], Jose De Vega[3] and **John Riascos**[1], (1)Colombian Sugarcane Research Center, CENICAÑA, Cali, Colombia, (2)Universidad de los Andes, Bogotá, Colombia, (3)Earlham Institute, Norwich, United Kingdom
The *Saccharum* genus is characterized by the polyploid nature of its different species. Modern sugarcane hybrids have an interspecific origin (*Saccharum officinarum* x *Saccharum spontaneum*) and consequently harbor a polypoid and highly aneuploid genome. Although such complexity difficult the implementation of breeding strategies based on molecular markers, the advent of the next generation sequencing technologies (NGS) and the possibility of identifying thousands of SNP markers opened a new window of opportunity for molecular breeding. Previous studies developed at CENICAÑA, aimed at the identification of SNP markers using the sorghum (*Sorghum bicolor*) genome as a reference, showed that only between 15%-30% of RADSeq and GBS derived reads could be used for this purpose. This low percentage of alignment, in addition to the lack in information about its organization, evidences the need to assemble a sugarcane genome.
For this work, we chose to sequence a commercial sugarcane hybrid developed by CENICAÑA. The genotype of interest produces on average 20 more tons of cane per hectare (mean of 146.5 *vs*. 126.5) and 2 more tons of sucrose per hectare (mean of 15.7 *vs*. 13.6) than the historically outstanding and most currently cultivated genotype in Colombia. This, within the environmental conditions and agronomical practices of The Cauca River Valley province at the southwest of the country. Its genome size was measured by flow cytometry resulting in a total DNA content of 11.21 Gbp (SD of 0.374), with a monoploid genome size of 1.019 Gbp, estimated on the bases of an 11X ploidy. From this genotype, 45.2 Gbp of long PacBio reads and 107.2 Illumina short reads were produced and used for genome assembly. Using the software Canu (v 1.5) 17,672 contigs were assembled with an N50 of 42.3 kbp and a total genome size length of 642 Mbp. The previously produced GBS and RADSeq data were aligned against the assembly and variants were identified using the NGSEP pipeline. The alignment rates increased to 78% and 61% for the RADSeq and GBS datasets, respectively.
Based on our results we have produced an additional 50 Gbp of PacBio Data. Future works aim to improve this assembly using the new PacBio reads and emerging scaffolding technologies such as 10x and Hi-C.

W962: Sugar Cane Sequencing Initiative
Whole Genome Sequencing of Sugarcane: Building off the Foundation of the Single Haplotype Path
Jeremy Schmutz[1], Adam Healey[2], Jane Grimwood[3], Jerry Jenkins[3], Yesesri Cherukuri[3], Kerrie W. Barry[4], Olivier Garsmeur[5], Robert J. Henry[6], Angélique D'Hont[5] and Karen S. Aitken[7], (1)Hudson Alpha, Huntsville, AL, (2)HudsonAlpha Institute For Biotechnology, Huntsville, AL, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)DOE Joint Genome Institute, Walnut Creek, CA, (5)CIRAD, UMR AGAP, Montpellier, France, (6)University of Queensland/QAAFI, Brisbane, Australia, (7)CSIRO Agriculture and Food, St Lucia, Australia
Sugarcane is one of the world's most important economic crops, however despite its importance, research into its genetics has been limited due to the complexity of its genome. Modern sugarcane genotypes are derived from interspecific hybridization of two highly polyploid *Saccharum* species (*S. officinarum* and *S. spontaneum*), with repeated backcrossing to yield cultivars with high and variable ploidy (aneuploidy), housing unbalanced proportions of the parental genomes (~80% S.o; 10% S.s; 10% recombined), with a total genome size of approximately 10 Gb. However, despite its complexity, the *Saccharum* genome is largely syntenic with *Sorghum*, which has allowed us to generate a minimal tiling path of BAC clones corresponding to the sugarcane monoploid genome, organized into 10 chromosomes (382 Mb). In order to assemble a whole genome shotgun based reference for sugarcane, we are leveraging this tiling path to map the heterozygous complexity of R570, a commonly used sugarcane variety. We will report on a dense map of ~21 million non-overlapping unique 80bp kmers across 2.4 million loci identified from Illumina fragment and mate-pair libraries which reveal complex kmer networks and structural rearrangements relative to the

tiling path. Additionally, we have kmer counts among 48 selfed progeny (~12X haploid coverage) that allows up to identify pairwise correlations and linkage, based on reductions of heterozygosity, beyond our maximum library insert sizes. These efforts underpin the strategy to unravel the complexity of the sugarcane genome and facilitate the construction of a haplotype specific assembly of the genome.

W963: Sugar Cane Sequencing Initiative
Development of Affymetrix Axiom SNP Array for Polyploid Sugarcane
Jianping Wang, Agronomy Department, University of Florida, Gainesville, FL

W965: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
Session Welcome and Genomic Tools for Sweetpotato (GT4SP) Improvement Project Update
Craig Yencho, North Carolina State University, Raleigh, NC
Sweetpotato is a widely recognized food security and cash crop with highly recognized potential to alleviate hunger, vitamin A deficiency, and poverty in Sub-Saharan Africa (SSA). It is also a crop predominantly grown in small plot holdings by poor women farmers across SSA. The Genomic Tools for Sweetpotato (GT4SP) Improvement Project is working to develop modern genomic, genetic, and bioinformatics tools to facilitate sweetpotato improvement. Our goal is to developing a set of "next generation" breeder tools for African sweetpotato breeders. In this 5 min. talk I will provide an overview of GT4SP project and introduce the sweetpotato speakers in the workshop.

W966: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
Linkage and QTL Mapping in Complex Polyploids: New Algorithms and Methods
Guilherme Da Silva Pereira[1], Marcelo Mollinari[1], Bode Olukolu[1], Dorcus C. Gemenet[2], Craig Yencho[1] and Zhao-Bang Zeng[1], (1)North Carolina State University, Raleigh, NC, (2)International Potato Center (CIP), Lima 12, Peru
Due to their intrinsic complexity, autopolyploid plant species have been left behind compared to the diploid ones regarding the usage of molecular markers for genetic studies and their effective application in breeding. Only with the advent of single nucleotide polymorphism chip arrays and ultimately next-generation sequencing-based protocols, has it become possible to properly overcome the limitations imposed by the use of single-dosage markers coupled with diploid-like approaches. The need to call higher dosage markers has led to advances on linkage and quantitative trait loci (QTL) mapping analyses in tetraploid species, mainly in potato. However, these methods are not readily applied to higher ploidy levels. For example, the number of possible genotypes of a bi-parental cross offspring increases from 36 to 400 to 4,900 in tetra-, hexa- and octoploid species, respectively. By using hexaploid sweetpotato mapping populations, we were able to (i) sample higher dosage markers in about half of the genome, (ii) build integrated linkage maps taking into account all markers simultaneously and obtain conditional genotype probabilities using a hidden Markov model, and (iii) map multiple QTL using a random model approach. To deal with the computational burden of analyzing a hexaploid genome, we have developed new tools called polymap and polyqtl to make the process relatively fast yet statistically accurate and genetically comprehensible.

W967: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
Linkage and QTL Analysis in the Hexaploid New Kawogo x Beauregard Mapping Population
Bonny Michael Oloka[1], Bode Olukolu[2], Benard Yada[3], Milton O. Anyanga[4], Doreen Chelangat[3], Paul Musana[3], Agnes Alajo[4], Guilherme Da Silva Pereira[2], Marcelo Mollinari[2], Zhao-Bang Zeng[2] and Craig Yencho[2], (1)North Carolina State University, RALEIGH, NC, (2)North Carolina State University, Raleigh, NC, (3)National Crops Resources Research Institute (NaCRRI), Kampala, Uganda, (4)National Crops Resources Research Institute, Kampala, Uganda
Genetic improvement of sweetpotato, *Ipomoea batatas* (L.) Lam., for important agronomic traits has been slow over the years in the global arena especially in sub-Saharan Africa where the crop is a staple. This is largely due to the crop's complex hexaploid (2n = 6x = 90) genetics, its large genome size and out-crossing nature with significant self and cross incompatibilities, high heterozygosity, and a wide array of biotic and abiotic stresses. We developed a bi-parental mapping population from the sweetpotato cultivars "New Kawogo" x "Beauregard" (NKB), consisting of 287 segregating F1 progeny, and used next-generation sequencing, computing and bioinformatics technology to develop high quality SNPs, which we used for linkage mapping and QTL analysis. Using a modified genotyping by sequencing (GBS) pipeline, we were able to mine 1,409,131 SNPs from the alignment of sequence files from the NKB population to the *I. trifida* reference (version 3.0) at a rate of 74.4%. We were able to call 132,201 SNPs using SuperMASSA software and after filtering for segregation distortion and missing data we retained 5,624 high quality SNPs along with their respective dosage information. We used these SNPs to build a genetic linkage map and have identified all 15 linkage groups. QTL analysis for sweetpotato weevil, sweetpotato virus disease and storage root yield is in progress. These tools will facilitate more efficient introgression of important traits and subsequent faster genetic gain of key traits in this complex yet globally important crop.

W968: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
Advances in Sweetpotato Phenotyping in the USA and SSA
Edward Carey[1], Maria Andrade[2], Omar Benites[3], Doreen Chelangat[4], Bryan Ellerbrock[5], Raul Eyzaguirre[3], Dorcus C. Gemenet[3], Mercy N. Kitavi[6], Nicolas Morales[5], Lukas A. Mueller[5], Paul Musana[4], Robert O.M. Mwanga[7], Alex C. Ogbonna[5], Ivan Perez[3], Reuben T. Ssali[7], Jolien Swanckaert[1], Titima Tantikanjana[5], Benard Yada[4], Craig Yencho[8] and Luka Wanjohi[6], (1)International Potato Center (CIP), Kumasi, Ghana, (2)International Potato Center, Maputo, Mozambique, (3)International Potato Center (CIP), Lima 12, Peru, (4)National Crops Resources Research Institute (NaCRRI), Kampala, Uganda, (5)Boyce Thompson Institute, Ithaca, NY, (6)International Potato Center (CIP), Nairobi, Kenya, (7)International Potato Center (CIP), Kampala, Uganda, (8)North Carolina State University, Raleigh, NC
Sweetpotato breeders have long recognized the benefits of standardized data collection for sharing and comparing results. Recent support for breeding efforts under the Sweetpotato Action for Security and Health in Africa (SASHA) and Genomic Tools for Sweetpotato Improvement

(GT4SP) projects, have advanced development of methods and tools for phenotyping, including trait ontologies, and database and breeding program management. Training and capacity development through breeding communities of practice, and demand for high throughput phenotyping methods are strengthening user uptake and further development of these tools. Field data collection on mobile electronic devices is eliminating the need for cumbersome data transcription from paper to computer, and the use of bar code labelling is reducing errors and enhancing quality control at all stages of the breeding process. Sweetpotatobase (https://sweetpotatobase.org) a relational database designed to handle phenotypic and genotypic data, provides an increasingly versatile platform for data storage and tools for management and analysis. Linked to Sweetpotatobase, the fieldbook app (http://wheatgenetics.org/field-book) and the Highly Interactive Data Analysis Platform (HIDAP; https://research.cip.cgiar.org/gtdms/hidap/) provide increasingly integrated tools for data collection, quality control and analysis. Some traits of interest, such as yield are shared across all breeding programs, while others such as disease and pest resistances and quality attributes are specific to regions where they occur or to specific breeding programs and populations. Methods of phenotyping are and will continue to evolve with methods of data capture and understanding of traits as we move toward increasingly high throughput methods.

W969: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
The Potential for Crop Improvement in Yam Phylogenomics
Juan Viruel[1], Marybel Soto Gomez[2], Ilia J Leitch[3] and Paul Wilkin[1], (1)Royal Botanic Gardens, Kew, London, United Kingdom, (2)University of British Columbia, Vancouver, BC, Canada, (3)Jodrell Laboratory, Richmond, United Kingdom
Up to seven species of *Dioscorea* are widely cultivated worldwide in an extensive variety of agronomic environments and social contexts. Considerable efforts in crop improvement are being developed based on genomic approaches. However, the enormous species diversity within *Dioscorea* (ca. 650 accepted species) constitutes an unexplored potential for crop breeding, as many wild species possess traits of interest or even the potential to become cultivated entities themselves. Evolutionary patterns within the pantropical *Dioscorea* have been explored using one nuclear region and five plastid regions for approximately 20% of the species, showing congruent phylogenetic topologies and subsequent divergences correlated with different kinds of underground storage organs. These organs constitute a pivotal trait with direct economic impact as species edibility is usually linked to annual tuber replacement in some species, or the potential to accumulate secondary compounds of interest in perennial underground organs. Next-generation sequencing methods provide new opportunities to gather huge amounts of genomic data to explore novel biological questions within a phylogenetic context. We will discuss the potential of HybSeq and transcriptomics methods to identify the crop wild relatives and to understand the origin of the cultivated species.

W970: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
Developing Genomic Resources for the Water Yam, *Dioscorea alata* L.
Ranjana Bhattacharjee, International Institute of Tropical Agriculture, Ibadan, Nigeria and Jessica B. Lyons, University of California, Berkeley, Berkeley, CA
D. alata is an important food security and income generation crop for millions of smallholder farmers in the tropics and sub-tropics of West Africa, Asia, Latin America, and the Pacific islands. Water yam boasts desirable characteristics such as high nutritional content, yield under low soil fertility, and low post-harvest losses. However, production is constrained by biotic and abiotic stresses, and breeding for desired traits is arduous. In the context of surging population growth and climate change, modern genomic and genetic tools are urgently needed for the more efficient improvement of water yam. Here we report our progress towards a *D. alata* reference genome sequence, constructed from long- and short-read shotgun data combined with long-range linking information. Comparison between the *D. alata* draft genome and the recently published *D. rotundata* genome assembly reveals extensive conservation, but also highlights differences between the two yam species. Towards an initial SNP catalog for *D. alata*, we have shotgun sequenced seven accessions. These accessions are the parents of eight F1 mapping populations that comprise over 1500 offspring and segregate traits important to smallholder farmers. Genotyping of these populations is underway. In order to assess global genetic diversity of *D. alata*, we are also gathering material from around the world for genotyping.

W971: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
Greater Yam (*Dioscorea alata* L.) Pre-Breeding and Breeding: Use of Genomic Tools to Decipher the Genetic Diversity and Identify Wild Relatives.
Hana Chaïr[1], Ranjana Bhattacharjee[2], Claudie Pavis[3], Marilyne Summo[1], Fabien Cormier[4], Gemma Arnau[1], Vincent Lebot[5] and PAG author's group, (1)CIRAD, Montpellier, France, (2)International Institute of Tropical Agriculture, Ibadan, Nigeria, (3)INRA Antilles-Guyane, Petit-Bourg, France, (4)BIOGEMMA, Chappes, France, (5)CIRAD, Port-Vila, Vanuatu
The greater yam (*Dioscorea alata* L.) is the most widespread edible yam species and is cultivated throughout sub-tropical and tropical areas. The species is an important food in West Africa, the Caribbean and the Pacific where it has considerable social and cultural importance, and it is also grown in parts of upland Asia. Although *D. alata* production is expanding in West Africa because of its ease of cultivation, one of the major constraints to further development is its suitability for “fufu”, a traditional dish necessitating tubers with high dry matter and specific starch contents. Moreover, some varieties with agronomic importance are susceptible to anthracnose (*Colletotrichum gloeosporioides*). Added to that, *D. alata* genetic improvement is constrained by access to well documented germplasm. Nevertheless, collections of *D. alata* and related species exist in international and national genebanks. Thus, to overcome these main limitations, rationalise the ex situ collections, and facilitate breeding for tuber quality and anthracnose tolerance, using Genotyping By Sequencing (GBS), we are investigating the genetic diversity of a worldwide sample of more than 500 *D. alata* accessions. Using targeted genotyping approaches on chloroplast and nuclear genomes, we are also investigating the relationship between *D. alata* and *D. nummularia* considered as one of its wild relatives from the Pacific. To identify gene/QTLs related to key agronomic traits, genetic mapping is on-going as well. Finally, the genomic resources produced are assembled to build up a “Yam Genome Hub”

W972: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement
New Insights into Yam (*D. rotundata*) Diversity and Domestication through 167 Whole Genome Re-Sequencing

Yves P. Vigouroux[1], Nora Scarcelli[1], Jude Obidiegwu[2], Emmanuel Otoo[3] and PAG author's group, (1)IRD, Montpellier, France, (2)National Root Crops Research Institute (NRCRI) Umudike, Umudike, Nigeria, (3)CSIR-Crops Research Institute, Kumasi, Ghana

Dioscorea rotundata is the most important yam species produced in West-Africa but we still know very little about its origin and evolutionary history. Thanks to the fast development of NGS technology, it is know possible to analyse the genetic diversity at genome-wide level, and to go deep insight into yam diversity and history. In this study, we analyzed 167 whole genome re-sequencing of cultivated yams and its two wild relatives (*D. abyssinica* and *D. praehensilis*). Sampling covers a wide geographical area, from Ghana to Cameroon. A mapping to the *D. rotundata* genome reference reveals more than 3 million high quality SNP. Wild yam diversity is higly structured while cultivated yam diversity reveals only weak genetic structure. The analysis and modelling of cultivated yam diversity allows to present for the first time some statistically supported hypothesis for yam domestication history.

W973: Sweetpotato and Yam Genomics: Next Generation Omics technologies for sweetpotato and yam improvement Workshop Discussion Session

Craig Yencho, North Carolina State University, Raleigh, NC, Ranjana Bhattacharjee, International Institute of Tropical Agriculture, Ibadan, Nigeria and Dorcus C. Gemenet, International Potato Center (CIP), Lima 12, Peru

General discussion of workshop presentations and wrap-up comments

W974: Swine

Ensembl Sscrofa11.1 Genome Annotation

Thibaut Hourlier, European Molecular Biology Laboratory - EBI, Cambridge, United Kingdom

Ensembl provides high-quality, reference annotation resources for publicly available genome assemblies, including domestic pig (*Sus scrofa*). Pig is an important model for cardiovascular disorders, infectious diseases, and xenotransplantation, and also an economically important species for meat production.

The Swine Genome Sequencing Consortium recently produced an improved reference assembly Sscrofa11.1 (GCA_000003025.6). We have fully updated all pig genome resources in Ensembl, including producing a curated gene annotation using new Illumina and PacBio transcriptome data.

Transcriptomic data sets improve annotations by adding alternative isoforms to genes and giving the possibility to discover new genes, either protein coding or non-coding. One of the new challenges is to provide methods to assign the type and functionality of the genes recently found. We will provide an insight on how it was incorporated in the Ensembl Annotation System.

These data are available in Ensembl. Data and tools to facilitate research on pig will be accessible through our website (www.ensembl.org), REST API (http://rest.ensembl.org), Variant Effect Predictor (www.ensembl.org/Tools/VEP), BioMart (http://www.ensembl.org/biomart) and our public MySQL server (ensembldb.ensembl.org).

Ensembl supports upload and visualisation of data in multiple file formats such as BAM or GFF3. For groups who wish to share their datasets and view them alongside Ensembl data, we have developed a TrackHubRegistry (http://trackhubregistry.org/) to enable discovery of publicly accessible track data hubs. You no longer need to host your own datafiles: you can now submit CRAM files to the ENA, register your track hub with the ENA accessions, and view your results in Ensembl.

W975: Swine

Additional Annotation of the Pig Transcriptome Using Integrated Iso-Seq and Illumina RNA-Seq Analysis

Hamid Beiki, Iowa State University, Ames, IA

Alternative splicing is a well-known phenomenon that dramatically increases eukaryotic transcriptome diversity. The extent of mRNA isoform diversity among porcine tissues was assessed using Pacific Biosciences single-molecule long-read isoform sequencing (Iso-Seq) and Illumina short read sequencing (RNA-seq) from a single individual White cross-bred pig. Isoseq data for nine tissues (brain, hypothalamus, liver, muscle, thymus, pituitary, small intestine, spleen and diaphragm) was error-corrected using RNA-seq data from the same RNA samples. Alignment of RNA-seq data to the current Ssc11.1 build (Ensembl release 90) for all 9 tissues revealed 401 tissue specific (TS) genes (50-fold higher FPKM level in one tissue compared with all others) and 8,309 housekeeping genes (FPKM≥1 in all tissues). Interestingly, 262 TS genes had no Gene Ontology annotation. Integration of IsoSeq and RNAseq data in liver and brain tissues identified 17,086 expressed (RNAseq FPKM≥1) novel isoforms (isoform that have at least one novel splice junction with an annotated transcript). Many expressed isoforms were detected within annotated intergenic (1,271) or intronic (801) regions. Many of these novel genes were validated using H3K36me3 (gene body mark) and H3K4me3 (promoter mark) ChIP-seq in liver. Analyses are ongoing for the other seven tissues, as well as tissue specific alternative splicing and long non-coding RNAs across all tissues. In summary, the results of this study can improve the incompletely annotated pig genome through the addition of alternative splicing complexity and identification of new features that are not included in the current pig genome annotation. The USDA is an equal opportunity provider and employer.

W976: Swine

Effect of sncRNA on Gene Expression on the Homeostatic Status of Pigs Infected with Highly Pathogenic PRRSV

Damarius S. Fleming, ORAU, USDA National Animal Disease Center, Texas A&M, Ames, IA and Laura C. Miller, NADC-ARS-USDA, Ames, IA

It has been established that reduced susceptibility to porcine reproductive and respiratory syndrome virus (PRRSV) has a genomic component. This component, however, is a multi-faceted composition of coding and non-coding genetic elements that function as regulators of immune function. Our study focuses on the small non-coding (sncRNA) side of this response in pigs because of the emergence of various sncRNAs shown to play important roles in human viral immunity. Among these sncRNAs are the microRNA (miRNA) and transfer RNA (tRNA) molecules. Our study looks at changes in expression of these sncRNAs to produce information on how gene function in the pig can become dysregulated and subsequently respond to the virus.

The objective of the study is to identify differences in miRNA and tRNA gene expression between healthy and highly pathogenic PRRSV challenged pigs. The study was conducted using total RNA extracted from pig whole blood taken from a total of 24 pigs split into either control (sham inoculation) or infected pigs at 1, 3, and 8 days post infection. Sequencing of the samples produced 100bp single end libraries for transcriptomic analysis of sncRNA gene expression.
The results indicated statistically significant changes in sncRNA expression were dependent on time and treatment. The miRNA expression was variable and showed perturbation of multiple signaling pathways; while tRNA expression declined steadily post-infection. The results of this study highlights changes in sncRNA expression that have the potential to unlock new targets for understanding the effect of PRRSV on pig homeostasis.

W977: Swine

Development of Targeted GBS Panels for Breeding and Parentage Applications in Cattle and Swine

Angela Burrell[1], Prasad Siddavatam[1], Michelle Swimley[1], Roy C. Willis[1], Maarten de Groot[2], Ryan Ferretti[3] and Rick Conrad[1], (1)Thermo Fisher Scientific, Austin, TX, (2)VHLGenetics, Wageningen, Netherlands, (3)Neogen/GeneSeek, Lincoln, NE

Parentage testing and genomics-assisted breeding are critical aspects of successful herd management. Due to its highly accurate and reproducible results, targeted GBS is becoming an increasingly favored technology for SNP genotyping. With the utilization of next-generation sequencing, labs can test hundreds of samples across thousands of SNPs simultaneously in a simple high throughput workflow starting from either extracted nucleic acid or crude lysis samples.
We developed targeted sequencing panels for both cattle parentage, based on 200 SNP markers selected by the International Society of Animal Genetics (ISAG), and swine breeding using a 1500 SNP imputation panel. Utilizing the AgriSeq™ HTS Library Kit, a high-throughput targeted amplification and re-sequencing workflow, each panel's performance was tested on >96 diverse cattle and swine DNA samples. Libraries were sequenced on the Ion S5™ using an Ion 540™ chip with genotyping calling generated using the Torrent Variant Caller (TVC) plugin
The mean genotype call rate of markers across the samples was >98% for the cattle panel and >96% for the swine panel. Concordance across replicate library preparations and independent sequencing runs was >99.9% for both panels. Panel results were compared with results from a DNA array and the genotype call concordance was >99% with the AgriSeq workflows. The cattle panel was also used on field samples by a Netherland service lab to successfully determine the parentage relationships of 45 calves with 48 potential mother cows.
The data demonstrates the utility of the AgriSeq targeted GBS approach for cattle and swine SNP genotyping applications.

W978: Swine

3D Genomic Mapping Reveals Transcription Regulation and Chromatin Organization of Skeletal Muscles in Swine

Jianhua Cao, Huazhong Agricultural University, Wuhan, China

The three-dimensional (3D) chromatin organization is important for gene transcription. However, the 3D chromatin organization and its effect on gene transcription still remain largely unknown in pig genome, and the dynamics of 3D genome of skeletal muscles is much less investigated. We applied long-reads ChIA-PET strategy to comprehensively map higher-order chromatin interactions and dynamics mediated by transcription factors CTCF and RNA Polymerase II (RNAPII) in porcine longissimus dorsi muscles (LDM). We found that 1,010, 142 CTCF-defined chromatins boundaries functioned as the fundamental scaffolds for genomic 3D architectures of skeletal muscles in swine, whereas RNAPII-mediated chromatin interactions primarily served as 927,236 transcription factories to reshape the chromatin organizations in porcine myogenesis activities. Our ChIA-PET data confirmed that the boundaries delimited by CTCF-binding sites play an important role in consolidating the stability of chromatin structures via the topological associated domains (TAD). The long-range interactions within CTCF topological foci were crucial in determining gene transcription related with postnatal muscle growth. Of particular, combined with RNA-Seq data and ChIP-Seq data, the muscle specific genes like *miR-1/133*, *MSTN* have novel elucidation on 3D genome aspect with ChIA-PET data. Our findings in this study thus provide new clues and potential targets to explore key elements related to muscle growth in swine and are also likely to shed light on elucidating chromatin organization and dynamics underlying the process of mammalian myogenesis.

Key Words
Pig; ChIA-PET; 3D genome; transcription factory; chromatin organization

Highlights

- ChIA-PET has great contribution to elucidate 3D chromatin organizations of skeletal muscles in swine.
- CTCF-defined chromatin TADs and RNAPII associated transcription factories co-regulated gene transcription including skeletal muscle specific genes.
- CTCF-binding sites at genome topological boundaries play an important role in skeletal muscle growth in swine.
- A good example shows how chromatin dynamics fundamentally impact on the gene transcription in muscle specific gene transcription in myogenesis.

W979: Swine

A Comprehensive Map of *cis* Regulatory Elements and 3D Structure of the Pig Genome

Yu Luan[1], Yunxia Zhao[1], Ye Hou[1], Yueyuan Xu[1], Huanhuan Zhou[1], Hongbo Yang[2], Feng Yue[2], Xinyun Li[1] and Shuhong Zhao[1], (1)Key Lab of Agricultural Animal Genetics, Breeding, and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan, China, (2)Department of Biochemistry and Molecular Biology, College of Medicine, The Pennsylvania State University, Hershey, PA

Pig is an important livestock species for meat supply. Moreover, as ~80% of the protein-coding genes are conserved between human and pig, pig is also an ideal model for biomedical studies.However, the functional annotation of pig genome is largely behind comparing with human and mouse. Here we took a similar approach adopted by the encode and roadmap epigenomics projects, and performed RNAseq and ChIPseq for H3K27ac and H3K4me3 histone markers to generate a comprehensive map of transcriptomes and regulatory element in a variety of pig tissues including muscle, fat, liver, heart, and spleen from Large White, Duroc, Meishan and Enshi Black pigs. We obtained over 10,000 cis regulatory elements in the pig genome, the most comprehensive functional annotation effort made so far in the pig. By comparing the data

generated by the ENCODE and Roadmap Epigenomics projects, we also defined a set of functionally conserved and species-specific regulatory sequences among pig, mouse and human.To further explore the three dimension(3D) structure of pig genome, we performed (high-throughput chromosome conformation capture (Hi-C)) experiment using pig muscle and liver tissues. Our HiC matrix showed that topologically-associating domains were also conserved among pig, mouse, and human. Muscle and liver specific enhancer and promoter interactions predicted from CHIP-seq data are further validated by muscle Hi-C matrix.
In summary, we generated a great genomics / epigenomic resource for the functional annotation and the 3D structure of the pig genome and further expanded the value of pig

W980: Swine
Genomic Co-Localization of microRNA Expression QTL Target Genes with Phenotypic QTL in the Michigan State University Duroc x Pietrain Pig Resource Population
Kaitlyn Daza, Deborah Velez-Irizarry, Sebastian Casiro, Juan P. Steibel, Nancy E. Raney, Ronald O. Bates and Catherine W. Ernst, Department of Animal Science, Michigan State University, East Lansing, MI
MicroRNAs (miRNAs) are a class of non-coding RNAs known to post-transcriptionally regulate gene expression through binding with target mRNAs, ultimately affecting a multitude of biological processes and phenotypes. Combining miRNA and gene expression profiles with genotypic and phenotypic data from the same animals enables the elucidation of regulation of complex traits important to the pig industry. The objective was to identify genomic co-localization events of miRNA expression Quantitative Trait Loci (miR-eQTL) miRNA target genes from *Longissimus dorsi* muscle samples with previously-identified phenotypic QTL (pQTL) in the MSU Duroc x Pietrain population. Animals were previously characterized for over 60 phenotypes and genotyped with Illumina PorcineSNP60 BeadChips. In total, 295 mature miRNA expression profiles were included in a GBLUP-based GWA analysis. Target genes for 15 miR-eQTL miRNAs were identified using TargetScan and filtered based on transcript abundance data from the same samples. Target genes negatively correlated (FDR≤0.05) with their associated miRNA's expression were co-localized with pQTL, yielding three miR-eQTL miRNAs with 29 total target genes overlapping pQTL across seven chromosomes. One miR-140-5p target, *RRP36*, co-localized with pQTL for dressing percentage (SSC7), while three targets of miR-6782-3p co-localized with pQTL for number of ribs (SSC7). Targets of miR-874 co-localized with a large pQTL on SSC15 for meat quality traits including juiciness, tenderness, Warner-Bratzler shear force, protein content, pH at 24 h, drip loss, and cook yield. Continuing to investigate miR-eQTL and their effects on downstream phenotypes will contribute to deciphering mechanisms controlling complex pig production traits.

W981: Swine
Integration of Gene Expression Profiling of Hypothalamic Arcuate Nucleus with Genome-Wide Associations to Discover Functional Variants Associated with Age at Puberty in Gilts
Hiruni R Wijesena[1], Clay A Lents[2], Brittney N Keel[2], Jean-Jack Riethoven[1], Matt L. Spangler[1], Stephen D. Kachman[1] and Daniel C. Ciobanu[1], (1)University of Nebraska - Lincoln, Lincoln, NE, (2)USDA, ARS, USMARC, Clay Center, NE
Age at puberty (AP) in gilts is a moderately heritable trait ($h^2 = 0.42$) and the earliest indicator of sow reproductive longevity. Therefore, quantifying the pleiotropic sources that influence both AP and reproductive longevity is important in understanding the differences in sow fertility. In this study, we integrated genome-wide associations (GWAS), whole genome sequencing and gene expression profiling of the micro-dissected hypothalamic arcuate nucleus using RNA sequencing to identify genetic variants associated with AP in the UNL resource population (n=1,644). The arcuate nucleus plays a major role in regulating the onset of puberty through controlled secretion of gonadotropins. Seventy differentially expressed genes (DEG) were identified ($P_{adj} < 0.1$) between early (n=11) and late (n=6) onset of puberty gilts. Three of these genes (*CDADC1, FAM111B* and *HERPUD2*) overlapped with major (top 1%) QTL regions for AP from GWAS. Genetic variants located upstream of transcription start site (<1000 bp) affecting potential cis-binding motifs were identified as possible sources of differential expression and variation in onset of puberty. For example, SNP-affected motifs for two transcription factors (*CTCF* and *SP2*) known to regulate the expression of both *CDADC1* and *FAM111B* were identified in the proximal promoter of these genes. There were 363 upstream regulators of the 70 DEG identified. Thirty-eight upstream regulators of six DEG (*CDKN1A*, *DPP4*, *FFAR2*, *LCN2*, *PGK1* and *SAMHD1*) overlapped with major QTL regions for AP. Missense SNP were identified in four regulators (*APC*, *CDCA2, IL17B* and *RAD9A*), which could be potential trans-modulators of DEG in gilts with differences in AP.
USDA is an equal opportunity provider and employer.
Keywords: gilts, age at puberty, hypothalamus, RNA, transcription motifs

W982: Swine
BARC Station Report
Joan K. Lunney, APDL, BARC, ARS, USDA, Beltsville, MD

W983: Swine
Host Variability and the Longitudinal Diversity of Microbiota Composition in Swine
Christian Maltecca, NC State University, Raleigh, NC, Bryce Lu, NCSU, Raleigh, NC, Francesco Tiezzi, NCSU, raleigh, NC, Nathan McNulty, Matatu Inc., Saint Louis,, MO and Clint Schwab, The Maschhoffs LLC, Carlyle, IL
In pigs, gut bacteria have been shown to play important roles in nutritional, physiological and immunological processes in the host. However, the contribution of their metagenomes or part of them, whas yet to be fully investigated. Fecal samples, collected from a population of pigs at 3 time points, including weaning, week 15 post weaning , and end-of-feeding test, were used to evaluate changes in the composition of the fecal microbiome of each animal over time. There were 1039 animals that had samples collected at all three time points, and also had phenotypic records on back fat thickness (BF) and average daily body weight gain (ADG). Firmicutes and Bacteroidetes were the most abundant phyla at all 3 time points. The most abundant genera at all 3 time points included *Clostridium*, *Escherichia*, *Bacteroides*, *Prevotella*, *Ruminococcus*, *Fusobacterium*, *Campylobacter*, *Eubacterium* and *Lactobacillus*. Two enterotypes were identified at each time point. However, only

enterotypes at week 15 and off-test were significantly associated with BF. We report herein 2 novel findings: (i) alpha diversity and operational taxonomic unit (OTU) richness were moderately heritable at week 15, h2 of 0.15 ± 0.06 to 0.16 ± 0.07 and 0.23 ± 0.09 to 0.26 ± 0.08, respectively, as well as at off-test, h2 of 0.20 ± 0.09 to 0.33 ± 0.10 and 0.17 ± 0.08 to 0.24 ± 0.08, respectively; whereas very low heritability estimates for both measures were detected at weaning; and (ii) alpha diversity at week 15 had strong and negative genetic correlations with BF, -0.53 ± 0.23 to -0.45 ± 0.25, as well as with ADG, -0.53 ± 0.32 to -0.53 ± 0.29. These results suggest fecal microbiota diversity can be used as an indicator trait to improve traits that are expensive to measure.

W984: Swine

Utility of the Affymetrix Axiom Porcine Genotyping Array for Genome-Wide Studies

Catherine W. Ernst, Department of Animal Science, Michigan State University, East Lansing, MI and Scott A. Funkhouser, Genetics Program, Michigan State University, East Lansing, MI

A new high-density genotyping array for the pig, the Affymetrix Axiom™ Porcine Genotyping Array (Axiom_PigHDv1), assays more than 600 thousand SNPs. We have used genotypes from four U.S. pig breeds to show how the greater SNP density of this array can be used to estimate short-range linkage disequilibrium within breeds and short-range persistence of phase across breeds. We have also evaluated how well the array works in conjunction with existing pig SNP genotyping chips by assessing genotype call comparisons between platforms and imputation accuracy when imputing up to 600K density. The Axiom_PigHDv1 provides ample data to estimate average linkage disequilibrium (LD) between SNP pairs that are 1 kb apart or more, enabling estimation of short-range LD decay, which was not possible using lower density pig SNP chips. Genotype calls on the Axiom_PigHDv1 are highly consistent with the Illumina Porcine SNP60 Beadchip, although a small percentage (0.4%) of SNPs show evidence of different reference alleles between platforms. Despite this, the feasibility to impute 600K density from 50K density is relatively high, as we observe average SNP-wise imputation accuracy (R^2) of 0.857. By enabling estimation of short-range LD and short-range persistence of phase, and by working in conjunction with existing pig genotyping platforms, the Axiom_PigHDv1 has great potential for use in genetic association studies and genomic prediction.

W985: Swine

Developing Tools for Evaluating Chromatin Preps for Porcine Functional Genomics

Christopher K. Tuggle[1], Jianzhen Huang[2,3], Hamid Beiki[1] and Martine Schroyen[4], (1)Iowa State University, Ames, IA, (2)Jiangxi Agricultural University, Nanchang, China, (3)Iowa State University, Department of Animal Science, Ames, IA, (4)Université de Liège - Gembloux Agro-Bio Tech, 5030 Gembloux, Belgium

Post-translational covalent modifications on the histones in nucleosomes play important roles in regulating gene expression at the chromatin level. The patterns of epigenetic modifications across the genome can be used to filter SNPs for functional relevance. To study the interplay between RNA expression and histone modification in pig tissues, we are developing experience in liver chromatin preparations. One question we have addressed is methods to test for quality of such preparations using quantitative PCR of specific locations at liver-specific or muscle-specific genes (control) genes we had previously validated using rt-Q-PCR of liver and muscle RNA.

Trimethylation of lysine 36 of histone H3 (H3K36me3) and trimethylation of lysine 4 of histone H3 (H3K4me3) are associated with specific transcriptional states, and we predicted that H3K4me3 marks would be enriched in liver chromatin at liver-specific gene promoters, but not at muscle-specific gene promoters. Likewise, we predicted H3K36me3 would be enriched in the 3'UTR only for liver-specific genes. Liver chromatin preparations from four replicates were immunoprecipitated (ChIP) using antibodies to these two marks. Q-PCR tests (at the promoter or the 3'UTR) for 17 expression-validated genes (9 liver-specific and 8 muscle-specific) showed that the H3K36me3 levels measured at the 3'UTR were high for all liver-specific genes but low for muscle-specific genes. H3K4me3 levels measured at the promoter were high for only liver-specific genes. We have also verified that H3K4me3 peaks map preferentially 5 prime to annotated promoters in the pig genome, and that RNAseq expression level is highly associated with H3K4me3 peaks at gene promoters.

These data indicated that H3K36me3 and H3K4me3 are active marks that correlate well in pigs with their known connections to functional gene components in human and thus are likely to play a vital role in epigenetic control of porcine gene expression. We further show that Q-PCR testing can demonstrate chromatin quality prior to ChIP-seq. Funding acknowledgement: NIFA-AFRI-2011-68004-30336.

W986: Swine

Genome Wide Association Analyses for Dry Cured Ham Related Traits in Italian Duroc Pigs

Francesca Bertolini[1], Giuseppina Schiavo[2], Giuliano Galimberti[3], Samuele Bovo[2], Mariasilvia D'Andrea[4], Maurizio Gallo[5], Luca Buttazzoni[6], Max F. Rothschild[1] and Luca Fontanesi[2], (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, (3)Department of Statistical Sciences, University of Bologna, Bologna, Italy, (4)Department of Agricultural, Environmental and Food Sciences - University of Campobasso, Campobasso, Italy, (5)Associazione Nazionale Allevatori Suini (ANAS), Roma, Italy, (6)CREA Research Centre for Animal Production and Aquaculture, Roma, Italy, Italy

The heavy pig production chain is a very important source of niche pork products, particularly in Southern Europe. In Italy, about 70% of born pigs are raised to produce Protected Designation of Origin dry-cured hams. For this purpose, a specific breeding program designed to improve meat quality for this production has included as key traits the level of intermuscular fat between the leg muscles and ham weight loss during the seasoning period in the Italian Duroc heavy pig breed. These traits may also be of interest to other niche markets, like the Tennessee ham in the US.

In our work, Genome-wide association studies were carried out using Random Residuals of those traits in 573 performance-tested Italian Duroc to find Quantitative Trait Loci (QTL) associated with these traits. The analyses were carried out using a 1 Mb windows based Bayes B approach and a parallel independent single-SNP analyses with a linear mixed model approach. The analyses allow detecting seven QTL on chromosome 2, 4, 6, 7, 8 and 12 for ham weight loss at first salting and three QTL on chromosome 1 and 8 for visible intermuscular fat, that do not overlap with QTL of other production traits normally considered for Italian Duroc evaluation. This first QTL map for these traits in Italian Duroc, in the medium-term, will be useful for developing a more accurate genomic selection program including QTL information.

W987: Synthetic Biology
A Versatile Semi-Synthetic Tool-Fungal Artificial Chromosome (FAC)
Cheng-Cang Charles Wu, Intact Genomics, Inc., St Louis, MO
We have successfully developed a semi-synthetic tool-fungal shuttle bacterial artificial chromosome (BAC) or fungal artificial chromosome (FAC). Combining shuttle BAC/FAC, advanced heterologous host, and chemical analysis-metabolomic scoring (MS), we screened 56 secondary metabolite BGCs from diverse fungal species for expression in *Aspergillus nidulans*. We discovered 15 new metabolites and assigned them with confidence to their biosynthetic gene clusters. Using the FAC-MS platform, we extensively characterized and structure-elucidated a new macrolactone, valactamide A, and its hybrid nonribosomal peptide synthetase–polyketide synthase (NRPS–PKS). The shuttle BAC/FAC tool also enables facile the dissection of biosynthetic gene clusters. The ability to regularize access to microbial secondary metabolites at an unprecedented scale stands to revitalize drug discovery platforms with renewable sources of natural products.

W988: Synthetic Biology
Breaking Self-Incompatibility in Diploid Potato using the CRISPR/Cas9 System
Felix E. Enciso, Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI

W989: Synthetic Biology
Characterizing a Rice Diversity Panel with a 7K SNP Chip and Flowering Time Evaluation
Karina Y. Morales[1], Stephon Warren[1], John Carlos I. Ignacio[2], Yuxin Shi[3], Rodante Tabien[4], Tobias Kretzschmar[2], Susan McCouch[3] and Michael J. Thomson[1], (1)Texas A&M University, College Station, TX, (2)International Rice Research Institute, Los Baños, Philippines, (3)Cornell University, Ithaca, NY, (4)Texas A&M AgriLife Research, Beaumont, TX
Rice (Oryza sativa L.) is an essential food crop with demands for increased yield as it provides the daily caloric intake of over 50% of the world's growing population. Flowering is one of the most sensitive stages of rice growth and is highly variable among varieties and across environments. In Texas, farmers often desire early flowering varieties as these can avoid peak temperatures of the summer months and give sufficient time for the ratoon crop to mature before the cold temperatures of winter begin. This experiment took place at the Texas A&M AgriLife Research Center in Beaumont, TX where 208 rice varieties of diverse origins were planted in spring 2017 and were grown through the summer of 2017. Beginning approximately 50 days after planting, notes were collected once a week on flowering percentage to estimate days to 50% flowering. Each variety was genotyped using the Illumina 7K rice SNP chip developed at Cornell University. This project aims to identify genetic loci which contribute to extremely early and late flowering time. Upon identifying these loci, we will use the CRISPR/Cas9 genome editing system to validate candidate genes in diverse genetic backgrounds to gain a better understanding of how each locus may contribute to days to heading in rice. Ultimately, the improved knowledge on manipulating flowering time genes will lead to more precise tools to provide early flowering in any genetic background for each target environment.

W990: Synthetic Biology
Chinese Tongue Sole: Genome to Breeding
Songlin Chen, Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

W991: Synthetic Biology
Optimization of *Agrobacterium*-Mediated Transformation in Soybean
Shuxuan Li[1], Yahui Cong[1], Yaping Liu[1], Tingting Wang[1], Qin Shuai[1], Nana Chen[1], Junyi Gai[2] and Yan Li[3], (1)Nanjing Agricultural University, Nanjing, China, (2)Soybean Research Institute, National Center for Soybean Improvement, Key Laboratory of Biology and Genetic Improvement of Soybean (Ministry of Agriculture), State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, (3)National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China
High transformation efficiency is a prerequisite for study of gene function and molecular breeding. Agrobacterium tumefaciens-mediated transformation is a preferred method in many plants. However, the transformation efficiency in soybean is still low. The objective of this study is to optimize Agrobacterium-mediated transformation in soybean by improving the infection efficiency of Agrobacterium and regeneration efficiency of explants. The results showed that an infection efficiency of over 96% was achieved by collecting the Agrobacterium at a concentration of OD_{650} = 0.6, then using an Agrobacterium suspension medium containing 154.2 mg/L dithiothreitol to infect the half-seed cotyledonary explants (from mature seeds imbibed for 1 day), and co-cultured them for 5 days. On the other hand, the rates of shoot elongation were compared among six different concentration combinations of gibberellic acid (GA_3) and indole-3-acetic acid(IAA). The shoot elongation rate of 34 and 26% was achieved when using the combination of 1.0 mg/L GA_3 and 0.1 mg/L IAA for Jack Purple and Tianlong 1, respectively. This rate was higher than the other five concentration combinations of GA_3 and IAA, with an 18 and 11% increase over the original laboratory protocol (a combination of 0.5 mg/L GA_3 and 0.1 mg/L IAA), respectively. The transformation efficiency was 7 and 10% for Jack Purple and Tianlong 1 at this optimized hormone concentration combination, respectively, which was 2 and 6% higher than the original protocol, respectively.

W992: Systems Biology and Ontologies
Cross-Species Semantics to Inform Rare Disease Diagnosis and Discovery
Melissa Haendel, Oregon Health & Science University (OHSU), Portland, OR
Correlating phenotypes with genetic variation and environmental factors is a core pursuit in biology and biomedicine. However, numerous challenges impede our progress: patient phenotypes may not match known diseases, candidate variants may be in uncharacterized genes, model

organisms may not recapitulate human or veterinary diseases, and filling evolutionary gaps is difficult. Additionally, many resources must be queried to find potentially significant genotype–phenotype associations. Advanced informatics tools can identify phenotypically relevant disease models in research and diagnostic contexts. Large-scale integration of model organism and clinical research data together can provide a breadth of knowledge not available from individual sources. The Monarch Initiative (monarchinitiative.org) is a collaborative, international open science effort that aims to semantically integrate and curate genotype–phenotype knowledge from many species and sources in order to support precision medicine, disease modeling, and mechanistic exploration. To this aim, Monarch has created an integrated knowledge graph, analytic tools, and web services that enable diverse users to explore relationships between phenotypes and genotypes across species. Recent developments include the release of a new BioLink API, and the creation of a suite of phenotype annotation tools (phenotyper.monarchinitiative.org). The BioLink API allows researchers and clinicians to search for relationships between various types of data from patients and animal models, including genes, variants, diseases, phenotypes, clinical notes, environmental conditions, and metadata. Researchers can also use the Phenotyper tool to obtain Human Phenotype Ontology (HPO) terms from research articles and export those terms as a PhenoPacket (phenopackets.org) for data analytics.

W993: Systems Biology and Ontologies

The Tomato Expression Atlas: A New Platform for Biological Discovery with Cell-Type Resolution

Jocelyn KC Rose, Cornell University, Ithaca, NY

The development of fleshy fruits involves the distinct, but coordinated and synergistic, functions of the multiple differentiated tissues and cell types. However, most biochemical and molecular studies of fruit biology use a homogenized mixture of tissues, which limits insights into cell specialization, and lower abundance molecules that are present only in certain cell types are often diluted below the level of detection. In addition, interior fruit tissues are often not included in fruit analyses. There is therefore a critical 'information void' when it comes to annotating and presenting gene expression data. We have been addressing this challenge in the context of understanding the entirety of gene expression during tomato fruit development, by coupling RNA-seq analysis with laser capture microdissection (LCM), which allows the precise isolation of individual fruit cells/tissue types. In addition to resolving gene expression down to the level of cell/ tissue type, this approach has enabled: *(i)* the identification of previously unannotated genes, demonstrating the value of LCM as a tool for gene discovery; *(ii)* inferences regarding gene functions, based on the patterns of tissue- or cell type-related expression. We have also been developing computed tomography as a non-invasive imaging tool to create a 3D 'virtual tomato', which includes internal structures, to provide digital a scaffold upon which to present transcriptome, or other 'omics' data sets as a 4D display. All data will be publicly accessible in a new database, the Tomato Expression Atlas. This database includes a novel user interface with a correlation matrix that reveals patterns of co-expressed genes at an unprecedented level of spatiotemporal resolution, thereby optimizing the identification of functionally related suites of genes.

W994: Systems Biology and Ontologies

Oligosaccharides from Plant Sources: Linking Bioactive Components to Products, Processes, and Effluents

Tian Tian and Daniela Barile, UC Davis, Davis, CA

Oligosaccharides are targets of new investigations because of their highly specific functions such as acting as prebiotics by feeding beneficial bacteria, blocking attachment of pathogens in the gut, and interacting directly with intestinal cells. Therefore, these non-digestible oligosaccharides have great potential to improve the quality of many foods. However, few natural sources for bioactive oligosaccharides have been identified and used. Yet some plant products and their industrial residues after nutrient extraction are rich in dietary fiber and can provide good sources for bioactive oligosaccharides. This work presents the methods for purification and identification of plant oligosaccharides and their application to coffee beans and spent coffee grounds (residue obtained during the brewing process). The work was expanded to the study of various streams from the nutraceutical industry (including milled chia seed meal, feedstock & side stream of sting nettle root, and citrus fiber), as well as apple pomace as waste products of the apple juicing industry. Over 40 novel oligosaccharides were identified through mass spectrometry and a comprehensive oligosaccharides library was established. The observed variations among the identified structures are linked to products, processes, and effluents of different types of food matrices.

W995: Systems Biology and Ontologies

Data-Driven Plant Breeding – Getting the Most from Your Resources: Strategies for Leveraging Big Data for the Genetic Improvement of Rice and Sorghum

Jennifer Spindel, DOE Joint Genome Institute, Walnut Creek, CA

In the field of crop genomics, it is easier than ever to generate highly dense genotype and phenotype datasets on plant populations, however, our ability to generate large datasets has not kept pace with the application of data driven science to breeding programs, particularly for crops such as rice and sorghum. Genomic selection (GS), in particular, holds great promise for improving rice breeding efficiency, but has not yet been widely deployed among public rice breeding programs. In this talk, I will review the promise GS holds for rice breeding, and present strategies that would help breeders and geneticists to get the most from their data, including the benefits that could be seen from widespread data sharing among genomic assisted breeding programs.

W996: Systems Biology and Ontologies

Translational Dynamics of Gene Expression in the Maize Leaf

Indrajit Kumar, Donald Danforth Plant Science Center, Saint Louis, MO

Quantification and comparison of gene expression by RNAseq is one of the most frequently used techniques in a majority of biological studies. However, RNAseq serves only as a proxy for gene expression as it is based on the assumption that transcript levels directly correlate to that of protein. A small but significant number of transcripts could also be under translational control. The maize leaf develops from the tip to the base and a comparative leaf gradient study provides an opportunity to monitor the process of photosynthetic differentiation. We utilized RNAseq and the recently developed Ribo-Profiling technique to understand the relationship between transcriptional and translational dynamics of nuclear genes along 4 predefined sections of maize leaf (Base, sink-source transition, maturing and Tip). Our analysis revealed more than 1000

genes with significant change in translational efficiency (TE). We have also investigated various potential features/mechanisms that could be involved in translational regulation of these genes.

W997: Systems Biology and Ontologies

Plant Reactome Database: A Portal for Plant Pathways Resources

Sushma Naithani[1], Parul Gupta[1], Justin Preece[1], Peter D'Eustachio[2], Justin L. Elser[1], Antonio Fabregat[3], Doreen Ware[4], Lincoln Stein[5] and Pankaj Jaiswal[1], (1)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, (2)NYU School of Medicine, New York, NY, (3)European Bioinformatics Institute, Hinxton, United Kingdom, (4)USDA ARS, Cold Spring Harbor, NY, (5)Ontario Institute of Cancer Research, Toronto, ON, Canada

Plant Reactome (http://plantreactome.gramene.org/), a pathway and network portal of Gramene database, employs the structural framework of a plant cell to show metabolic, genetic, transport, developmental and signaling pathways for 75 species including crops and model plants. Plant Reactome features *Oryza sativa* (rice) as a reference species for the curation of pathways based on published literature and orthology driven functional annotation of the genes. The curated rice pathways are used for generating homology-based projections of pathways for additional 74 plant species. Plant researchers can i) search and browse various components of the database, ii) compare projected pathways with those from the reference species to identify potential gaps in projection or biological differences, iii) visualize curated baseline and differential gene expression data in the context of pathways, and iv) upload and analyze Omics datasets generated in their laboratories to identify differentially expressed pathways and associated genes. The presentation will discuss the development of the Plant Reactome and its utility for omics data analysis and visualization. The project is supported by the Gramene database award (NSF IOS-1127112), the NIH (U41 HG003751), ENFIN (LSHG-CT-2005-518254), the Ontario Research Fund, and the EBI Industry Programme.

W998: Systems Genomics

Chemical Genetics Dissection of Interference between Pathogen and Drought Stress Tolerance Signaling in Plants

Jiyoung Park, University of California, San Diego, La Jolla, CA

The plant hormone abscisic acid regulates adaptation to environmental stresses, particularly drought. How plants cope with multiple stresses, especially when challenged with pathogen infection and then drought, remains largely unknown. The tolerance mechanisms against the two stresses often negatively affect each other. However, the underlying mechanisms remain unknown. Using a chemical genetics approach that can address genetic redundancy and network robustness, a novel small molecule “DFPM” was identified that down-regulates abscisic acid signaling by activating plant immune responses. To dissect this interference signaling, an Arabidopsis thaliana reporter line harboring an ABA-inducible marker pRAB18:GFP was EMS mutagenized and screened for hyposensitive responses to DFPM. *rda* (***r***esistant to ***D***FPM inhibition of ***A***BA signaling) mutants were isolated and mapped to a putative receptor-like kinase. Further characterization of the functions of this receptor-like kinase in plant immune signaling and interference mechanisms with ABA signaling will be presented. This research will help understand how plants exposed to both pathogen and drought can coordinate effective tolerance responses which will be relevant for plant survival and crop yield.

W999: Systems Genomics

Molecular Mechanisms of Ginsenoside Biosynthesis Revealed by Genome-Wide Systems Analysis of the Genes Significantly Associated with Ginsenoside Contents in *Panax ginseng* C.A. Meyer

Mingzhu Zhao[1], Xiangyu Li[1], Yanping Lin[1], Yanfang Wang[2], Jingjing Qin[1], Yilai Han[1], Kangyu Wang[1], Chunyu Sun[1], Rui Yin[1], Yi Wang[1], Hong-Bin Zhang[3] and **Meiping Zhang**[1,3], (1)College of Life Science, Jilin Agricultural University, Changchun, China, (2)College of Chinese Medicinal Materials, Jilin Agricultural University, Changchun, China, (3)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX

Ginseng, *Panax ginseng* C.A. Meyer, is one of the most precious medicinal herbs for human health and medicine in which ginsenosides are known to play critical roles. The cytochrome P450 (*PgCYP*)-, oxidosqualene cyclase (*PgOSC*)- and UDP-glycosyltransferase (*PgUGT*)-encoding gene families have been shown to play important roles in ginsenoside biosynthesis, but only a few of genes have been so far cloned from each of the families in ginseng. Here we report genome-wide identification of the genes from these gene families significantly correlated in expression activity with the variation of ginsenoside contents and the molecular mechanisms underlying ginsenoside biosynthesis by systems analysis of these ginsenoside biosynthesis-related genes. We found that at least 100 *PgCYP* genes, 7 *PgOSC* genes and 44 *PgUTG* genes of these gene families, including the published ginsenoside biosynthesis genes previously cloned from the families, were significantly correlated in expression activity with the variations of nine mono- and total-ginsenoside contents in the roots of a number of different cultivars. Mutation analysis and gene regulation analysis both have further confirmed the association of these genes with ginsenoside biosynthesis. Therefore, we systems analyzed these genes in several aspects. These genes formed a co-expression network and functioned in a correlative manner, even though they belong to three different families, are categorized into a diversity of functional categories and express differentially. These results not only led to identification of a number of new candidate genes involved in ginsenoside biosynthesis, but also provided a genome-wide insight into the molecular mechanisms underlying ginsenoside biosynthesis in ginseng.

W1000: Systems Genomics

Novel Unilateral Incompatibility in *Brassica rapa* is Regulated by Duplicated Self-Incompatibility Genes, *PUI1* and *SUI1*

Masao Watanabe, Tohoku University, Sendai, Japan

Most of flowers pollinated by insects prevent the undesirable pollen from different species. Within species, for maintain the genetic diversity, several outcrossing systems, dioecy, dichogamy and self-incompatibility (SI), etc., are established. About half of plants have SI trait. The SI is defined as the inability of a fertile hermaphrodite plant to produce zygotes after self-pollination. In *Brassica* species, this SI trait is sprophytically regulated by a single locus with *S*-multiple-alleles. To date, SP11 (small cysteine-rich protein) and SRK (receptor-type protein kinase) have been identified as the male and female *S* determinants, respectively. These molecules have a role in the pollen-pistil interaction of SI. In the case of crossing between the plants with the same *S*-alleles, incompatibility phenotype is observed.

We have already identified novel unilateral incompatibility (UI) phenomenon between Japanese and Turkish lines within *B. rapa*. Even if combination between plants with different *S*-alleles, pollination of Turkish pollen on Japanese pistil leads to UI, whose phenotype is similar to SI.
In order to dissect this UI, we performed genetic mapping and genome analysis. A set of genes homologous to *SP11* and *SRK* was identified at the *UI* locus, which might have been duplicated from the different chromosome. PUI1 and SUI1, which were demonstrated as pollen and pistil determinants, respectively, should trigger the UI between Japanese and Turkish lines. In this presentation, we will show the functional and phylogenetic analyses of this UI.

W1001: Systems Genomics
Phylogenomics and Systems Biology Approaches Reveals Conserved Adaptive Processes in Atacama Desert Plants
Soledad Undurraga[1], Daniela Soto[1], Kranthi Varala[2], Gil Eshel[3], Erika Viviana Araus[1], Francisca Diaz[1], Gabriela Carrasco[1], Chase W. Nelson[4], Dennis Wm. Stevenson[5], Gloria Coruzzi[3], Claudio Latorre[6] and **Rodrigo A. Gutiérrez**[1], (1)FONDAP Center for Genome Regulation. Millennium Nucleus Center for Plant Systems and Synthetic Biology. Department of Molecular Genetics and Microbiology. Pontificia Universidad Católica de Chile, Santiago, Chile, (2)Purdue University, West Lafayette, IN, (3)New York University, New York, NY, (4)5. Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, (5)New York Botanical Garden, Bronx, NY, (6)Ecology Department. Pontificia Universidad Católica de Chile, Santiago, Chile
The Atacama Desert in Chile is the oldest and driest desert on Earth. Despite being one of the harshest environments in the world, plants have colonized and adapted to its extreme abiotic conditions. In order to identify genes underlying the adaptive traits of these “extremophile” plants, we developed a systems biology and phylogenomics pipeline that uses transcriptome data to annotate genes, determine evolutionary relationships, identify orthologues, and measure natural selection. Our approach identifies the evolutionary divergence of “extremophile” Atacama plants by comparing their sequences to publicly available phylogenetically-related (“sister”) species, which are not adapted to Atacama’s extreme conditions. We identified, collected, and sequenced the transcriptome of 32 plant species within an altitudinal transect that spans the limits for life in Atacama. All these species are seed plants, with a preponderance of Angiosperms. An overall set of 70 species (32 from Atacama, 32 sisters, and 6 model plants) was processed, generating nearly 1.7 million predicted proteins and over 30 thousand orthologue families. The resulting phylogenomic tree contained twenty nodes that account for the divergence of extremophile Atacama plants from their non-Atacama sister species. We identified thousands of genes that provide support for independent origins of environmental adaptation. We identified a subset of genes that gave recurring support (independent origins). This subset was enriched (p-value < 0.05) in processes related to response to stress, response to radiation, photosynthesis, and nitrogen compound metabolic processes, among others. Data were further used to identify key genes involved in adaptation to marginal soils. Functional characterization of key candidate genes is ongoing in selected plant model systems.

W1002: Systems Genomics
Functional Genomics on Medicinal Plants
Amit Rai[1], Mami Yamazaki[1] and Kazuki Saito[1,2], (1)Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan, (2)Metabolomics Research Group, RIKEN Center for Sustainable Resource Science, Yokohama, Japan
In the last decade, we have seen functional genomics playing a major role towards holistic understanding of different biological processes in agronomically important crops, and have contributed towards increase in food production and to meet upcoming ecological and environmental challenges. On the other hand, medicinal plants, producing valuable bioactive metabolites, are yet to see this revolution, primarily due to the lack of high quality whole genome sequence resources. In recent years, next-generation sequencing technologies have made available genomic datasets of thousands of medicinal plants in the form of transcriptome assemblies or whole genome sequencing datasets, and are playing an important role in understanding biosynthesis and regulation of specialized metabolites in-planta (Plant J. 2017 May; 90(4):764-787). Here, I will present my laboratory recent results and on-going research using functional genomics to understand biosynthesis of bioactive metabolites. Beginning from the studies on model plant Arabidopsis, I will show how the tools and analysis pipelines developed can be effectively used to mine potential candidate genes involved in the biosynthesis of specialized metabolites for functional characterization. I will then bring through several of our recent studies using transcriptome profiling, comparative transcriptome analysis, integrative-omics analysis and multi-variant regression modeling on omics datasets of medicinal plants to elucidate biosynthesis of key specialized metabolites. I will also discuss our on-going and recently published whole genome sequencing of medicinal plants and our strategy on genome mining. Towards the end, I will bring though our on-going study and recent results on elucidating camptothecin biosynthesis and regulation.

W1003: Systems Genomics
Accurately Dissecting Genetic Effects for Complex Plant Trait Analysis
Patrick X. Zhao, Noble Research Institute, Ardmore, OK
Complex plant traits, such as yield, tolerance to abiotic and biotic stresses, are often governed by many individual genes (G), the gene-gene interactions (GxG) and gene-environment interactions (GxE). Therefore, analysis of complex plant trait demands accurately dissecting these genetic causal effects, i.e. accurately identifying individual genes (loci), gene-gene and gene-environment interactions that are consistently associated with the observed phenotypes. In this report, we describe our recently developed a trio of genotype-phenotype association analysis tools, namely 1) GWASPRO (bioinfo.noble.org/GWASPRO/), which adopts a simple linear mixed model (LMM) for the analysis of *additive* genetic effects (referred as one-dimensional mapping or 1D GWAS here) and is specially optimized for the analysis of “big data” generated from large-scale genome-wide association studies (GWASs); 2) PEPIS (bioinfo.noble.org/PolyGenic_QTL/), which adopts a full polygenic linear mixed model to analyze the *additive* (1D GWAS), *dominance effects* (1D GWAS) and *epistatic effects* such as *additive* x *additive*, *additive* x *dominance*, *dominance* x *additive*, *dominance* x *dominance* (referred as two-dimensional mapping or 2D GWAS here) in GWASs and quantitative trait loci mapping; and 3) PATOWAS (bioinfo.noble.org/PATOWAS/), which further extends the 2D GWAS LMMs for broader associative ‘omics’ studies, i.e. the LMMs can not only be applied to GWASs, but also transcriptomics-wide association studies

(TWASs) and metabolomics-wide association studies (MWASs). Our case analysis of a set of publically available Immortalized F2 (IMF2) associative 'omics' studies, which consisted of comprehensive GWASs, TWASs and MWASs, demonstrated the high performance of our developed LMMs and tools, enabling genotype-phenotype association discovery and genetic variances analysis for complex plant traits.

W1004: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Statistical Tests on Genetic Hypotheses
Zhenyu Jia, University of California Riverside, Riverside, CA
Important genetic laws, such as Mendelian inheritance patterns, were discovered and proved using statistical hypothesis testing. Although the history of genetics has been woven with the application of statistical methods, it is challenging for instructors without a strong statistics background to teach these methods in introductory genetics classes. 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 undergrads in my genetics classes.

W1005: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Sequencing to Success: Engaging Community College Students in Life Sciences Research Projects
Alejandro Cortez, University of California Riverside, Riverside, CA

W1006: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Can You Hear the Buzz? Incorporating an Authentic Research Experience on Bee-Associated Micro Biomes in Non-Majors Introductory Biology at the University of California, Riverside
Katie Burnette, University of California, Riverside, Riverside, CA

W1007: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Teaching Genetics One Mutant at a Time
Jelena Brkljacic, The Arabidopsis Biological Resource Center, Center for Applied Sciences, The Ohio State University, Columbus, OH

W1008: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
FPsc: A Novel Plant-Based (rapid-cycling *Brassica rapa*) Model System to Help Students Connect Mutant *Phenotypes* with Underlying, DNA Sequence-Based *Genotypes*
Scott Woody, UW-Madison, Madison, WI

W1009: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Science Tools in the Classroom: Partnering with Teachers in Bioinformatics
Joann Mudge, National Center for Genome Resources (NCGR), Santa Fe, NM
The "Science Tools in the Classroom" (STC) project, funded by NIH SEPA, partners with all levels of public school teachers to enhance life science content knowledge, provide bioinformatics classroom activities, and help teachers develop their own bioinformatics classroom activities, thereby giving their students a first glimpse of bioinformatics. Bioinformatics is a cross-cutting tool that incorporates both biology and computing, and can be applied across many different biological disciplines. It is, therefore, easy to find "hooks" to engage student interest and create relevant and authentic activities. Engagement and relevancy, in turn, have been shown to increase student interest in STEM careers. Furthermore, the next generation sequencing revolution has made DNA sequencing inexpensive enough that all students can expect encounter it in their own lives, whether at the doctor's office or through personal genome sequencing. But education is lagging behind. By partnering with teachers through professional development workshops, we aim to help fill in this gap. We introduce teachers to DNA and bioinformatics content, areas and skills that were often limited or absent in their training. We teach with and provide DNA and bioinformatics case studies for classroom use that are age appropriate for their students. Finally, we assist teachers as they create their own DNA and bioinformatics case studies. This allows them to take ownership, choose topics that interest their students and that are complementary to their curriculum, incorporate science standards and fill in gaps in their curriculum. This allows the teachers to teach their students about DNA and bioinformatics in an engaging, active learning approach, exposing students to a field that is both highly relevant and that develops marketable skills needed across the biological sciences and other STEM fields.

W1010: The Analysis and Role of the Microbiome
Host Genotype Control of the Maize Microbiome
Maggie R. Wagner, North Carolina State University, Raleigh, NC
Genetically enhancing plants to attract more beneficial microbes from their environment is a promising avenue for sustainable agriculture research. Recent research has shown that plant-associated microbiomes are heritable and can dramatically alter host health. However, most studies to date have used arbitrarily chosen genotypes and descriptive approaches that only minimally improve our ability to predict how microbiomes might respond to specific genetic changes in the host. To address this problem, I apply predictions from quantitative genetics to analyze microbiome composition in field-grown maize lines from various stages of real, ongoing breeding programs.
First, we present experimental evidence that the introgression of disease-resistance alleles into elite inbred lines has side effects on non-pathogenic microbial endophytes. We compared leaf microbiome composition of a disease-susceptible inbred line to NILs with introgressed multiple-disease-resistance (MDR) loci. Fungal microbiome diversity was 7% to 40% higher in MDR NILs relative to their disease-susceptible inbred parent. Over 16% of leaf-associated fungi responded to the introgression of MDR alleles.

Second, we describe ongoing experiments to compare the leaf- and root-associated microbiomes of F1 hybrids to those of their inbred parental lines. Despite the tremendous value of hybrid vigor, and although hybridization is the basis of many breeding techniques, it is unknown whether microbiome composition exhibits heterosis like many other quantitative traits in maize. High-throughput amplicon sequencing, shotgun metagenome sequencing, and re-inoculation experiments will test whether the maize microbiome contributes to the superior performance of hybrid lines. Together, these experiments will improve our ability to predict interactions between plant genomic variation, microbiome diversity, and crop health.

W1011: The Analysis and Role of the Microbiome
Harvesting the Root Microbiome of Grasses Toward Sustainable Increase Crop Production
Tatiana S. Mucyn, University of North Carolina at Chapel Hill, Chapel Hill, NC
The global demand for increased food commodities and energy supply highlights the necessity to further improve crop productivity. A variety of plant functions and traits are co-dependent on the microbial communities that exist within and around them. Characterizing, manipulating and /or mimicking the plant microbiota provide a promising path toward increasing crop production. Grasses constitute our main source of food, feed, and bioenergy, but how their microbiomes relate to, or may differ from other plant microbiomes remain unclear. We have compared the well-studied bacterial root microbiome of *Arabidopsis thaliana*, with those of two monocot species *Brachypodium distachyon* and *Setaria viridis* from plants grown in the same wild soil, using high-throughput bacterial 16S rDNA, as well as fungal ITS2 profiling. We determined that the root microbiome is relatively conserved across plant species. However it remains to be assessed if the microbiome may also present functional conservation across species. To better assess the relationship between the microbiome and the plant nutritional stress response, and increase our odds to identify bacterial strains promoting plant health under nutrient stress we established a field experiment is which Setaria was grown under low and high Nitrogen side by side with a Sorghum accession. We propose to establish a microbial transplant experiment to identify microbes with beneficial effect on plant growth.

W1012: The Analysis and Role of the Microbiome
The Cattle Rumen Microbiome: Insights from Metagenome Assembled Genomes
Robert Stewart, Roslin Institute, Edinburgh, United Kingdom

W1013: The Analysis and Role of the Microbiome
Shedding Light on Microbial Dark Matter: Lessons Learned from Annotating Novel Microbial Genomes from Metagenomes
Mick Watson, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom
Mick Watson, The Roslin Institute, University of Edinburgh
Recent advances in bioinformatics have enabled the rapid assembly of genomes from metagenomes (MAGs), and there is a need for reproducible pipelines that can annotate and characterise thousands of genomes simultaneously. Here we present MAGpy, a Snakemake pipeline that takes FASTA input and compares MAGs to several public databases, checks quality, assigns a taxonomy and draws a phylogenetic tree.

W1014: The Analysis and Role of the Microbiome
Solving the Metagenome Assembly Problem with Hi-C
Ivan Liachko, Zev Kronenberg, Andrew Wiser, Maximilian Press, Kaylee Mueller and Shawn Sullivan, Phase Genomics, Seattle, WA
Sequencing metagenomic samples yields large numbers of disconnected DNA sequence fragments. Without a direct source of information on which sequences originated from the same cells, typical metagenomes have to be binned using indirect measures of relatedness, such as copy number and nucleotide composition. These methods have limited power and accuracy at reconstructing individual genomes from microbiome samples and fail at reconstructing genomes with multiple chromosomes and plasmids.
We have developed ProxiMeta, a Hi-C-based metagenomic platform that enables significant improvements in the assembly of genomes from mixed populations. The Hi-C method creates proximity-ligated DNA junctions within intact cells. The resulting *in vivo* proximity-ligation data provides direct evidence as to which sequences originated from the same cell and, therefore, same strain and species. ProxiMeta utilizes Hi-C data to yield hundreds of genomes from diverse microbiome samples with neither culturing nor *a priori* information, making it ideal for microbial discovery efforts. A unique advantage of this system is its ability to accurately associate plasmids with their host genomes and to assemble genomes with multiple chromosomes.
We show that this method drastically outperforms normal binning tools in terms of both increased numbers of recovered genomes as well as the quality of resultant genome assemblies. We also show that ProxiMeta is versatile and applicable to very diverse sample types, including fecal, agricultural, and environmental samples.

W1015: Translational Genomics: Redesigning genetic improvement projects
Maximizing Quantitative Traits in the Mating Design Problem
Susan Hunter, Purdue University, West Lafayette, IN

W1016: Translational Genomics: Redesigning genetic improvement projects
Genetic Algorithm for Variable Selection, with Application to Breeding
Shizhong Xu, Department of Botany & Plant Sciences, University of California, Riverside, CA
Genetic algorithm (GA) is not a procedure for genetic experiments; rather, it is an optimization procedure in computer science and operations research. Genetic algorithm is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary

algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as random mating, genetic drift, mutation, crossover and selection. Variable selection is important in the big data era because the majority of variables are simply noise. Genetic algorithm is an ideal tool to separate signals from noises. In this tutorial, I first introduce the concept of genetic algorithm and then demonstrate how to use GA to select variables in linear regressions involving a large number of predictors. Finally, I report an application of GA to select markers in a best linear unbiased prediction (BLUP) model to predict genetic values of hybrid rice.

W1017: Translational Genomics: Redesigning genetic improvement projects
Comparing Genomic Selection Approaches in a Plant Breeding Simulator
Lizhi Wang, Iowa State University, Ames, IA
Conventional genomic selection approaches use a truncation selection strategy based on the breeding values (e.g., GEBV, WGEBV, and OHV) of the plant individuals. More recently, new genomic selection approaches use a group-based metric (e.g., GB and OPV) to select the optimal group of breeding parents rather than optimal individuals. A limitation of these new approaches is the static nature of the metric, i.e., the metric is not updated over time to reflect the evolution of the plant genomes. We present a new genomic selection approach that maximizes the probability to produce an outstanding progeny by a certain project deadline, and the probability is dynamically calculated and updated in every generation. To test the effectiveness of this new approach, we developed a plant breeding simulator and used it to compare the performance of the new approach with conventional ones.

W1018: Translational Genomics: Redesigning genetic improvement projects
Multi-Objective Optimized Breeding Strategies
Deniz Akdemir[1], William D. Beavis[2], Asheesh K. Singh[2], Roberto Fritsche Neto[3] and Julio Isidro-Sanchez[4], (1)StatGen Consulting, Ithaca, NY, (2)Iowa State University, Ames, IA, (3)Universidade de São Paulo, NA, Brazil, (4)University College Dublin, Dublin, Ireland
The purpose of most plant and animal breeding programs is to make decisions that will lead to sustainable genetic gains in more than one traits while controlling the amount of co-ancestry in the breeding population. The decisions at each cycle in a breeding program involve multiple, usually competing, objectives; these complex decisions can be supported by the insights that are gained by applying multi-objective optimization principles in breeding. Multi-objective optimization is an emerging field in mathematical optimization which involves optimization a set of objective functions simultaneously. The discussion in this talk includes the definition of several multi-objective optimized breeding approaches and the comparison of these approaches with the standard multi-trait breeding schemes such as tandem selection, culling and index selection. Proposed methods will be demonstrated with two empirical data sets and with simulations.

W1019: Transposable Elements
Transposable Elements Facilitate Adaptive Genome Evolution in Plant Pathogenic Microbes
Michael Seidl, Wageningen University, Wageningen, Netherlands
Transposable elements (TEs) can impact genome structure and function, for example by inducing gene deletions or altering gene expression leading to reduced fitness. Nevertheless, TE-induced genomic variability has been associated with adaptive genome evolution in many animals, plants and microbes – in particular under stress conditions. Many pathogenic microbes have evolved genomes with highly variable regions enriched in TEs and pathogenicity-related genes. We study the impact of TEs on adaptive genome evolution in the fungal genus *Verticillium* that contains economically and ecologically important plant pathogens, among which *Verticillium dahliae* is the most notorious pathogen that causes vascular wilt disease on >200 plant species. We have recently unraveled the impact of TEs on adaptive genome evolution in *V. dahliae*. Segmental duplications extensively occur in TE-rich genomic regions, which evolve by genomic rearrangements mediated by erroneous double-strand breaks, often over TEs. In contrast to the remainder of the genome, these regions are also enriched in 'active' TEs that further contribute to local plasticity, for example by disruptions of pathogenicity-related genes likely to avoid recognition by plant immune receptors. Comparing the evolutionary dynamics of TEs within the genus *Verticillium* revealed extensive expansions of distinct TE families in few *Verticillium* species leading to significant differences in TE content. In particular in TE-rich *Verticillium* genomes, TEs co-localize in distinct genomic regions, similar to *V. dahliae*, suggesting that TE-rich regions in *Verticillium* are cradles for adaptive genome evolution. Notably, plants similarly have genomes with TE-rich regions in which plant immune receptors reside, and thus TEs direct the coevolution of plants and their pathogenic microbes.

W1020: Transposable Elements
Patterns and Consequences of Subgenome Differentiation Provide Insights into the Nature of Paleopolyploidy in Plants
Meixia Zhao, Biao Zhang, Damon Lisch and Jianxin Ma, Purdue University, West Lafayette, IN
Polyploidy is an important feature of plant genomes, but the nature of many polyploidization events remains to be elucidated. Here, we demonstrate that the evolutionary fates of the subgenomes in maize (*Zea mays*) and soybean (*Glycine max*) have followed different trajectories. One subgenome has been subject to relaxed selection, lower levels of gene expression, higher rates of transposable element accumulation, more small interfering RNAs and DNA methylation around genes, and higher rates of gene loss in maize, whereas none of these features were observed in soybean. Nevertheless, individual gene pairs exhibit differentiation with respect to these features in both species. In addition, we observed a higher number of chromosomal rearrangements and higher frequency of retention of duplicated genes in soybean than in maize. Furthermore, soybean "singletons" were found to be more frequently tandemly duplicated than "duplicates" in soybean, which may, to some extent, counteract the genome imbalance caused by gene loss. We propose that unlike in maize, in which two subgenomes were distinct prior to the allotetraploidization event and thus experienced global differences in selective constraints, in soybean, the two subgenomes were far less distinct prior to polyploidization, such that individual gene pairs, rather than subgenomes, experienced stochastic differences over longer periods of time, resulting in retention of the majority of duplicates.

W1021: Transposable Elements

Epigenetic Control of Transposable Elements during Plant-Cyst Nematode Interactions

Tarek Hewezi, University of Tennessee, Knoxville, TN, USA,, Knoxville, TN

Transposable elements (TEs) are mobile DNA molecules that constitute significant portion of plant genomes. Recent studies have shown that TEs are epigenetically controlled and that this mechanism of control may contribute to the transcriptional activity of adjacent genes. The impact of DNA methylation and transcriptional activity of TEs on the expression of neighboring genes during plant-cyst nematode interaction was illustrated. Infection of Arabidopsis plants with beet cyst nematode (BCN, *Heterodera schachtii*) triggered differential DNA methylation in certain TEs families in a sequence context–specific fashion, particularly those TEs located nearby protein-coding genes. Gene expression quantification revealed an association between BCN-induced hypomethylation in TEs and low expression of proximal genes, which are enriched for secondary metabolic processes. A significant number of the differentially expressed TEs was found to locate nearby protein-coding genes known to change expression in the nematode feeding cells. Collectively, these data suggest that DNA methylation and differential expression of TEs may contribute to transcriptome reprogramming of nematode feeding cells.

W1022: Transposable Elements

Assessing Genome Assembly Quality Using the LTR Assembly Index (LAI)

Shujun Ou, Michigan State University, East Lansing, MI

W1023: Transposable Elements

Transposable Element Contributions to Dynamics of the Maize Genome and Transcriptome

Nathan M. Springer[1], Sarah N. Anderson[1], Jaclyn Noshay[2] and Michelle Stitzer[3], (1)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (2)University of Minnesota, St. Paul, MN, (3)University of California, Davis, Davis, CA

Many eukaryotic genomes are replete with transposable elements (TEs). These TEs have drastically influenced genome size and organization in crop species. There is also evidence for substantial polymorphism for TE insertion sites. Recent de novo genome assemblies of multiple maize genotypes provide the opportunity to survey the extent of variation for TE insertions in multiple maize genomes. While there are many conserved insertions, we find extensive variation in specific insertion sites between any two inbreds. This variability in TE content has the potential to create natural variation for gene expression through genetic or epigenetic influences. The analysis of TE family expression has revealed numerous families with tissue-specific or environment-specific expression. In some cases, the expression patterns of TEs can be associated with similar expression patterns for nearby genes. The TEs also create alterations in chromatin at the insertion site. We have been working to document the stability of TE-associated chromatin changes across multiple genotypes to understand whether TEs might contribute to partially heritable changes in the expression of nearby genes.

W1024: Transposable Elements

Replicative Transposition of the Miniature Inverted Repeat Transposable Element *mPing*

C. Nathan Hancock[1], Lisette Payero[2], David Gilbert[2] and Edward V. McAssey[3], (1)University of South Carolina AIken, Aiken, SC, (2)University of South Carolina Aiken, Aiken, SC, (3)Center for Applied Genetic Technologies, University of Georgia, Athens, GA

mPing is a miniature inverted repeat transposable element from the *PIF/Pong/Harbinger* superfamily that actively transposes in some rice cultivars. Like other DNA transposable elements, *mPing* is mobilized using a cut-and-paste mechanism which does not necessarily result in additional copies, however *mPing* has increased in copy number over time in rice. We hypothesize that replicative transposition of this element results from homologous repair of the excision site with a homologous *mPing* containing template.

We are testing this hypothesis using a previously developed yeast transposition assay that produces colonies when repair of the excision site does not include *mPing* (non-replicative transposition). Using genomic *mPing* reporter constructs we found that the number of colonies formed in a haploid strain with no possibility of replicative transposition was significantly higher that the number of colonies formed in diploid yeast that can perform homologous recombination. This result is consistent with our hypothesis and is also supported by our finding that a diploid with only one *mPing* reporter copy and a knocked out homologous sequence produces very low number of colonies (likely due to homologous repair with the non-functional template). To confirm the role of homologous repair, we performed transposition assays in *rad51* deficient strains of yeast that cannot perform homologous repair. This experiment produced equal numbers of colonies for haploids and single reporter copy diploids. This result confirms that homologous repair is responsible for the observed differences in the yeast transposition assays. We are currently in the process of sequencing the resulting colonies to detect replicative transposition events.

W1025: Tripal Database Network and Initiatives

Cucurbit Genomics Database: Integration Genetic and Genomics Resources for Cucurbit Breeding

Yi Zheng[1], Shan Wu[1], Rebecca Grumet[2], Amnon Levi[3], Yiqun Weng[4,5], Michael Mazourek[6], James D. McCreight[7], Umesh K. Reddy[8], Nurit Katzir[9], Jordi Garcia-Mas[10] and Zhangjun Fei[1,11], (1)Boyce Thompson Institute, Cornell University, Ithaca, NY, (2)Department of Horticulture, Michigan State University, East Lansing, MI, (3)USDA-ARS, Charleston, SC, (4)Department of Horticulture, University of Wisconsin, Madison, WI, (5)USDA-ARS, Vegetable Crops Research Unit, Madison, WI, (6)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (7)USDA-ARS, Crop Improvement and Protection Research, Salinas, CA, (8)Department of Biology, West Virginia State University, Institute, WV, (9)Agricultural Research Organization, Dept. of Genetics and Vegetable Crops, Newe Ya'ar Research Center, Ramat Yishay, Israel, (10)IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB, Carretera de Cabrils Km2, 08348 Cabrils, Barcelona, Spain, (11)USDA-ARS Robert W. Holley Center for Agriculture and Health, Ithaca, NY

The recent genome sequence assemblies of major cucurbit crop species, including cucumber, melon, watermelon and pumpkin have made it feasible to use advanced genomic approaches in cucurbit breeding. Under the CucCAP project (http://cuccap.org), the Cucurbit Genomics

Database (CuGenDB; http://cucurbitgenomics.org) will be a critical resource for accelerating the breeding progress of cucurbit crops. To manage, store, distribute and analyze the large amount of recently generated genotype and phenotype data, we used Tripal, with Chado backend database schema, as a solution to integrate genetic, genomics, transcriptomics and other biological data into CuGenDB. The CuGenDB is currently implementing various extension modules of Tripal such as BLAST UI, Analysis Unigene, Blast Analysis, InterProScan Analysis, and Daemon. Other modules that will be implemented in the database include Elasticsearch and Breeding API. The database also incorporates custom modules for gene classification, ontology enrichment analysis, synteny block visualization among different genomes, and exploration of expression data of mRNA and small RNAs. These custom modules will be shared with the Tripal community.

W1026: Tripal Database Network and Initiatives

Updates on Tripal Mapviewer and the Tripal Breeding Information Management System (BIMS)

Dorrie Main[1], Katheryn Buble[2], Taein Lee[3], Sook Jung[1], Jodi L. Humann[1], Jing Yu[1], Stephen P. Ficklin[1], Ksenija Gasic[4] and B. Todd Campbell[5], (1)Washington State University, Pullman, WA, (2)Washington State University, Pulmman, WA, (3)WSU, Pullman, WA, (4)Clemson University, Clemson, SC, (5)USDA-ARS, Florence, SC

We report progress on two Tripal modules, Tripal MapViewer and Tripal BIMS. Tripal MapViewer 1.0 provides an interactive visualization for genetic maps. Similar in functionality to the existing GMOD-CMap map comparison tool, TripalMap offers the benefit of using map data directly stored in GMOD-Chado, a common database schema for Tripal. New features in TripalMapViewer 1.0 include displaying corresponding linkage groups from different maps and generating dotplots and a matrix table of the corresponding maps. In addition, the admin page allows configuration of a custom query for building materialized views, which provides both better performance and flexibility in the way data is stored in Chado. TripalMap MapViewer 1.0 is available at Cool Season Food Legume Database (https://www.coolseasonfoodlegume.org/MapViewer) with source code available at https://github.com/tripal/tripal_map/releases/latest. The Breeding Information Management System in Tripal (Tripal BIMS), will allow individual breeders to integrate their 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 provide faster and flexible data entry. The new functionality of TripalBIMS allows breeders to generate input files for Field Book as well as to upload data collected using the App. In addition, the current TripalBIMS allows breeders to manage user access to their breeding program, to import breeding data using web forms or template file loader, search and download data, view the statistical data for a specific cross or trial and generate various lists to use in other tools. The files that they have uploaded or generated can be saved in their account. TripalBIMS has been implemented in two databases, GDR (https://www.rosaceae.org) and CottonGEN (https://www.cottongen.org).

W1027: Tripal Database Network and Initiatives

The Tripal Gateway Project: Supporting Exchange, Transfer and Analysis of Large-Scale Data for Online Biological Databases

Stephen P. Ficklin[1], Chun-Huai Cheng[1], Shawna Spoor[1], Ming Chen[2], Abdullah Almsaeed[3], Bradford Condon[4], Nick Mills[5], Nick Watts[6], Connor Wytko[1], Emily Grau[7], Nic Herndon[7], Brian Soto[1], Sook Jung[1], Kuang-Ching Wang[6], Alex Feltus[8], Margaret Staton[9], Jill Wegrzyn[10] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)University of Tennessee, Knowxville, TN, (3)University of Tennessee, Knoxville, TN, (4)University of Tennessee Knoxville, Lexington, KY, (5)Electrical & Computer Engineering, Clemson University, Clemson, SC, (6)Clemson University, Clemson, SC, (7)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (8)Genetics & Biochemistry, Clemson University, Clemson, SC, (9)University of Tennessee, Knoxville, Knoxville, TN, (10)University of Connecticut, Storrs, CT

The Tripal Gateway Project was initiated to address challenges related to "big data" for the online community genomics database. The project uses Tripal, a popular, open source toolkit to assist with the construction of online community databases. Tripal is particularly suited for the online publication of biological data using community-developed standards and provides data pages, search tools, and a graphical interface for site-specific customizations. Additionally, Tripal provides an Application Programming Interface (API) to support development of new extensions for individual community databases. Those extensions can then be shared with others. Thus, a network of multi-site developers continues to grow the functionality provided for the Tripal Community. Tripal is primarily used for the construction of online genomics, genetics and breeding websites for non-model species. These databases often fill niche's specific to their respective research community by providing data with tools, custom searches, and visualizations not provided by larger online repositories. However, the high availability of large-scale, high-throughput data creates challenges for research communities. These include challenges such as storage, transfer, and analysis of large data. The Tripal Gateway Project addresses these challenges by offering three new capabilities: 1) mechanisms for exchange of data between related Tripal sites, 2) best practices and protocols for optimizing data transfers across local and regional networks, and 3) infrastructure for complicated analysis through integration with the Galaxy Project (an interface for execution of complicated scientific workflows). The Tripal Gateway project is implemented by several Tripal-based tree databases including the Genome Database for Rosaceae (GDR), TreeGenes and HardwoodGenomics.org. The Tripal Gateway Project is funded by the US National Science Foundation (NSF) award #1443040.

W1028: Tripal Database Network and Initiatives

Tripal Extensions Facilitate Association Mapping Studies on Forest Trees with CartograTree

Nic Herndon, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Forest trees are long-lived and immobile individuals that serve as ideal models to assess population structure and adaptation to the environment. Despite the availability of comprehensive data, the researchers who study them are challenged to integrate data describing genotype, phenotype, and the environment. To address this, CartograTree, an open-source analytic web-based framework, allows the query and analysis of these data. It uses a map-based setting for association mapping, ecological genomics, and landscape genomics analyses, primarily on forest trees (with other plants also supported), to address the numerous threats facing trees worldwide. It is developed as a Drupal module to enable the seamless integration with Tripal and to take advantage of some of the features provided by the latter. It uses genotypic and

phenotypic data provided by three clade organism databases - TreeGenes, Hardwood Genomics, and Genome Database for Rosaceae - that use the Tripal v3 platform, which enables the data to be shared between them using Semantic Web technologies. It complements these data with environmental data from WorldClim, Ameriflux, and other public repositories, which can be displayed on maps using the open-source GIS platforms, GeoServer and OpenLayers. Once a user selects the data to use for analysis, CartograTree places it in the Tripal data collections, and then transfers it to a high performance computing instance running the Galaxy server. There, using the Tripal Galaxy module, the data is analyzed using one of the workflows developed for CartograTree, based on the user's selection. The workflows perform either association genetics or landscape genomics analysis.

W1029: Tripal Database Network and Initiatives

Tripal Elasticsearch: Bringing Simple and Powerful Sitewide Search to Tripal Websites

Abdullah Almsaeed[1], Ming Chen[2], Bradford Condon[1], Stephen P. Ficklin[3] and Margaret Staton[1], (1)University of Tennessee, Knoxville, TN, (2)University of Tennessee, Knowxville, TN, (3)Washington State University, Pullman, WA

The Hardwood Genomics Project (HWG, www.hardwoodgenomics.org) houses genomic resources for over 25 hardwood tree species. Built with Tripal[1], a growing toolkit for constructing biological databases, HWG also provides a variety of tools that enables scientists to visualize, analyze and download genomic data including assembled genomes, transcriptomes and genetic mapping information. As the size of data grows in genomic databases, it becomes slower and harder for the user to find relevant information. Therefore, we built Tripal Elasticsearch, a Tripal extension module that utilizes the open source search engine Elasticseach[2] to enhance the search experience and retrieve relevant data in a fast and powerful manner. Elasticsearch also addresses multiple search issues such as full text search, misspelled keywords and wildcard searching. The Tripal Elasticsearch extension module makes it easy for Tripal sites to implement Elasticsearch. It provides default settings for full site search, gene search and cross-site search. Site administrators can choose these pre-built options, or use the administrative pages to create custom searches The customizable cross-site search feature allows the user to search across multiple genomic databases from a single page. Utilizing modern web development technologies, Tripal Elasticsearch can asynchronously search millions of records across multiple databases and retrieve the most relevant results.

1- Ficklin, Stephen P., et al. "Tripal: a construction toolkit for online genome databases." Database 2011 (2011): bar044.
2- Elasticsearch. https://www.elastic.co/. Accessed Oct 26, 2017

W1030: Tripal Database Network and Initiatives

National Center for Genome Analysis Support (NCGAS) use and development of Tripal Genome Browsers on XSEDE's Jetstream

Sheri A. Sanders[1], Carrie Ganote[2], Bhavya Papudeshi[2] and Tom Doak[2], (1)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (2)Indiana University, Bloomington, IN

The National Center for Genome Analysis Support (NCGAS) helps the biological community analyze, understand, and make use of the vast amount of genomic information now available. To this end, NCGAS develops and supports genome browsers for several genomics projects, using Tripal as its front end for the last year. Since the adoption of Tripal, we've modified some of the tools to develop features our users find useful. We will introduce our groups' projects and give a demo of these tools in our browsers. These tools include: 1) Modifications to tripal_blast module to link blast reports to external JBrowse sites through URL manipulation—allowing visualization against potentially any browser with publicly available data with which to build a blast database. 2) A GUI based web tool to spin up new JBrowse instances, with on-the-fly GUI-based track addition/removal, which allows for more flexible community visualization. 3) Finally, we will demo the virtual machine image built on the XSEDE cloud (Jetstream) with Tripal, JBrowse, and these tools installed - allowing for free genome browser hosting with minimal command line use. It is our hope that these tools will reduce the learning curve required to make use of Tripal genome visualization tools.

W1031: Tripal Database Network and Initiatives

An Open and Community Oriented Web Portal for Subtropical Fruit Trees Genomic, Phenotype and Breeding Data Analysis

Alicia Talavera[1], Hugo Pineda[2], J. Inaki Hormaza[1] and Antonio J. Matas[1,2], (1)IHSM "La Mayora" - UMA - CSIC, Málaga, Spain, (2)Departamento de Biología Vegetal. Universidad de Málaga, Málaga, Spain

The availability of plant genomes and high throughput sequencing approaches to characterize genome-wide patterns from natural diversity populations or breeding programs constitutes a trove of useful data for breeders. The challenge now is to relate these data with plant phenotype under different growing conditions and to allow for fast and accurate reporting of results to establish solid breeding programs. Numerous initiatives have adopted the use of open source tools that share a common standard database scheme (Chado), a web content manager (Drupal) and several specifically designed modules to facilitate the communication between both (Tripal). Genomic, phenotypic and breeding information of several Rosaceae species, cacao, banana, cotton, among many others, is stored in a relational database and accessed through a web based interface, providing several cross-compatible platforms that collect, standardize, relate and analyze the data. In our current project, aimed at characterizing the genomic diversity of avocado, cherimoya and other subtropical fruit trees varieties, we deployed an infrastructure based in several components from the Generic Model Organism Database (GMOD) and Tripal, modifying an extended plant ontology to fit avocado traits commonly used in variety descriptions. Here we describe work done by collecting phenotype and genotype data from germplasm collections. The project, designed as a community based approach, could easily incorporate a wide range of data types, genome or phenotype related, from other research projects. Moreover, this approach could act as a model for similar work in other woody perennial fruit species where good germplasm collections and phenotype data are available.

W1032: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics

Assembly of Large Plant Genomes from Hybrid Long and Short Read Data.

Aleksey Zimin[1], Daniela Puiu[2] and Steven L. Salzberg[2], (1)University of Maryland, College Park, MD, (2)Johns Hopkins University, School of Medicine, Baltimore, MD
The third generation (PacBio SMRT and Oxford Nanopore) genome sequencing data opened a large new realm of possibilities in de novo genome assembly due to long (10kb+) read lengths and no sequencing bias. However the inherent high error rates of about 15% present a challenge to using these data to assemble highly repetitive or heterozygous plant genomes. In this presentation we describe a hybrid technique that is capable of overcoming the assembly challenges by effectively combining the third generation long PacBio or Nanopore reads with the second generation short but accurate Illumina reads. We have successfully applied this technique to create and publish complete and accurate assemblies of several challenging plant genomes such as ancestral wheat A. tauschii, Loblolly Pine, and the hexaploid wheat T. aestivum from PacBio and Illumina data. We report on our progress on using Nanopore and Illumina data for assembly of giant sequoia.
The technique is implemented in publicly available MaSuRCA assembler. You can learn about the assembler and download the code from http://masurca.blogspot.com.

W1033: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics
Hi-C and Chromosome-Scale Assembly to Detect Large Chromosomal Rearrangements in Wheat Genomes
Cécile Monat[1], Sudharsan Padmarasu[2], Axel Himmelbach[3], Kobi Baruch[4], Markus C. Kolodziej[5], Thomas Wicker[6], Sean Walkowiak[7], Jennifer Ens[7], Chu Shin Koh[8], Andrew G. Sharpe[9], Kirby Nilsen[7], Klaus F.X. Mayer[10], Simon Krattinger[5], Beat Keller[5], Martin Mascher[11], Nils Stein[11] and Curtis J Pozniak[7], (1)Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, Germany, (2)Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Gatersleben, Germany, (3)Institute of Plant Genetics and Crop Plant Research (IPK), Gaermany, Germany, (4)NRGENE LTD., Ness-Ziona, Israel, (5)University of Zurich, Zurich, Switzerland, (6)Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland, (7)University of Saskatchewan, Saskatoon, SK, Canada, (8)Global Institute for Food Security, Saskatoon, SK, Canada, (9)National Research Council Canada, Saskatoon, SK, Canada, (10)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (11)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany
High-quality sequence assemblies of multiple individuals have emerged as important tools towards a deeper understanding of the full diversity within a species. The comprehensive and robust assessment of the different type of variation based on sequence assemblies can only be achieved through the collection of genome-wide sequence data with complementary technologies that retain information about physical linkage between sequence fragments. In the context of the Wheat Ten plus Genomes Project, in which we are assembling at pseudomolecules level wheat genomes with the final goal to get on overview of the wheat pan-genome, we are studying structural variation in bread wheat (*Triticum aestivum* L.). We have developed a pipeline to use chromosome capture (Hi-C) sequencing data to construct chromosome-scale sequence assemblies. Hi-C uses three-dimensional contact probabilities of chromatin in the nucleus to reconstruct the linear order of sequence scaffolds. The first step of our pipeline is the alignment of Hi-C data to a genetically anchored sequence assembly. Subsequently, residual misassemblies are detected and corrected and chromosome-scale physical genome maps are constructed from Hi-C contact matrices. Here, we present the outcome of this computational pipeline for three wheat cultivars. Highly contiguous sequence assemblies of three wheat genomes with scaffold N50 values > 10 Mb were constructed using the DeNovoMAGIC™ technology (NRGene, Ness Ziona, Israel) from paired-end and mate-pair Illumina data as well 10X Chromium linked-reads. These assemblies were ordered into 21 pseudomolecules representing > 95 % of the genome using Hi-C. Sequence alignments and Hi-C contact matrices revealed megabase-scale chromosomal rearrangements between cultivars such as inversions and inter-chromosomal translocations.

W1034: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics
Tetraploid Wheat Germplasm Diversity Scan based on the Durum Wheat Genome Assembly
Marco Maccaferri[1], Anna-Maria Mastrangelo[2], Elisabetta Mazzucotelli[3], Raj K Pasam[4,5], Reem Joukhadar[4], Danara Ormanbekova[6,7], Silvio Salvi[8], Daniela Marone[9], Simona Corneti[10], Sara G. Milner[11], Shiaoman Chao[12], Steven S. Xu[12], Nicola Pecchioni[13], Matthew Hayden[14], Roberto Tuberosa[7], Luigi Cattivelli[15] and The International Durum Wheat Genome Sequencing Consortium[16], (1)DipSA, Department of Agricultural Science, University of Bologna, Bologna, Italy, (2)CREA-MAC, SS42, 24126 Bergamo, Foggia, Italy, (3)CREA-GB,, Fiorenzuola d' Arda, Italy, (4)Department of Economic Development, Jobs, Transport and Recourses, AgriBio Centre for AgriBioscience, Bundoora, VIC, 3083, Australia., Bundoora, Australia, (5)DEDJTR, Biosciences Research, AgriBio, Melbourne, Australia, (6)DipSA, University of Bologna, Bologna, Italy, (7)Department of Agricultural Sciences, University of Bologna, Bologna, Italy, (8)DipSA - University of Bologna, Bologna, Italy, (9)CREA-CI,, Foggia, Italy, (10)DipSA, University of Bologna,, Bologna, Italy, (11)IPK Gatersleben, Stadt Seeland, Germany, (12)USDA-ARS, Fargo, ND, (13)CREA - Research Centre for Cereal and Industrial Crops, Foggia, Italy, (14)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia, (15)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (16)CREA Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy
The genome of modern durum wheat (DW) cultivar *Svevo* has been assembled based on a combination of whole genome shotgun sequencing (270X), NRGene deNovoMagic assembler, high-resolution genetic mapping obtained from the cross between Svevo DW and Zavitan wild emmer wheat (WEW) and scaffold ordering based on chromosome conformation capture sequencing (Hi-C). The assembly consisted of 9.96 Gb of ordered sequences with 66,559 high-confidence (HC) genes. We used this resource to investigate the genetic diversity and ancestry of tetraploid wheat germplasm. iSelect 90K SNP array was used to genotype a global collection of 1,858 non-redundant accessions covering the whole range of tetraploid genetic resources from WEW, cultivated emmer (CEW), durum landraces (DWL) and modern durum cultivars (DWC). We performed a whole-genome scan for population genetic structure, selective sweeps together with the tetraploid QTLome projection. Average whole-genome genetic diversity were p_{WEW}= 0.285, p_{CEW} = 0.254, p_{DWL} = 0.201, p_{DWC} = 0.192, with an overall WEW-DWC decrease in diversity equal to 32.6%. Diversity depletions were more relevant in peri-centromeric regions (p_{WEW_C} = 0.269, p_{DWC_C} = 0.151) as compared to the highly-recombinogenic distal regions (p_{WEW_R} = 0.287, p_{DWC_R} = 0.250). From WEW to DWC, 68 chromosome regions were

subjected to diversity depletion, affecting up to 38% of the genome in total: 19 of these were associated to WEW-CEW transition, 41 to CEW-DWL and 8 to DWL-DWC. The gene content of these regions is being explored in relation to known QTL content and haplotype analysis. Overall, the analysis pointed out the chromosome regions subjected to strong selective sweeps during the domestication and breeding selection, on one side, and those regions that would benefit from targeted genetic diversity restoration on the other side.

W1035: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics
Mining Natural Variation in Triticeae to Improve Plant Immunity
Burkhard Steuernagel[1], Sanu Arora[1], Oadi Matny[2], Ryan Johnson[3], Jesse Poland[4], Alison R Bentley[5], Jonathan D. G. Jones[6], Jake Enk[7], Eric Olson[8], Steven S. Xu[9], Evans Lagudah[10], Brian J. Steffenson[3] and Brande B.H. Wulff[1], (1)John Innes Centre, Norwich, United Kingdom, (2)Department of Plant Pathology, University of Minnesota Twin Cities, St. Paul, MN, (3)University of Minnesota, St. Paul, MN, (4)Department of Plant Pathology, Kansas State University, Manhattan, KS, (5)NIAB, Cambridge, United Kingdom, (6)The Sainsbury Laboratory, Norwich, United Kingdom, (7)Arbor Biosciences, Ann Arbor, MI, (8)Michigan State University, East Lansing, MI, (9)USDA-ARS, Fargo, ND, (10)Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia
It is widely acknowledged that during domestication many crops went through a genetic bottleneck leading to loss of large parts of intraspecific diversity. Modern agriculture therefore seeks to recruit genetic diversity from wild relatives to improve crops. One important area of crop improvement is breeding for resilience to biotic stress. Mining resistance genes from crop wild relatives, however, is a laborious endeavour due to their poor agronomy, ploidy differences, and limited genomic resources. Gene cloning projects are usually long procedures involving the creation of populations dedicated to only a single gene. Here we report Association Genetics using Resistance gene ENrichment SEQuencing (AgRenSeq) to rapidly clone resistance genes from a wild diversity panel through association genomics of selectively captured and sequenced resistance gene analogues. We demonstrate our concept by mining a panel of 151 accessions of *Aegilops tauschii*, a wild relative of wheat, for resistance genes against the wheat stem rust causing fungus *Puccinia graminis* f. sp. *tritici*.

W1036: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics
Exploring Epigenomic Diversity in Polyploid Wheat
Laura-Jayne Gardiner, Earlham Institute, Norwich, United Kingdom
Wheat has been domesticated into a large number of agricultural environments, a key question is what drives the ability for crops to rapidly adapt. To address this question, we survey genotype and DNA methylation across the core Watkins bread wheat landrace collection that is representative of global wheat genetic diversity. We identify independent variation in methylation, genotype and transposon copy number. These three sources of variation are likely to be driving phenotypic differences across this diverse wheat collection. Methylation and transposon diversity could therefore be used alongside single nucleotide polymorphism (SNP) based markers for breeding.

W1037: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification
Gene Regulatory Networks Reveal Novel Genes Controlling Senescence in Wheat
Philippa Borrill, John Innes Centre, Norwich, United Kingdom
Monocarpic senescence in crops is essential to enable nutrient remobilisation from photosynthetic tissues to the grain. This process must be tightly regulated to prevent premature senescence adversely affecting yields, however few genes controlling senescence have been identified in wheat. We are using a combination of approaches to identify novel regulatory genes affecting the early processes controlling senescence. We have generated a high-resolution RNA-Seq time-course of ten time-points from anthesis until the first visible signs of flag leaf senescence. To understand the key genes driving transcriptional changes, we used a combination of gene regulatory network analyses to identify modules of co-expressed genes and hub genes regulating the transcriptional processes across this time-course. From these networks, we selected ten transcription factors as candidate genes for further characterisation. We have generated double knock-out mutants of these candidate genes using the sequenced tetraploid TILLING population. Preliminary results show that two out of five candidate genes tested to date have roles in monocarpic senescence. Further studies are in progress to characterise the effects of these novel senescence regulators on nutrient remobilisation. The availability of new genomic resources, such as high-quality genome sequences and TILLING knock-out mutants, has enabled the study of genes regulating senescence at an unprecedented resolution. These genes may represent new breeding targets to adapt senescence to the environment and to modulate grain nutrient content which is influenced by the rate of senescence.

W1038: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification
Walking a Waxy Path: Molecular Characterisation of Barley *Eceriferum-yy*
Chiara Campoli[1], Miriam Schreiber[2], Zenith Tandukar[3], Gary J. Muehlbauer[3] and Robbie Waugh[2], (1)University of Dundee, Dundee, United Kingdom, (2)The James Hutton Institute, Invergowrie, Dundee, United Kingdom, (3)University of Minnesota, St. Paul, MN
The cuticle is a waterproof barrier that covers aerial parts of land plants contributing to the reduction of non-stomatal water loss and conferring protection to many biotic and abiotic stresses. The cuticle is mainly composed by cutin and cuticular waxes, a complex mixture of aliphatic compounds derived from the lipid biosynthetic pathway. The composition of cuticular waxes is species- and organ-specific and can be modulated by environmental factors. Cuticular waxes form crystal structures on plant surfaces that confer a bluish-white (glaucous) or a glossy-green (non-glaucous) colour depending on its composition. In barley 1,580 *eceriferum* (*cer*) loci have been identified and described based on the glaucous appearance of the leaves, leaf-sheaths and spikes. Here I describe the characterisation and fine mapping of *Cer-yy*, a dominant, organ-specific suppressor of wax accumulation in barley.
We analysed the composition and crystal structure of cuticular waxes of leaves, leaf-sheaths and spikes of wild type and *Cer-yy* mutants; this showed that the non-glaucous appearance of the mutant spikes is due to the absence of β-diketones. An RNA-seq analysis corroborated this finding and revealed that in *Cer-yy* mutants the expression of the *Cer-cqu* gene cluster, a key element in the synthesis of β-diketones, is strongly down-regulated. We mapped *Cer-yy* to a sub-centimorgan region on the top of barley chromosome 1H, allowing the identification of

candidate genes. I will conclude by discussing the major findings of this research and outlining future developments aimed at identifying *Cer-yy* and understanding its contribution to the regulation of wax biosynthesis in barley.

W1039: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification
Isolating a Gene that Suppresses Stem Rust Resistance in Wheat
Colin W. Hiebert[1], Matthew J. Moscou[2], Timothy Hewitt[3], Inmaculada Hernández-Pinzón[2], Burkhard Steuernagel[4], Jaroslav Dolezel[5], Jan Vrana[6], Miroslava Karafiatova[5], Tom Fetch[7], Vincent Pujol[8], Peng Zhang[9], Matthew N. Rouse[10], Yue Jin[11], Robert A. McIntosh[9], Sridhar Bhavani[12], Li Huang[13], Brande B.H. Wulff[4], Evans Lagudah[3] and Wolfgang Spielmeyer[14], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)The Sainsbury Laboratory, Norwich, United Kingdom, (3)Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia, (4)John Innes Centre, Norwich, United Kingdom, (5)Institute of Experimental Botany, Olomouc, Czech Republic, (6)Experimental Botany, Olomouc, Czech Republic, (7)Agriculture and Agri-Food Canada, Brandon, MB, Canada, (8)CSIRO Agriculture and Food, Canberra, Australia, (9)Plant Breeding Institute, University of Sydney, Cobbitty, Australia, (10)USDA-ARS Cereal Disease Lab, St Paul, MN, (11)USDA-ARS, University of Minnesota, St Paul, MN, (12)CIMMYT, Nairobi, Kenya, (13)Montana State University, Bozeman, MT, (14)CSIRO Agriculture & Food, Canberra, Australia
Stem rust, caused by *Puccinia graminis* f.sp. *tritici* (*Pgt*), is an important disease of wheat that can be controlled by developing cultivars that carry effective resistance genes. However, as the pathogen evolves virulence, new resistance must be identified and deployed. Previously, a gene on chromosome arm 7DL was described that suppresses resistance to some races of *Pgt* in the cultivar Canthatch. When the suppressor is knocked-out by mutagenesis resistance is expressed that is normally silenced. Our goals were to map and isolate the suppressor gene, *SuSr-D1*. Two EMS-induced mutants, NS1 and NS2, were each crossed to Thatcher, the recurrent parent of Canthatch, and two doubled haploid (DH) mapping populations were generated. The DH populations were phenotyped with *Pgt* race QTHJC at the seedling stage. Chromosome 7D was isolated from wild-type and mutant stocks by flow cytometry and then sequenced using Illumina technology. Sequences were assembled and SNVs were called between wild-type and mutant contigs. SNVs and *SuSr-D1* were mapped in DH populations to define a physical region for detailed bioinformatic analysis. A single gene was found to have nonsense mutations in NS1 and NS2 that also co-segregated with the *SuSr-D1* phenotype. Sequencing an additional five independent mutants showed the same gene carried a nonsense mutation in each instance. *SuSr1-D1* encodes a subunit of the Mediator complex, which plays a key role in regulating transcription of protein-coding genes.

W1040: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification
Sr21* and *Sr13* NLR Genes Confer High-Temperature Resistance to Wheat Stem Rust Race *Ug99
Jorge Dubcovsky[1], Shisheng Chen[1], Wenjun Zhang[1] and Matthew N. Rouse[2], (1)University of California, Davis, CA, (2)USDA-ARS Cereal Disease Lab, St Paul, MN
The *Puccinia graminis* f. sp. *tritici* (*Pgt*) Ug99 race group poses a threat to global wheat production. We report here the cloning of Ug-99 resistance genes *Sr13* from durum wheat and *Sr21* from *Triticum monococcum*. Using map-based cloning, we identify CC-NBS-LRR genes completely linked to the resistant phenotypes in two large segregating populations. We confirmed the correct identification of these genes by multiple susceptible mutants and by independent transgenic resistant events of the susceptible variety Fielder. Hexaploid transgenic plants carrying more than one copy of *Sr21* were more resistant than the hexaploid lines with the *Sr21* introgression. Quantification of the sporulation areas and of pathogen growth five days post inoculation revealed significant interactions between temperature and stripe rust inoculation. Both genes were more effective at higher than at lower temperatures. *Sr13* and *Sr21* transcript levels were either not affected or slightly upregulated at higher temperatures, respectively. By contrast, six pathogenesis-related (*PR*) genes were highly upregulated at high temperatures when both the resistance genes and the pathogen were present. We hypothesize that this coordinated up-regulation of *PR* genes may contribute to the high temperature partial resistance mechanism observed in *Sr13* and *Sr21*. Based on the sequences of resistant and susceptible alleles identified in this study, we developed diagnostic markers for *Sr13* and *Sr21* to accelerate their deployment and pyramiding in wheat breeding programs. The cloning of these two genes also expands the number of *Pgt*-resistance genes that can be incorporated into multi-gene transgenic cassettes to control this devastating disease.

W1041: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification
The Cloned *Yr15* Gene (WTK1) Encodes Two Kinase-like Protein Domains, Both Required for Conferring Broad-Spectrum Resistance to Stripe Rust
Tzion Fahima[1], Elitsur Yaniv[2], Lin Huang[1], Valentina Klymiuk[1], Dina Raats[1], Lihua Feng[1], Shisheng Chen[3], Zeev Frenkel[1], Andrii Fatiukha[1], Tamar Krugman[1], Gabriel Lidzbarsky[1], Marko Jääskeläinen[4], Wei Chang[4], Christian Schudoma[5], Lars Paulin[4], Pia Laine[4], Harbans Bariana[6], Hanan Sela[7], Chris Khadgi Sørensen[8], Mogens S. Hovmøller[8], Assaf Distelfeld[7], Boulos Chalhoub[9], Jorge Dubcovsky[3], Abraham B. Korol[1] and Alan H. Schulman[2], (1)Institute of Evolution and the Department of Evolutionary and Environmental Biology, Faculty of Science and Science Education, University of Haifa, Haifa, Israel, (2)LUKE & University of Helsinki, Helsinki, Finland, (3)University of California, Davis, CA, (4)Institute of Biotechnology, University of Helsinki, Helsinki, Finland, (5)Earlham Institute, Norwich, United Kingdom, (6)University of Sydney, Cobbitty, Australia, (7)School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel, (8)Department of Agroecology, Aarhus University, Flakkebjerg, Slagelse, Denmark, (9)UEVE-INRA, Evry, France
Stripe rust, caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a devastating fungal disease that threatens global wheat production. The wild emmer wheat gene *Yr15*, located on chromosome 1BS, confers resistance to a broad spectrum of *Pst* races. Comparative genomics, chromosome walking, BAC libraries (wild emmer and bread wheat), whole genome assemblies, EMS mutagenesis and transgenic approaches enabled us to clone *Yr15* and validate its function. The Yr15 protein has a novel structure for R-genes in wheat with two kinase-like domains in tandem, designated here Wheat Tandem Kinase 1 (WTK1). We have shown that both kinase domains are essential for conferring *Pst* resistance. Macro- and microscopic observations of development and accumulation of fungal biomass suggest that the hypersensitive response

plays a central role in the resistance mechanism, limiting the development of fungal feeding structures. Non-functional alleles of *Yr15* in *T. dicoccoides*, *T. durum* and *T. aestivum* differ from the functional allele of DIC G25 by indels, creating truncated proteins. Therefore, we designed diagnostic markers that differentiate between functional and non-functional *Yr15* alleles. Our results suggest that *Yr15* has the potential to improve stripe rust resistance in a wide range of tetraploid and hexaploid wheat germplasm. The absence of the functional *Yr15* in tested durum and common wheat varieties highlights the value of DIC germplasm as a reservoir of resistance genes for wheat.

W1042: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification
Bi-Phasic Regulation of Immunity during Infection of Barley with the Powdery Mildew Pathogen
Roger Wise[1], Priyanka Surana[2], Ruo Xu[2,3], Greg Fuerst[1], Antony V. E. Chapman[2] and Dan Nettleton[2], (1)Corn Insects and Crop Genetics Research, USDA-Agricultural Research Service, Iowa State University, Ames, IA, (2)Iowa State University, Ames, IA, (3)Google, Inc., Mountain View, CA
Powdery mildew pathogens colonize over 9,500 plant species, causing critical yield loss. The Ascomycete fungus, *Blumeria graminis* f. sp. *hordei (Bgh)*, causes powdery mildew disease in barley *(Hordeum vulgare* L.*)*. Successful infection begins with penetration of host epidermal cells, culminating in haustorial feeding structures, facilitating delivery of fungal effectors to the plant and exchange of nutrients from host to pathogen. We used expression Quantitative Trait Locus (eQTL) analysis to dissect the temporal control of immunity-associated gene expression in a doubled haploid barley population challenged with *Bgh*. Two highly significant clusters of trans eQTL were identified near the telomeric ends of chromosomes 2HL and 1HS. Within these clusters reside two diverse *R* loci: the first derived from barley landrace, *H. laevigatum* (*MlLa)*, and the second originating from *H. vulgare* cv. Algerian (*Mla1)*. Notably, *MlLa* and *Mla1* are uniquely associated with the altered expression of 961 and 3,296 barley genes during fungal penetration and haustorial development, respectively. Regulatory control of transcript levels for 299 of the 961 genes is re-prioritized from *MlLa* on 2HL to *Mla1* on 1HS as infection progresses; with 292 of the 299 alternating the allele responsible for higher expression. These findings suggest a bi-phasic control mechanism that activates a network of conserved, as well as unique genes, in response to key stages of fungal infection.
Research supported by Joint NSF-PGRP/ERA-CAPS #1339348: Host Targets of Fungal Effectors as Keys to Durable Disease Resistance

W1043: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
A Megabase-Scale Comparative Analysis of Wheat Chromosome 2D from Two Wheat Cultivars Unravels Molecular Mechanisms of Genome Evolution
Simon Krattinger, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
Recent improvements in DNA sequencing technologies and assembly algorithms have paved the way to generated high-quality genome assemblies of the large and complex genomes of wheat and its wild relatives. These genome assemblies form the basis to study the evolutionary dynamics of wheat genomes on a megabase-size scale.
Here, we provide a comparative analysis of two high-quality assemblies of the 729-megabase-sized chromosomes 2D of wheat landrace Chinese Spring (IWGSC RefSeq v1.0) and the modern Swiss spring wheat line 'CH Campala *Lr22a*'. In general, there was a high degree of sequence conservation along the chromosome. Analysis of large insertions and deletions (InDels) revealed four large InDels of a total size of 2.2 Mb. The molecular signatures at their breakpoints enabled to identify unequal crossing over and double-strand break repair as the molecular causes of these InDels. In addition, the gene content of the two chromosomes were compared. This comparison revealed that 99% of the genes were present and collinear in both the cultivars. The fraction of unique single-copy genes observed was 0.44% for Chinese Spring and 0.71% for 'CH Campala *Lr22a'*. Hence, our analysis provides evidence that the number of genotype-specific genes is considerably smaller than previously estimated.

W1044: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Environmental Association in Barley Landraces: Identifying the Genetic Basis of Low Temperature and Drought Tolerance
Peter L. Morrell, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN
Barley is cultivated across a broad latitudinal range. In the Northern Hemisphere, the range of cultivation extends from the equator to inside the Arctic Circle (0 - 66˚ N). This broad range of cultivation is especially noteworthy because the wild progenitor species, *Hordeum vulgare* ssp. *spontaneum* occupies a relatively narrow latitudinal range (~30 - 40˚ N), and typically occurs at low elevation, < 1500 m. Changes in growth habit and induction of flowering have been important to cultivation at higher latitudes and a number of genes involved in flowering time have been isolated. Cold temperature tolerance has been less extensively explored but is amenable to study through environmental association in barley landraces (primitive cultivars). We report an environmental association study involving 803 landraces genotyped with the 9K iSelect genotyping platform. A subset of these lines has been used for exome resequencing. We identify allele frequency outliers across both elevation and latitudinal gradients and make use of mixed model association analysis relative to bioclimatic variables. Using resequencing data, we test for linkage disequilibrium between the SNPs queried in genotyping and SNPs in neighboring loci. In many cases, patterns of linkage disequilibrium are consistent with the causative variant occurring in the immediate vicinity of the queried SNP. While we identify a number of SNPs that potentially contribute to low temperature tolerance, variants potentially associated with drought tolerance are more abundant in our study.

W1045: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Starp: A User-Friendly and Broadly Applicable Technique for SNP Genotyping in Wheat and Other Crops
Yunming Long[1], Lili Qi[2], Xiwen Cai[1], Justin D. Faris[2] and Steven S. Xu[2], (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND
Single nucleotide polymorphisms (SNPs) are widely distributed in the genome of every organism. Recent advances in DNA sequencing technology have accelerated the discovery of variations in DNA sequences. Multiplex chip-based technology for genome-scale SNP genotyping has made great progress in the past two decades. However, PCR-based genotyping of individual SNPs remains a challenge task.

Here, we report a novel SNP genotyping method designated semi-thermal asymmetric reverse PCR (STARP), which combines all of the advantages in accuracy, throughput, simplicity, and operational costs as well as compatibility with multiple platforms. STARP assays employ two universal priming element-adjustable primers (PEA-primers) and one group of three locus-specific primers: two asymmetrically modified allele-specific primers (AMAS-primers) and their common reverse primer. The two AMAS-primers are used to specifically amplify target alleles and generate PEA-primer binding sites. The two PEA-primers are common for all genotyping assays to stringently target AMAS-primer amplicons with similar PCR efficiencies and for flexible detection using either gel-free fluorescence signals or gel-based size separation. STARP is a broadly applicable and more user-friendly alternative to KASP. We developed numerous STARP markers associated with important agronomic genes for low cadmium accumulation and resistance to Hessian fly, Fusarium head blight, and stem rust in wheat. These STARP markers have being employed in wheat breeding. In addition, STARP technique has been successfully extended to analyze the differential expression of the homologous genes and specifically amplify target DNA fragments in highly repetitive regions. STARP will facilitate genomic research in wheat and other species with large and complex genomes.

W1046: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Genomics of Barley Tiller Development, Many Branches on a Theme

Allison M. Haaning[1], Kevin P. Smith[1], Gina Brown-Guedira[2,3], Shiaoman Chao[4], Priyanka Tayagi[5] and Gary J. Muehlbauer[6], (1)University of Minnesota, Saint Paul, MN, (2)CropScience NCSU, Raleigh, NC, (3)USDA/ARS, Raleigh,, NC, (4)USDA-ARS, Fargo, ND, (5)CropScience NCSU, Raleigh, NC, Raleigh, NC, (6)University of Minnesota, St. Paul, MN

Shoot architecture of barley is largely defined by the number and vigor of tillers, and the majority of grain harvested comes from tillers. Despite this, genetic regulation and other sources of variation that impact tillering throughout development are not well characterized. The main goals of this work were to (1) identify primary sources of variation that impact tiller number and rate of outgrowth and (2) to identify genes and regulatory networks important for early tiller development. For the first goal 768 genetically diverse lines, split equally between two- and six-row spike morphology, from the USDA National Small Grains Collection were genotyped by GBS and 50K SNP array and grown in the field in 2014 and 2015; and data for the following traits was collected: tiller number from two to seven weeks past-emergence, productive tiller number, plant height, days to heading, seeds per spike, fifty seed weight, stem diameter, and leaf width. Results of genome-wide association mapping and phenotypic analyses suggest tiller number and rate of development are primarily influenced by environment, spike row-type, and days to heading. For the second goal, tissue from Bowman and Morex seedlings was harvested by laser microdissection from shoot apical meristems (SAM) and axillary meristems (AXM), which form the main stem and tillers, respectively. RNA sequencing analysis identified 102 genes that were differentially expressed between SAM and AXM in Bowman and Morex, and annotation and GO term enrichment of these genes suggests that early tiller development is mediated by ethylene signaling similar to submergence response in rice.

W1047: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Speed Breeding with Genomic Selection to Accelerate Wheat Variety Development

Amy Watson, The University of Queensland, Brisbane, Australia

Genomic selection (GS) in wheat could accelerate yield gain principally through a reduction in breeding cycle duration. A method for rapid generation advance called ‘speed breeding’ (SB) enables up to six generations of spring wheat per year, and could be used to accelerate breeding population development and be combined with GS in various breeding schemes to enable even further gains. To improve the accuracy of selection for improved yields, many heritable traits that are genetically correlated with yield could be measured in the field and used in multi-trait models to improve genetic gain (over that of traditional single-trait models only containing yield data of the training population). To test these hypotheses, a 260 multi-parent spring wheat population, genotyped with 8,000 DArT polymorphic markers, underwent yield trials over three years. Trial plots were also phenotyped for height and normalized difference vegetation index (NDVI) using a hand-held GreenSeeker sensor. Yield prediction accuracy was accessed using five-fold cross validation and predicting into different years. Results indicate multi-trait GS prediction including field proxy traits improved selection for field-based yield over that of single-trait models. These traits could be phenotyped in the field following rapid line development under SB and used with training population yield data to advance genetic gain and wheat variety development.

W1048: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Completion of the ‘Jagger’ Wheat Genome Leads to Identification of *Aegilops ventricosa* 2NS Translocation and Its Impact in Kansas and US Wheat Breeding

Liangliang Gao[1], Kevin M. Dorn[1], Trevor W. Rife[1], Xu Wang[1], Cristiano Lemes da Silva[2], Marshall Clinesmith[2], Paula Silva[1], Allan K. Fritz[3], Nils Stein[4], Martin Mascher[4] and Jesse Poland[5], (1)Kansas State University, Manhattan, KS, (2)Kansas State University, Department of Agronomy, Manhattan, KS, (3)Department of Agronomy, Kansas State University, Manhattan, KS, (4)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany, (5)Department of Plant Pathology, Kansas State University, Manhattan, KS

Since its release 23 years ago, the winter wheat cultivar Jagger has had a huge impact on wheat production in the US and around the globe and became a parental germplasm in many of the current cultivars of the central U.S. Importantly, Jagger also possesses the *Aegilops ventricosa* 2NS translocation fragment, which is associated with disease resistance against multiple wheat pathogens including the devastating wheat blast fungus. Here we present the first *de novo* assembly and anchoring of the Jagger genome based on the NRGene DeNovoMAGIC 3.0 pipeline and chromosome conformation capture technology (Hi-C). We were able to successfully anchor ~3000 scaffolds with a cumulative length of 14.2 Gb and an N50 of 10.5 Mb to build 21 draft pseudomolecules. Overall, these draft pseudomolecules showed high collinearity with the reference Chinese Spring genome (IWGSC RefSeq v1). The Jagger genome will be a powerful tool in the analysis of the wheat pan genome and the dissection of important agronomic traits. To illustrate its utility, we delineated the *Aegilops ventricosa* 2NS translocation fragment in Jagger, consisting of a ~30 Mb fragment identified through alignment of Jagger chromosome 2A to ‘Chinese Spring’ chromosome 2A. We also developed a pipeline for identification of 2NS translocations in wheat breeding lines using genotyping by sequencing (GBS) data and reference

genomes. Our results suggest that the 2NS translocation segment is widely present in KS and other US breeding materials, and is potentially providing a yield benefit over the span of 26 years.

W1049: US National Animal Genome Research Program (NRSP8)
Towards Understanding the Function of the Porcine Genome
Christopher K. Tuggle, Iowa State University, Ames, IA
Maximizing the value of genome analysis of livestock species is dependent on knowing the function of the component parts. In pigs, progress by many groups has been made in defining the transcriptome in a number of tissues and cells. The research includes early work in expressed sequence tags for comparative and physical mapping, microarray and RNAseq analysis of many tissues, as well as early work on epigenetic and genetic control of gene expression. A major focus in my laboratory has been the study of gene expression in blood to inform studies on the immune system as well as develop biomarkers predicting future phenotypes. This talk will focus on such transcriptomic studies in several projects and look toward the future of functional genomics, including current and anticipated results from the Functional Annotation of Animal Genomes (FAANG) consortium. An overarching theme of early FAANG efforts is the collaborative approach and substantial funding required for such comprehensive research, thus another emphasis of this talk will be our community efforts to work together to develop the environment for successful FAANG research. Recently expanded global funding opportunities for functional genomics in livestock show the value in such an organized and concerted effort in animal functional genomics, and bodes well for improvement in the use of genome information in animal genetics.

W1050: US National Animal Genome Research Program (NRSP8)
Comparative and Integrative Genomics of Bovine and Human Tuberculosis: A One Health Perspective
David E. MacHugh, University College Dublin, Dublin, Ireland
Human tuberculosis, caused by *Mycobacterium tuberculosis*, is second only to HIV/AIDS as the greatest killer worldwide due to a single infectious agent. Bovine tuberculosis, caused by the closely related *Mycobacterium bovis* (99.95% sequence identity), is an economically-important disease affecting global cattle production, particularly in many developing countries where it also represents a significant zoonotic disease.
We have taken a One Health approach to tuberculosis by using network-based approaches to compare and integrate host transcriptome data from human and bovine macrophages infected with *M. tuberculosis* and *M. bovis*, respectively. These analyses have shed light on host-pathogen interaction for human and bovine tuberculosis disease and have provided information that can also be used to prioritize genome-wide association data and enhance detection of genomic variants for susceptibility/resistance to mycobacterial infections in cattle and humans. In addition, we have used high-throughput epigenomics to functionally dissect transcriptional control of the bovine alveolar macrophage response to infection with *M. bovis*. Results from this work reveal mechanisms underlying mammalian macrophage M1/M2 polarisation in response to mycobacterial infection and provide a novel perspective on host-pathogen interaction. Finally, we have demonstrated that peripheral blood transcriptomics can provide a route to development of novel biomarkers for mycobacterial infections in cattle with significant implications for diagnosis of human tuberculosis.

W1051: US National Animal Genome Research Program (NRSP8)
***De novo* Assembly of Plant and Animal Genomes with Chromosome-Length Scaffolds using Hi-C**
Erez Lieberman Aiden, Baylor College of Medicine, Houston, TX

W1052: US National Animal Genome Research Program (NRSP8)
Training the Next Generation of Animal Breeders in the Genomics Era
Jack C.M. Dekkers, Iowa State University, Department of Animal Science, Ames, IA
With the increasing application of genomics in industry breeding programs, the demand for geneticists that are trained in genomics, quantitative genetics, and statistics is increasing. The purpose of this presentation is to discuss the challenges, opportunities and strategies to meet learners, industry and community goals.

W1053: US National Animal Genome Research Program (NRSP8)
Emerging Careers in Data Science and Genomics
Katherine CH Amrine, Insight Data Science, Palo Alto, CA
Data Science is the hottest field of the 21st century. Genomics has created new opportunities in industry data science from health (e.g. determining disease markers in the human genome) to entertainment (tracking ancestry). Emerging application of genomics in non-human systems are less studied but very exciting. In this talk, I will focus on data science in systems like house pets and agriculture, and the skills necessary to transition into data science roles working with non-human genomic-related data. I'll provide an overview of emerging companies in the genomics space and discuss techniques for academics looking to transition into data science, based on my experience working with hundreds of Fellows in the Insight Data Science Fellows program (insightdatascience.com).

W1054: US National Animal Genome Research Program (NRSP8)
Genome to Phenome: A USDA Blueprint for Animal Production
Caird E. Rexroad III, USDA, Beltsville, MD and **James M. Reecy**, Department of Animal Science, Iowa State University, Ames, IA
In 2008 the animal genomics community developed the “Blueprint for USDA Efforts in Agricultural Animal Genomics 2008 – 2017” under the leadership of Dr. Ronnie Green (USDA-ARS) and Dr. Muquarrab Qureshi (USDA NIFA). Over the last decade the vision outlined in this document served to guide intra- and extramural research programs at USDA and in the broader international community. In 2017 ARS and NIFA initiated efforts to develop a second generation document that reflects on progress made and outlines meaningful and tangible goals for

the next decade. A workshop was convened in November to initiate the process of developing this document, this presentation will highlight activities to date and announce additional activities to obtain feedback from the broader international animal genomics community.

W1055: Weedy and Invasive Plant Genomics
Gene Expression Hotspots Contribute to Herbicide Resistance in *Amaranthus tuberculatus*.
Patrick Tranel, University of Illinois, Urbana, IL
In the last decade, waterhemp (*Amaranthus tuberculatus*) has evolved resistance to 2,4-D and HPPD inhibitors in multiple states across the midwestern US. Two populations resistant to both chemistries, one each from Nebraska (NEB) and Illinois (CHR), were studied using RNA-seq to identify candidate resistance genes. For each population, cDNA libraries were generated and sequenced from eight herbicide-resistant (HR) and herbicide-sensitive (HS) plants. Using both a waterhemp transcriptome assembly and a high-quality grain amaranth (*A. hypochondriacus*) genome as references, differential gene expression analysis was conducted to identify genes that were significantly over- or under-expressed in HR compared to HS. When these differentially expressed genes (DEGs) were mapped back to the grain amaranth genome, physical clustering of the DEGs was apparent at gene expression "hotspots" along several of the 16 grain amaranth scaffolds. SNP calling was performed across all 32 samples to look for condition-specific (CS) variants and all statistically significant CS variants were also mapped to the grain amaranth genome. In almost every one of the expression hotspots, allele-specific expression was also observed, allowing for the development of allele-specific assays for resistance diagnosis of field samples. Further work is underway to identify any potential cis-acting regulators or epigenetic variation leading to this localized difference in expression between R and S plants. These allele-specific expression hotspots are a potentially useful tool in future RNA-seq studies to narrow down the regions of true regulatory control leading to resistance, and they may also provide insights into the evolution of herbicide resistance in weeds

W1056: Weedy and Invasive Plant Genomics
Genome Alterations in Response to Intense Glyphosate Selection in Weedy and Invasive Species.
Mithila Jugulam, Educational, Manhattan, KS
Amplification of 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) gene, the molecular target of glyphosate, has been found to confer resistance to this herbicide in several weedy and invasive species. Copy number variation is often related to alteration to number and structure of chromosomes in several eukaryotes. Nonetheless, mechanisms underlying herbicide-induced chromosomal alteration in plants are still unclear. Our research on physical mapping of the glyphosate-resistant (GR) weed species, e.g. *Kochia scoparia*, *Amaranthus tuberculatus*, *Amaranthus palmeri* and *Lolium multiflorum* have identified variation in the distribution *EPSPS* gene copies among these species. In GR *K. scoparia*, the *EPSPS* gene was localized on the distal end of one pair of homologous chromosomes, arranged in the tandem configuration of ~40 to 70 kb apart with one copy in an inverted orientation. Whereas in GR *A. tuberculatus*, a cluster of *EPSPS* genes was detected in the pericentromeric region on one pair of homologous chromosomes as well as on an additional chromosome, besides the native chromosome pair. In contrast, the development of high-resolution physical maps in GR *A. palmeri* and *L. multiflorum* displayed a unique and distinguishable distribution of *EPSPS* genes in the genome. These results propose that the mechanism of amplification of *EPSPS* gene can be species specific, potentially different genetic elements may control the amplification. Overall, our findings demonstrate that changes in chromosome structure and number as a consequence of amplification of *EPSPS* genes may be a rapid adaptation of a plant genome to herbicide stress and may expedite evolution of resistance in new environments.

W1057: Weedy and Invasive Plant Genomics
The Mechanism of Dicamba Resistance and its Physiological Consequences in *Kochia scoparia*
Eric Lloyd Patterson, Colorado State University, Fort Collins, CO

W1058: Weedy and Invasive Plant Genomics
WeedGenomics: An Online Repository of Weed Species Genomic Information.
Scott McElroy, Auburn University, Auburn, AL
There is currently no online repository for genomic information dedicated to the field of weed science. Sequencing information is customarily uploaded to the National Center of Biotechnology Information (NCBI) as a means of providing public access to sequence data referenced in peer-reviewed manuscripts. While a new online repository would not be meant to supplant NCBI, it could serve the greater weed science community more specifically. Such online repositories serve to unite scientific communities around a specific biological topic which in turn aids research advance of that area or species. Closely related to the field of weed genomics is the International Survey of Herbicide Resistant Weeds (http://www.weedscience.com) which has developed into a world-wide source for herbicide resistance information and has help to build a community of researchers focused on herbicide resistant weeds. To this end, we have developed an online repository (WeedGenomics) currently available at http://www.weedgenomics.com. WeedGenomics currently has 22 available annoated weed transcriptomes available for download or searchable by keyword or DNA search string. A database of >1600 herbicide related genes from common weed genera are available on WeedGenomics. Sequences are can also be searched for target-site resistance mutations using an online search tool or a separate command line tool can be downloaded from GitHub. WeedGenomics was deployed on on 25 June 2017. This seminar will discuss the future growth and development of WeedGenomics with the goal of developing the site into a hub for the study of weed genetics, genomics, and transcriptomics.

W1059: Weedy and Invasive Plant Genomics
TBD
Amy Lawton-Rauh, Clemson University, Clemson, SC

W1060: Weedy and Invasive Plant Genomics

Comparative Population Genomics of Herbicide Resistance: Mating System, Ploidy, and Mechanistic Patterns of Adaptation.

Julia M. Kreiner, John Stinchcombe and Stephen I. Wright, University of Toronto, Toronto, ON, Canada

The evolution of herbicide resistance in weed populations is a highly-replicated example of adaptation surmounting the race against extinction, but the factors determining its rate and nature remain poorly understood. Population genetic theory predicts that both mating system and ploidy differences should influence the rate of adaptation and its probability from standing genetic variation, with important implications for the evolution of herbicide resistance. We took two approaches to address this question. First, in a meta-analysis, we used the extensive literature on the molecular basis of herbicide resistance to gather data on the extent that populations are mutation-limited. Across 118 studies and 79 species, our findings are consistent with theoretical predictions that self-fertilization reduces resistance adaptation from standing variation within populations, but increases independent adaptation across populations. Furthermore, we found evidence for a ploidy–mating system interaction that may reflect trade-offs in polyploids between increased effective population size and greater masking of beneficial mutations. Second, we conducted a genomic study of the mechanism of herbicide resistance in agricultural populations of the selfing allotetraploid, *Capsella bursa-pastoris*. We found at least three independent origins of ALS resistance in *C. bursa pastoris* in the Canadian Prairies, with differential adaptation occurring across parental subgenomes. Population genomic scans for genomic regions in Canada under divergent selection from Europe suggested significant enrichment for genes functioning in catalytic activity and transport, potentially important for NTSR. Herbicide resistance may thus be a foremost driver of adaptation, even for a selfing species with a small effective population size.