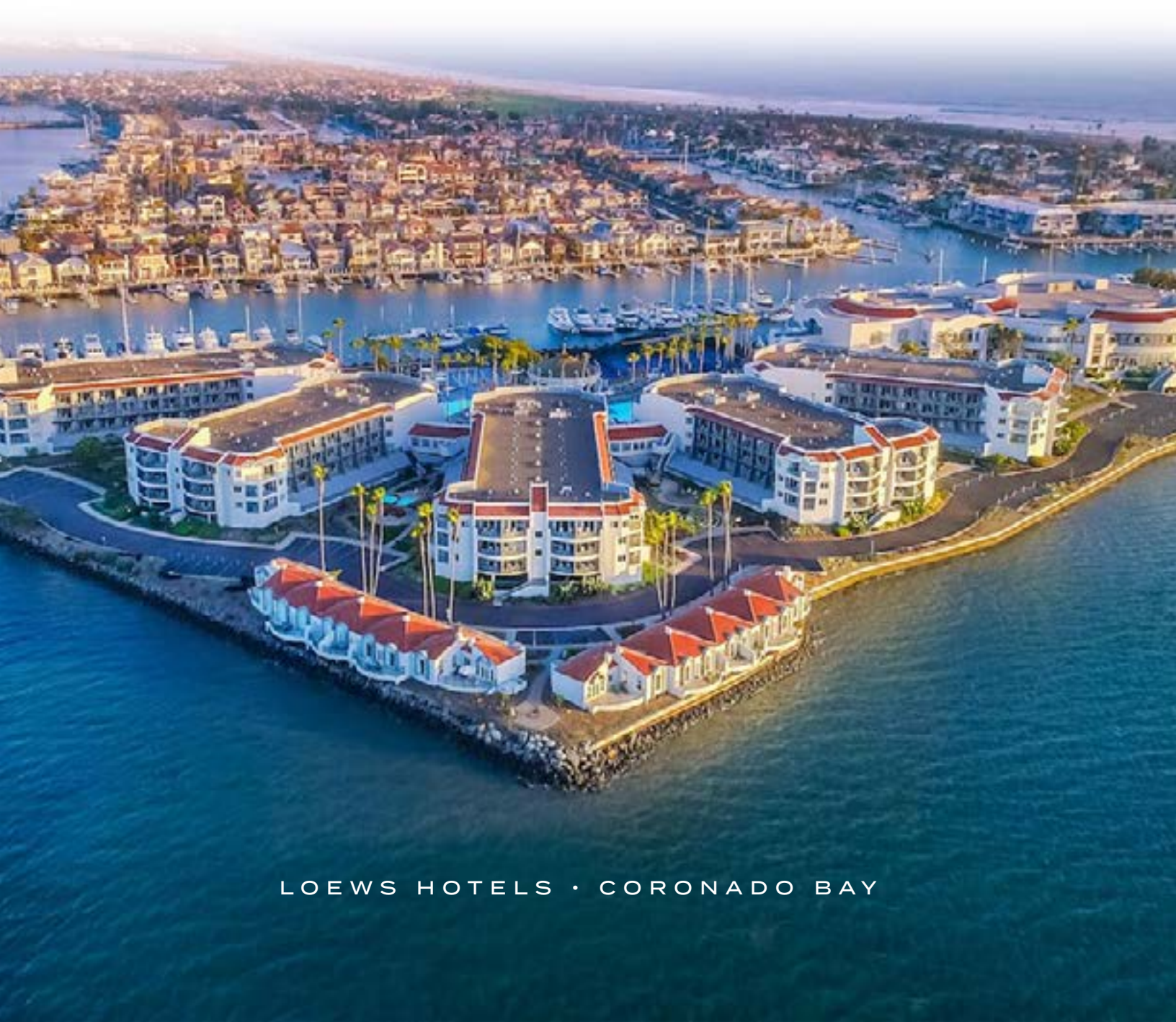
AGBT2022™
AGRICULTURE

April 4 - 6, 2022 • San Diego, CA



AGBT 2022™
AGRICULTURE

April 4 - 6, 2022 • San Diego, CA

It is our honor to welcome you to the **Advances in Genome Biology and Technology (AGBT)** inaugural Ag meeting. You are among the world's leading genome researchers, data scientists, breeders, policy influencers, funders and technology innovators from the global North and South embracing opportunities to redesign terrestrial and aquatic agriculture.

Over the next few days, we will focus on the integration of genomics and agriculture to address the escalating needs of a changing Earth.

AGBT Ag is the third global meeting organized by The Genome Partnership, a Saint Louis based non-profit that advances research, promotes education and expands commerce in genome science and technology. The Genome Partnership has been proudly organizing AGBT meetings since 1999.

Organizing Committee

We are incredibly grateful to Sarah Hearne (OC Chair) and the entire Organizing Committee for the countless hours they invested in making AGBT AG22 a content rich experience.

Over the next several days we will gather to collaborate, share the latest scientific discoveries, and discuss new technological breakthroughs thanks in large part to the work of our organizing committee.

Xiaofeng Cao
Principal Investigator-Institute of Genetics and Developmental Biology, *Chinese Academy of Sciences (CAS)*

Jack Dekkers
Distinguished professor, section leader of animal breeding and genetics, *Iowa State University*

Appolinaire Djikeng
Director, Centre for Tropical Livestock Genetics and Health, *The University of Edinburgh*

Michel Georges
Director, GIGA Institute, *Liège University*

Sarah Hearne
Principal Scientist, *International Maize and Wheat Improvement Center (CIMMYT)*

Charlie Johnson
Executive Director Genomics & Bioinformatics, *Texas A&M AgriLife*

Renee Lafitte
Deputy Director for Crops R&D, *Bill and Melinda Gates Foundation*

Nathan Lakey
President and Chief Executive Officer, *Orion Genomics*

Susan McCouch
Professor of Plant Breeding and Genetics, *Cornell University*

Len Pennacchio
DOE Joint Genome Institute, *Lawrence Berkeley National Laboratory*

Catherine Potenski
Chief Editor, *Nature Genetics*

Alison Van Eenennaam
Animal Biotechnology and Genomics, *Department of Animal Science, University of California, Davis*

Fred Van Eeuwijk
Professor in Applied Statistics, *Wageningen University, The Netherlands*

Bruce Walsh
Professor of Ecology and Evolutionary Biology, *University of Arizona*

Susan Wessler
Distinguished Professor of Genetics, *University of California, Riverside*

AGBT 2022™
AGRICULTURE

We gratefully acknowledge the generous support provided by the following companies

GOLD

ThermoFisher SCIENTIFIC

SILVER

PerkinElmer® *For the Better*

PacBio

BRONZE

CORTEVA™ agriscience

gencove

EXHIBITORS AND CONTRIBUTING SPONSORS

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

BioSpyder

BLUE CAT BIO

computomics®

dolomite bio

Element Biosciences

FLUIDIGM

illumina®

IDT
INTEGRATED DNA TECHNOLOGIES

L7 INFORMATICS

MatMaCorp®

MOLGEN

NEOGEN

NEW ENGLAND BioLabs Inc.

PARAGON GENOMICS

Quantabio

resolve biosciences

TWIST BIOSCIENCE

UgenTec

General Information

Registration

Saturday,	April 2, 2022	2:00pm - 6:00pm
Sunday,	April 3, 2022	7:00am - 6:00pm
Monday,	April 4, 2022	8:00am - 6:00pm
Tuesday,	April 5, 2022	8:00am - 6:00pm
Wednesday,	April 6, 2022	8:00am - 6:00pm

Name Badges

Please wear your name badge at all times. Security will refuse entry to anyone not associated with the AGBT meeting. Social Bands will also be available to offer a visible indicator of your personal comfort level in today's COVID environment.

Transportation

San Diego International Airport is 12 miles from the Loews Coronado Bay Resort. The transfer time to the resort is approximately 25 minutes. Airport transportation to and from the resort can be arranged by visiting **San Diego International Airport.**

Social Events

AGBT Welcome Dinner – **Monday, April 4, 6 – 8pm,** on Marina Terrace
AGBT Dinner – **Tuesday, April 5, 6:00 – 7:30pm,** Marina Terrace
Click2Win – **Tuesday, April 5, 7:30 – 10:00pm,** in the Sponsor Promenade and Exhibit Hall

Make sure to visit the Sponsor Promenade and Exhibit Hall to complete your Click2Win challenge. Attendees can earn CLICK2WIN points by completing different photo challenges to earn badges.

Grand Prize is One Complimentary 4-Day registration to AGBT 2023! The Grand Prize will be awarded during Wednesday's Plenary Session. (This prize is non-transferable.)

AGBT Farewell Dinner – **Wednesday, April 6, 6:00 – 10:00pm,** on Marina Terrace

Conference Technology

Download our official AGBT conference mobile app for easy access to the agenda, abstracts, and more. Check your email for the link to download the app or visit our registration desk for assistance.

Start Tweeting!

#AGBTAG22

Agenda

Saturday, April 2

AGBT Meeting and Pre-Conference Registration

2:00 p.m. – 6:00 p.m. • Lobby

Sunday, April 3

Pre-Conference

8:00 a.m. – 6:00 p.m. NRSP8 National Animal Genome Research Program
• Bay Pavilion

7:00 a.m. – 6:00 p.m. **AGBT Meeting Registration**
• Atrium and Commodore Foyer

Monday, April 4

Scientific Program

8:00 a.m. – 6:00 p.m. **Meeting Registration / Hospitality Desk**
• Atrium and Commodore Foyer

7:30 a.m. – 9:00 a.m. **Breakfast**
• Bay Terrace

9:00 a.m. – 9:10 a.m. **Opening Remarks**
Sarah Hearne, Principal Scientist, International Maize and Wheat Improvement Center (CIMMYT) and AGBT Ag Conference Chair

Plenary Session: **Agriculture Needs and Opportunities in the 21st Century, Session I**
(Susan McCouch, Professor of Plant Breeding and Genetics, Cornell University, Session Chair)
• Commodore Ballroom

9:10 a.m. – 9:40 a.m. Segenet Kelemu, Director General & CEO of the International Centre of Insect Physiology and Ecology (icipe)
"Innovations in agricultural research for development"

9:40 a.m. – 10:10 a.m. Carol Ibe, Researcher in crop genetics and pathology, John Innes Centre (UK) and Founding Director of JR Biotek Foundation (USA)
"Beyond subsistence agriculture: unlocking Africa's potential to feed the world"

10:10 a.m. – 10:30 a.m. John McKay (Abstract Selected)
Colorado State University
"Identification of genetic mechanisms to redesign crop production systems for sustainability"

10:30 a.m. – 11:00 a.m. Coffee Break
• Constellation Ballroom and Sponsor's Promenade

Plenary Session: **Agriculture Needs and Opportunities in the 21st Century Session II**
(Susan Wessler, Distinguished Professor of Genetics, University of California, Riverside, Session Chair
• Commodore Ballroom

11:00 a.m. – 11:30 a.m. Edward Buckler, USDA-ARS Research Geneticist and adjunct professor in Plant Breeding and Genetics, Cornell University
"Tackling maize's contributions to climate change by learning from all plant genomic diversity"

11:30 a.m. – 12:00 p.m. Panel-The Grand Challenges

Moderator:
Susan Wessler, Distinguished Professor of Genetics, University of California, Riverside

Panelists:
Sarah Hearne, Principal Scientist, International Maize, and Wheat Improvement Center (CIMMYT) and AGBT Ag Conference Chair

Jack Dekkers, Distinguished Professor, section leader of animal breeding and genetics, Iowa State University

Appolinaire Djikeng, Director, Centre for Tropical Livestock Genetics and Health, The University of Edinburgh

Renee Lafitte, Deputy Director for Crops R&D, Bill and Melinda Gates Foundation

12:00 p.m. – 1:30 p.m. **AGBT Lunch**

Plenary Session: **Plant Genomics and Breeding, Session III**
(Bruce Walsh, Professor of Ecology and Evolutionary Biology, University of Arizona, Chair)
• Commodore Ballroom

1:30 p.m. – 2:00 p.m. Nils Stein, group leader at Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), and Professor of Genomics of plant Genetic Resources at Georg-August University
"Small grains "personal genomics" – needs and feasibility for future crop improvement"

2:00 p.m. – 2:30 p.m. **Susan Wessler, Distinguished Professor of Genetics, University of California, Riverside**
"Transposable element-mediated structural variation: from McClintock to Pan-genomes"

2:30 p.m. – 2:50 p.m. Sally Mackenzie (Abstract Selected)
Pennsylvania State University
"Decoding phenotypic plasticity as an epigenetically programmed response in plants"

2:50 p.m. – 3:20 p.m.	Coffee Break • Constellation Ballroom and Sponsor's Promenade
3:20 p.m. – 4:20 p.m.	**Gold Sponsor Workshop - Thermo Fisher Scientific** Jason Santos, Ravi Ramadhar, Dr. Manish K Pandey, Krishna Reddy Gujjula, Ali Pirani *"Powering genetics and genotyping – enabling better agriculture"* • Commodore Ballroom
Plenary Session:	**Enabling Approaches and Tech, Session IV,** (Len Pennacchio, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Session Chair) • Commodore Ballroom
4:20 p.m. – 4:40 p.m.	Lyndsey Aguirre (Abstract Selected) Cold Spring Harbor Laboratory *"Uncovering quantitative epistasis in tomato stem cell control"*
4:40 p.m. – 5:00 p.m.	Katharine Jenike (Abstract Selected) Johns Hopkins University *"HetTrek: improving assembly accuracy and contiguity of highly heterozygous genomes"*
5:00 p.m. – 5:20 p.m.	Bart Nijland (Abstract Selected) Genetwister *"Graph-based pangenome analysis of a haplotype-resolved genome assembly"*
5:20 p.m. – 5:50 p.m.	Silver 1 Sponsor Workshop - PerkinElmer Charles D. Johnson, Ph.D. Texas A&M AgriLife Research *"A low cost solution for fast and accurate high-throughput genotyping"*
6:00 p.m. – 8:00 p.m.	**AGBT Welcome Dinner** • Marina Terrace
Tuesday, April 5	**(Morning)**
6:30 a.m. – 7:30 a.m.	Rise and Shine with AGBT – Optional Activity • Meet in Lobby
7:30 a.m. – 9:00 a.m.	**Breakfast** • Bay Terrace
8:00 a.m. – 6:00 p.m.	**Hospitality Desk** Atrium and Commodore Foyer
Plenary Session:	**Animal Ag Genomics, Session V** (Jack Dekkers, Distinguished Professor, section leader of animal breeding and genetics, Iowa State University, Session Chair) • Commodore Ballroom

9:00 a.m. – 9:30 a.m.	Alison Van Eenennaam, Cooperative Extension Specialist, Animal Genomics and Biotechnology, University of California, Davis *"Functional analysis and characterization of the hornless phenotype in cattle"*
9:30 a.m. – 9:50 a.m.	Jean Noel Hubert (Abstract Selected) National Research Institute for Agriculture, Food and Environment (INRAE) *"Identification of parental allele-specific methylation in pigs through targeted sequence DNA methylation enrichment methods"*
9:50 a.m. – 10:10 a.m.	Troy Rowan (Abstract Selected) University of Tennessee, Knoxville *"Detecting historic and ongoing selection in beef cattle populations"*
10:10 a.m. – 10:30 a.m.	Jeremiah Minich (Abstract Selected) Salk Institute *"Fish microbiomes 101: disentangling the rules governing marine fish mucosal microbiomes across 101 species and its implications for probiotic discovery in aquaculture"*
10:30 a.m. – 11:00 a.m.	Coffee Break • Constellation Ballroom and Sponsor's Promenade
Plenary Session:	**Informatics and Analytics Session VI:** (Nathan Lakey, Orion Genomics, Session Chair) • Commodore Ballroom
11:00 a.m. – 11:30 a.m.	Daniele Lourenco, Associate Professor in Animal Breeding, Genetics, and Genomics, University of Georgia *"Leveraging genomics to reshape animal improvement"*
11:30 a.m. – 12:00 p.m.	Adam Phillippy, Senior investigator and head of the Genome Informatics Section, National Human Genome Research Institute (NHGRI) *"Complete, telomere-to-telomere assembly of vertebrate genomes"*
12:00 p.m. – 1:30 p.m.	**AGBT Lunch**
Plenary Session:	**Data Science, Session VII** (Fred Van Eeuwijk, Professor in applied statistics, Wageningen University, The Netherlands, Session Chair) • Commodore Ballroom
1:30 p.m. – 1:50 p.m.	Poster Flash Talks
1:50 p.m. – 2:10 p.m.	Yana Safonova (Abstract Selected) University of California, San Diego *"Revealing how variations in cattle antibody repertoires correlate with responses to bovine respiratory disease vaccine"*

2:10 p.m. – 2:30 p.m.	Tianjing Zhao (Abstract Selected) Department of Animal Science, University of California Davis *"Extend mixed models to multi-layer neural networks for genomic prediction including intermediate omics data"*
2:30 p.m.– 2:50 p.m.	Brenda Murdoch (Abstract Selected) University of Idaho *"The functional annotation of the ovine reference genome"*
2:50 p.m. – 4:20 p.m.	**Poster Session A and Ice Cream Social / Exhibit Hall & Sponsor Promenade/ Avalon**

Tuesday, April 5

	(Afternoon)
Plenary Session:	**Leveraging Genome Complexity Session VIII** (Charlie Johnson, Executive Director Genomics & Bioinformatics, Texas A&M AgriLife, Session Chair) • Commodore Ballroom
4:20 p.m. – 4:40 p.m.	Marce Lorenzen, North Carolina State University (NCSU) (Abstract Selected) *"Lessons learned while developing genomic resources and tools for a hemipteran pest of maize"*
4:40 p.m. – 5:00 p.m.	Genevieve Hoopes (Abstract Selected) J.R. Simplot Company *"Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity"*
5:00 p.m. – 5:30 p.m.	**Silver 2 Sponsor Workshop - PacBio** Jeremy Schmutz *"PacBio plant genomics: an entire world of possibilities"*
5:30 p.m. – 5:45 p.m.	AGBT's Next Gen Leadership Awards
6:00 p.m. –7:30 p.m.	Dinner • Marina Terrace
7:30 p.m. – ???	**Click2Win Party** (Engage with sponsors and exhibitors to win great prizes including free registration to next year's AGBT Ag) • Constellation Ballroom and Sponsor's Promenade

Wednesday, April 6

	(Morning)
6:30 a.m. – 7:30 a.m.	Rise and Shine with AGBT – Optional Activity • Meet in Lobby
7:30 a.m. – 9:00 a.m.	**Breakfast** • Bay Terrace
8:00 a.m. – 6:00 p.m.	**Hospitality Desk** • Atrium and Commodore Foyer
Plenary Session:	**Building From Basic Science, Session, IX** (Appolinaire Djikeng, Director, Centre for Tropical Livestock Genetics and Health, The University of Edinburgh, Session Chair) • Commodore Ballroom
9:00 a.m. – 9:30 a.m.	Joanne Chory, Howard Hughes Medical Investigator (HHMI), professor and director of the plant Biology Laboratory, Co-director of the Harnessing Plants Initiative at the Salk Institute for Biological Studies *"Resilience and adaptation of plants to climate extremes"*
9:30 a.m. – 10:00 a.m.	Zach Lippman, Professor of Plant Biology, Cold Spring Harbor Laboratory (CSHL) and Howard Hughes Medical Institute (HHMI) Investigator *"Dynamic evolution of gene duplications and their interactions in shaping natural and engineered trait variation"*
10:00 a.m. – 10:20 a.m.	Michael Schatz (Abstract Selected) Johns Hopkins University *"Dissecting the dynamic evolution of paralogs in shaping trait variation across the Solanum pan-genome"*
10:20 a.m. – 10:40 a.m.	Poster Flash Talks
10:40 a.m. – 12:10 p.m.	**Poster Session B with Coffee and Snacks / Exhibit Hall & Sponsor Promenade / Avalon**
12:10 p.m. – 1:30 p.m.	**AGBT Lunch**
Plenary Session:	**Adaptation & Evolution, Session X** (Alison Van Eenennaam, Cooperative Extension Specialist, Animal Genomics and Biotechnology, University of California, Davis, Session Chair) • Commodore Ballroom
1:30 p.m. – 2:00 p.m.	Julia Bailey-Serres, Professor of Genetics and Director of the Center for Plant Cell Biology, University of California, Riverside *"At the root of cellular responses to water extremes"*
2:00 p.m. – 2:30 p.m.	Jeff Ross-Ibarra, Professor, College of Biological Sciences, University of California, Davis *"Ancient admixture from teosinte highlights the role of diversity from wild relatives to adaptation in maize."*

2:30 p.m. – 2:50 p.m.	Coffee Break • Commodore Foyer
Plenary Session:	**Environment and Late Breaking, Session XI** (Catherine Potenski, Chief Editor, Nature Genetics, Session Chair) • Commodore Ballroom
2:50 p.m. – 3:10 p.m.	David Jackson (Abstract Selected) Cold Spring Harbor Laboratory *"Validation of maize ear and root single cell data by spatial transcriptomics using Molecular Cartography"*
3:10 p.m. – 3:30 p.m.	Patrizia Ricca (Abstract Selected) Computomics *"Higher-order machine learning models act as an approximation of biological regulatory mechanisms"*
3:30 p.m. – 3:50 p.m.	Sergey Koren (Abstract Selected) National Institutes of Health *"So, you think you finished a genome?"*
3:50 p.m. – 4:10 p.m.	**Bronze 1 Sponsor Talk - Corteva** Gina Zastrow-Hayes and Heather Snowgren *"Corteva Agriscience Innovation, Bringing the Outside In"*
4:10 p.m. – 4:30 p.m.	Bronze 2 Sponsor Talk - Gencove Jesse Hoff, Ph.D. *"Lighting the path to routine sequencing for genomic evaluation"*
Plenary Session:	**Panel: Communicating Your Science, Meet the Editors,** Renee Lafitte, Deputy Director for Crops R&D, Bill and Melinda Gates Foundation, Moderator • Commodore Ballroom
4:30 p.m. – 5:15 p.m.	Panel Members: Chanfu An - Senior Editor of *Nature Communications* Nonia Pariente, PhD- Editor-in-Chief, *PLOS Biology* Christopher Surridge, Chief Editor, *Nature plants, Nature Research* Hillary Sussman-Executive Director, *Cold Spring Harbor Laboratory Press* Michelle Trenkmann, Senior Editor, *Nature*
5:15 p.m. – 5:30 p.m.	Closing Comments, Sarah Hearne, Principal Scientist, International Maize and Wheat Improvement Center (CIMMYT) and AGBT Ag Conference Chair
6:00 p.m. – 10:00 p.m.	Farewell Dinner Party • Marina Terrace

Abstract Table of Contents

page

AGBT 2022™
AGRICULTURE

Analytics Informatics, Enabling Omic Tools and Resources

Genomic and agronomic dissection of lodging in tef (Eragrostis tef)

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 300

Shiran Ben-Zeev[1], **Timo Hellwig**[1,2,3], **Muluken Demeile**[1], **Rachel Zuckerman**[1], **Sasha Vorobyova**[4], **Sariel Hübner**[4], **Yehoshua Saranga**[1]

1 Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, Israel
2 Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel
3 Institute for Plant Genetics, Heinrich Heine University, Düsseldorf, Germany
4 Galilee Research Institute (MIGAL), Tel-Hai College, Qiryat Shemona, Israel

Tef (Eragrostis tef (Zucc.) Trotter) is an allotetraploid (n=20) C4 cereal crop with a genome size of 622Mbp, originating from and primarily grown in Ethiopia. Being solely important in Ethiopia, tef has lacked the attention of the global scientific community and hence it is often categorized as an 'orphan crop'. In recent decades, tef gained worldwide interest as a gluten-free "super-food" grain and high-quality resilient forage crop. Lodging, the permanent displacement of the culm from its vertical position, is the major yield-reducing problem in tef cultivation. Lodging is responsible for up to 35% yield loss and reduction of grain quality in traditional rain-fed tef farming, this is exacerbated in modern irrigated and mechanized farming. Dissecting tef lodging to its agronomic and genomic components, targeted in our studies, may help to ameliorate of this problem. Our agronomic experiments revealed that reduced sowing rate (3kg/Ha) and deeper sowing (2cm) resulted in significantly lower lodging severity without any negative effect on the crop's productivity. Similarly, anti-gibberellin (CCC) application reduced plant height and lodging with no yield penalty. A recently assembled tef diversity panel (TDP300) was genotyped using a ddRAD sequencing approach followed by alignment to the tef reference genome. Variant calling yielded a total of 28,837 SNPs (representing an average density of 46 SNPs/Mbp) which were used, along with phenotypic data from four environment to conduct a genome wide association study (GWAS) for lodging. Genomic loci associated with lodging varied across crop developmental stages, suggesting a dynamic and apparently polygenic nature of lodging. This newly assembled panel, now available to the International scientific community, alongside the identified loci and agronomic findings pave the way towards genetic and agronomic amelioration of lodging in tef.

Combining population genomics and quantitative genomics to improve livestock sustainability

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 301

Jared E. Decker[1,2,3], **Troy N. Rowan**[1,2,4,5], **Sara M. Nilson**[1], **Harly J. Durbin**[1,2,6], **Caleb Grohmann**[3], **A.J. Knowles**[1], **John Miraszek**[2], **Robert D. Schnabel**[1,2,3]

1 Division of Animal Sciences, University of Missouri, Columbia, MO
2 Genetics Area Program, University of Missouri, Columbia, MO
3 Institute for Data Science and Informatics, University of Missouri, Columbia, MO
4 Department of Animal Science, University of Tennessee, Knoxville, TN
5 College of Veterinary Medicine, Large Animal Clinical Science, University of Tennessee, Knoxville, TN
6 Syngenta, Maryville, TN

Population genomics and quantitative genomics are related, but often disconnected, scientific disciplines. Whereas population genetics looks at what genetic changes have occurred in the past, applied quantitative genetics focuses on creating genetic change in the future. However, combining theories, concepts and open questions from these two disciplines allows us to better understand phenotypic variation and to create tools to select more sustainable livestock populations. Inefficient livestock poorly adapted to their environment loose revenue and jeopardize sustainability of the food supply. Genomic data allows us to rigorously analyze polygenic selection, local adaptation, and genotype-by-environment interactions. We mapped polygenic selection using Generation Proxy Selection Mapping (GPSM) in both cattle (~130,000 genotyped animals) and swine (66,000 genotyped pigs) datasets. In cattle datasets, the polygenic selection is usually breed-specific, whereas in swine, the loci changing via selection are often shared between lines. This suggests that genetic architectures and selection objectives are similar across lines of pigs, but this is not the case in beef cattle. We used local adaptation selection scans, genotype-by-environment genome-wide association analyses, and ecoregion-specific genomic predictions of growth traits to predict local adaptation in beef cattle. Analyzing ~130,000 cattle from two breed associations with ~850,000 high-accuracy imputed SNPs, we used linear mixed model selection mapping methods (envGWAS) to identify genomic loci associated with adaptation. Among these data sets, we identify 114 genes responding to local adaptation selection. When genotype-by-environment interactions exist, ranking animals using an ecoregion-specific genetic evaluation will be different from national cattle evaluations. We developed ecoregion-specific genomic predictions using both a genomic additive relationship matrix and a genotype-by-environment (GxE) relationship matrix. When using this model, prediction accuracy for additive effects remained the same or slightly improved, and we were also able to predict random GxE effects. Across all local adaptation and GxE analyses, genes related to immunity and the circulatory system were enriched. Prototype region-specific genomic predictions will identify cattle best suited to an environment."

Analysis of polygenic selection in commercial purebred and crossbred pig genomes using Generation Proxy Selection Mapping

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 302

Caleb J. Grohmann[1], **Caleb M. Shull**[2], **Tamar E. Crum**[2], **Clint Schwab**[2], **Timothy J. Safranski**[1], **Jared E. Decker**[1*]

1 1University of Missouri, Columbia, MO
2 The Maschhoff's, LLC, Carlyle, IL

*Corresponding author

Background
Artificial selection on quantitative traits using selection indices in commercial livestock breeding populations causes changes in allele frequency over time, termed selection signatures, at causal loci and other surrounding genomic regions. Researchers and managers of pig breeding programs are motivated to understand the genetic basis of phenotypic diversity across genetic lines, breeds, and populations using selection signature analyses. Here, we applied Generation Proxy Selection Mapping (GPSM), a genome-wide association analysis of SNP genotype (38,294 to 46,458 SNPs) on birth date, in four pig populations (15,457, 15,772, 16,595 and 8,447 pigs per population) to identify loci responding to artificial selection over a span of five to ten years. Gene-drop simulation analyses were conducted to validate GPSM results. Selection signatures within and across each population of pigs were compared in the context of commercial pork production.

Results
Forty-nine to 854 loci were identified by GPSM as under selection (q-values less than 0.10) across 15 subsets of pigs based on population combinations. The number of significant associations increased as populations of pigs were pooled. In addition, several significant associations were identified in more than one population. These results indicate concurrent selection objectives, similar genetic architectures, and shared causal variants responding to selection across populations. Negligible error rates (less than or equal to 0.02%) of false-positive associations were identified when testing GPSM on gene-drop simulated genotypes, suggesting that GPSM is robust to detection of random genetic drift in actual pig populations.

Conclusions
This work confirms the efficacy and accuracy of the GPSM method in detecting selection signatures in commercial pig populations. Our results suggest shared selection objectives and genetic architectures across swine populations. Identified polygenic selection highlights loci important to swine production.

Genotype by environment by management analyses indicate need for context-specific genetic predictions

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 303

A.J. Knowles[1], **Sara Nilson**[1], **Robert Schnabel**[1], **Jared Decker**[1]

1 Division of Animal Sciences, University of Missouri, Columbia, Columbia, MO

Cattle that are not adapted to their environment create losses in production via unrealized growth or reproductive performance. In turn, this reduces profitability of producers, as well as the food security of consumers. While the cause of many adaptations are currently undefined, use of genetic, environmental, and resource (management) availability data will allow for G x E x M analyses, in comparison to today's national breed evaluations, encouraging more rapid genetic progress to occur. In order to create these predictions, we analyze genotypes imputed to 850K SNPs and production data of 35,222 Red Angus animals in order to partition variance between the effects of genetics, environment, and management, plus their interactions. These animals were assigned to one of nine discrete ecoregions based on historical weather data, with animals born from 1976 to present. Preliminary GBLUP analyses indicate a heritability of 0.31 for yearling weight. Moreover, preliminary univariate G x E x M analyses including month of birth, creep feeding status, ecoregion, sex, and year show that there is an environmental effect of ecoregion ($p < 2e-16$) and year ($p < 2e-16$) on yearling weight. A significant management effect was found in month of birth ($p < 2e-16$), but it should be noted that the uneven distribution of animals may be the cause. Additionally, creep feeding ($p = 0.0202$) was found to be significant, with a difference of 7.98lbs between animals that had received supplementary feed pre-weaning. Along with these effects, ecoregions "Southeast" and "Fescue Belt" had the greatest effect size, -79.25lbs and 7.96lbs, respectively, indicating that there are strong environmental effects that animals may not be adapted to cope with. These differences indicate the potential for interventions by producers to increase profitability, welfare, and sustainability by utilizing animals that are adapted for the unique environments they will be kept in.

Identification of cis-eQTL in a large-scale study of gene expression in blood of healthy young pigs

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 304

L.M. Kramer[1], **J. Teng**[2], **K.S. Lim**[1], **Y. Gao**[3,4], **H. Yin**[5], **L. Bai**[5], **G.E. Liu**[3], **Z. Zhang**[2], **L. Fang**[6], **G.S. Plastow**[7], **C.K. Tuggle**[1], **J.C.M. Dekkers**[1]

1 Iowa State University, Ames, IA
2 South China Agricultural University, Guangzhou, Guangdong, China
3 Animal Genomics and Improvement Laboratory Beltsville Agricultural Research Center, United States Department of Agriculture (USDA), Beltsville, MD
4 University of Maryland, College Park, MD
5 Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China
6 University of Edinburgh, Edinburgh, Scotland
7 University of Alberta, Edmonton, Alberta, Canada

RNA sequencing and high-density genotyping or whole genome sequencing can be used in tandem to identify genomic regions associated with the control of gene expression, also known as expression quantitative trait loci (eQTL) mapping. Whole blood is a convenient tissue for many purposes due to its ease of collection and relation to health. In addition, blood as a tissue is interesting from a research position as it interacts with many other tissues throughout the organism. To improve the usefulness of blood as a tissue, it is important to characterize the genetic control of gene expression in blood to enable utilization of this information to improve detection of QTL associated with various phenotypes. The purpose of this study was to identify cis-eQTL associated with gene expression data in whole blood. Whole blood RNA-seq data and high-density SNP array genotypes were collected from ~1,600 healthy nursery pigs as part of a study of resilience to a polymicrobial disease challenge. QuantSeq libraries were multiplexed using mRNA and sequenced in two sets: the first ~900 animals using Illumina HiSeq 3000 Sequencing Systems and the second ~700 with NovaSeq 6000 Sequencing Systems from Illumina. These data were used for genotype imputation to whole genome sequence with the reference panel and bioinformatics pipeline developed by the PigGTEx consortium to identify cis-eQTL. Cis-eQTL were on average associated with six genes each, with some eQTL being found to be significantly associated with up to 53 genes. Chromosome 12 was enriched for cis-eQTL in whole blood. The results provide insights into the control of gene expression in whole blood, allowing further study of its relationship with disease resilience. Future studies will include identification of trans-eQTL associated with gene expression in blood, as well as the deconvolution of the whole blood gene expression profile into individual cell specific gene expression profiles for further eQTL analysis. Funding from USDA-NIFA grants # 2017-67007-26144 and #2021-67015-34562, Genome Canada, and PigGen Canada.

Mega-scale mixed models for genome-wide prediction and association with thousands of traits from high-throughput phenotyping

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 306

Jiayi Qu[1], **Daniel E. Runcie**[2], **Hao Cheng**[1]

1 Department of Animal Science, University of California Davis, Davis, CA
2 Department of Plant Sciences, University of California Davis, Davis, CA

Large-scale phenotype data are becoming increasingly accessible due to advances in high-throughput phenotyping platforms. Genetic analyses conducted through conventional multivariate linear mixed models present computational challenges when analyzing high-dimensional, highly correlated phenotype data. Here, we extend the conventional multivariate linear mixed models as Bayesian sparse factor models to provide a statistical framework for genome-wide prediction and association with thousands of traits. Both kernel-based models (e.g., GBLUP) and marker effects models (e.g., BayesC) have been extended to accommodate genome-enabled analysis with thousands of traits (e.g., MegaGBLUP, MegaBayesC). In both real and simulated data analysis, our mega-scale mixed models show significant improvement for genetic value prediction. For genome-wide association studies, MegaBayesC shows accurate estimation of marker effects and improved power for association inference. In conclusion, by leveraging information from thousands of traits jointly, our proposed mega-scale mixed models enhance the accuracy of genome-wide prediction and the power of association inference.

Genomics of heifer pregnancy, days to conception, and days open in Red Angus heifers

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 307

Esdras Tuyishimire[1], **Troy N. Rowan**[1,2], **Robert D. Schnabel**[1], **Jared E. Decker**[1]

1 Division of Animal Sciences, University of Missouri, Columbia, MO
2 College of Veterinary Medicine, Large Animal Clinical Science, University of Tennessee, Knoxville, TN

Heifer Pregnancy (HP)-the ability for a heifer to conceive by the end of the breeding season, Days Open (DO) - days of breeding season a heifer remained open, and Days to Conception (DC) - number of days it took for a heifer to become pregnant, are fertility traits that are believed to be crucial as production traits in the beef industry. In the past, it has been challenging to make genetic improvements on fertility traits because of their low heritability. The ubiquity of genotyping and sequencing technology and the development of advanced computational methods give hope that there is a possibility of determining the genetic influence on fertility traits. This study aims to identify the DNA markers tagging genes responsible for the relationship among HP, DO, and DC and the genetic influence on each trait among 18,039 Red Angus heifers with phenotypic information (4,097 with genotypes). GCTA and GEMMA software suites performed univariate and multivariate analyses for variance component estimation and Genome-wide Association Study (GWAS) to identify the associated genomic loci. The preliminary GCTA results from the univariate model show heritabilities of 0.116 ± 0.02, 0.107 ± 0.024, 0.098 ± 0.021 for HP, DC, and DO, respectively. The estimated heritabilities from GEMMA univariate are 0.131 ± 0.021, 0.112 ± 0.025, 0.101 ± 0.021 for HP, DC, and DO, respectively. Only the bivariate model of HP and DO could converge with -0.61 genetic correlation from GCTA and -0.629 from GEMMA. GWAS yielded 60 significant SNPs associated with HP at q-value ⩽ 0.05, while there were no significant SNPs with DC and DO at that q-value. The bivariate models (HP & DO) gave 47 and 38 significant SNPs (q-value ⩽ 0.05) with GCTA and GEMMA, respectively. The preliminary results confirm the low heritability nature of fertility traits and shed light on the future procedures to follow to increase the power for fertility traits. The upcoming studies should consider a large sample size and use de-regressed breeding values at phenotypes to increase power for the SNP association analyses.

AgriSum™ Toolkit 2.0: enabling multi-species panel analysis for AgriSeq™

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 308

Tiffany Wang, **Haktan Suren, Stéphane Daly, Krishna Reddy Gujjula**

Thermo Fisher Scientific, Austin, TX

The AgriSum™ Toolkit (AST) is a data summarization and visualization plugin for the Torrent Suite™ Software (TSS). AST was primarily developed for analyzing targeted genotyping-by-sequencing (tGBS) solutions for AgriSeq™. AST provides overall panel, markers and samples summary metrics with visualizations. AST also reports the actual genotype alleles, which can be filtered and exported in multiple formats (SNP table, TOP/BOTTOM etc) compatible for downstream analyses.

In AST 2.0, we have added a new feature which enables our customers to analyze multi species panel easily. In previous version, only one single report was created for the multi-species panel aggregating all the metrics. Aggregated summary metrics for multi species panel were often confusing customers due to confounding effect (e.g. a very poor performing sub panel dragging the overall average metrics significantly down). In multi species report, key performance metrics are calculated and reported explicitly for each individual sub panel within the run. Customer can navigate between reports accessing key performance metrics and data files for each sub panel as well as the overall panel within the report with ease.

AST 2.0 enables our customers to genotype multi species samples on the same chip. Thereby increasing the multiplexing capabilities without worrying about interpretation of the sequencing data. AST is publicly available and can be downloaded via Thermo Fisher Cloud and installed on TSS at no additional cost.

For Research Use Only. Not for use in diagnostic procedures.
Keywords: AgriSeq, GBS, Ion S5, NGS, genotyping, AgriSum, analysis, bioinformatics

Bulked segregant analysis RNA-seq to identify candidate genes for fruit firmness in blueberry (Vaccinium corymbosum L.)

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 309

Lena Wilson[1], **Hamid Ashrafi**[1]

1 North Carolina State University, Raleigh, NC

Texture is an important component of blueberry (Vaccinium corymbosum) fruit quality. In a recent survey, blueberry farmers ranked firmness as a priority trait. Additionally, firmer fruit allows for the use of machine harvesters that would bruise softer fruit. Likewise, consumers tend to prefer blueberries with crisp textures.

Breeding for fruit quality traits in blueberry is a lengthy process, especially because most breeding programs allow blueberry plants to mature up to 4 years before evaluating fruit. Molecular markers allow breeders to predict fruit quality traits much earlier at the seedling stage. This research seeks to identify candidate genes involved in blueberry fruit firmness, allowing for the subsequent development of molecular markers.

Prior to this study, southern highbush blueberries 'Arlen' (soft-fruited) and 'Reveille' (firm-fruited) were crossed to create a mapping population (n=365). Based on three years of evaluation by Firmtech II instrument, four offspring were selected: two that exhibited soft fruit and two that exhibited firm fruit. Total RNA was extracted from the leaf, immature blush fruit, and mature fruit of these four offspring and the parents. The cDNA libraries were sequenced by Illumina. Soft-fruited offspring and parent were compared with firm-fruited offspring and parent using RNA-Seq. Principal component analysis (PCA) showed that firm-fruited selections did not group separately from soft-fruited selections in leaf tissue. However, in both immature and mature fruit, PCA of gene expression showed separation between firm-fruited selections and soft-fruited selections. Differentially expressed genes with Bonferroni-corrected p-values <0.05 were cross-referenced to functional annotation of published 'Draper' blueberry genome.

Detecting historic and ongoing selection in beef cattle populations

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 311

Troy Rowan[1], **Bob Schnabel**[2], **Jared Decker**[2]

1 University of Tennessee, Institute of Agriculture, Knoxville, TN
2 University of Missouri-Columbia, Columbia, MO

Understanding the biological consequences of ongoing selection is a key goal of modern livestock and plant breeding programs. The widespread adoption of genotyping for use in genetic prediction provides an opportunity to observe ongoing changes to the genome as a result of artificial selection. We used Generation Proxy Selection Mapping (GPSM) to detect ongoing allele frequency changes in two large (100K+ animals, 11 million imputed SNPs), temporally-stratified, genomic datasets of U.S. beef cattle (Red Angus and Simmental). GPSM uses genome-wide linear mixed models to identify allelic associations with a proxy for an individual's generation number, like birth year. We identified 294 unique loci actively under selection when GPSM was applied to our datasets. In most cases, GPSM is detecting allele frequency changes due to very recent selection (< 10 years). GPSM identified variants within genomic regions associated with known breed characteristics, like fertility and maternal ability in Red Angus and carcass merit and coat color in Simmental. Over 60% of GPSM loci resided in or near (<50 kb) annotated genes. An additional 36% of selected loci overlap known epigenetic marks or putatively functional genomic regions. In addition to GPSM, we used two statistics to map selective sweeps in this data, the number of segregating loci (nSL) and Raised Accuracy in Sweep Detecting (RAiSD). These methods identify hundreds of additional putative selective sweeps. Sweep regions were enriched for genes involved in embryonic and skeletal developmental processes, likely important aspects of breed formation and improvement. Sweep loci showed minimal overlap with GPSM signatures and exhibited vastly different patterns of linkage disequilibrium. This suggests that selection mapping with GPSM can complement traditional sweep mapping methods when temporal genomic data exists. In addition to better understanding the biological processes and features that underlie artificial selection, GPSM signatures might serve as important genomic annotations.

Revealing how variations in cattle antibody repertoires correlate with responses to bovine respiratory disease vaccine

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 312

Yana Safonova[1,2], **Sung Bong Shin**[3], **Luke Kramer**[4], **James Reecy**[4], **Corey T. Watson**[2], **Timothy P.L. Smith**[3], **Pavel A. Pevzner**[1]

1 Computer Science and Engineering Department, University of California, San Diego, San Diego, CA
2 Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY
3 U.S. Meat Animal Research Center, United States Department of Agriculture (USDA) Agricultural Research Service (ARS), Clay Center, NE
4 Department of Animal Science, Iowa State University, Ames, IA

An important challenge in vaccine development is to figure out why a vaccine succeeds in some individuals and fails in others. Although antibody repertoires hold a key to answering this question, there have been very few personalized immunogenomics studies so far aimed at revealing how variations in immunoglobulin genes affect a vaccine response. We conducted an immunosequencing study of 204 calves vaccinated against bovine respiratory disease (BRD) with the goal to reveal variations in immunoglobulin genes and somatic hypermutations that impact the efficacy of vaccine response. Our study represents the largest longitudinal personalized immunogenomics study reported to date across all species, including humans. To analyze the generated dataset, we developed an algorithm for identifying variations of the immunoglobulin genes (as well as frequent somatic hypermutations) that affect various features of the antibody repertoire and titers of neutralizing antibodies. In contrast to relatively short human antibodies, cattle have a large fraction of ultralong antibodies that have opened new therapeutic opportunities. Our study revealed that ultralong antibodies are a key component of the immune response against the costliest disease of beef cattle in North America. The detected variants of the cattle immunoglobulin genes, which are implicated in the success/failure of the BRD vaccine, have the potential to direct the selection of individual cattle for ongoing breeding programs.

Refinement of the genetic basis for the scurs phenotype in cattle

Analytics Informatics, Enabling Omic Tools and Resources

POSTER 313

Guosong Wang, **Clare A. Gill**

Department of Animal Science, Texas A&M University, College Station, TX

Understanding the biological consequences of ongoing selection is a key goal of modern livestock and plant breeding programs. The widespread adoption of genotyping for use in genetic prediction provides an opportunity to observe ongoing changes to the genome as a result of artificial selection. We used Generation Proxy Selection Mapping (GPSM) to detect ongoing allele frequency changes in two large (100K+ animals, 11 million imputed SNPs), temporally-stratified, genomic datasets of U.S. beef cattle (Red Angus and Simmental). GPSM uses genome-wide linear mixed models to identify allelic associations with a proxy for an individual's generation number, like birth year. We identified 294 unique loci actively under selection when GPSM was applied to our datasets. In most cases, GPSM is detecting allele frequency changes due to very recent selection (< 10 years). GPSM identified variants within genomic regions associated with known breed characteristics, like fertility and maternal ability in Red Angus and carcass merit and coat color in Simmental. Over 60% of GPSM loci resided in or near (<50 kb) annotated genes. An additional 36% of selected loci overlap known epigenetic marks or putatively functional genomic regions. In addition to GPSM, we used two statistics to map selective sweeps in this data, the number of segregating loci (nSL) and Raised Accuracy in Sweep Detecting (RAiSD). These methods identify hundreds of additional putative selective sweeps. Sweep regions were enriched for genes involved in embryonic and skeletal developmental processes, likely important aspects of breed formation and improvement. Sweep loci showed minimal overlap with GPSM signatures and exhibited vastly different patterns of linkage disequilibrium. This suggests that selection mapping with GPSM can complement traditional sweep mapping methods when temporal genomic data exists. In addition to better understanding the biological processes and features that underlie artificial selection, GPSM signatures might serve as important genomic annotations.

Breeding Bioinformatics and Related Tools

Unleashing the utility of weather data for improving trait forecasting in strawberry

Diego Jarquin[1], **Juan Acosta**[2], **Luis Osorio**[3], **Vance Whitaker**[3]

1 Agronomy Department, University of Florida, Gainesville, FL
2 Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC
3 Gulf Coast Research and Education Center, Horticultural Sciences, University of Florida, Gainesville, FL

Genomic Selection (GS) has demonstrated utility for improving the selection of genotypes with desirable characteristics in breeding programs; however, the extend of these improvements in predictive ability is limited by considering only genomic variants in the process. In addition, elaborated prediction models that allow the inclusion of the genotype-by-environment interaction have been proposed in an attempt to leverage the availability of weather data.

The integration of climatic data has increased the expectations for improving predictive ability. However, since all genotypes tested in a location receive the same environmental stimuli (temperature, precipitation, solar radiation, etc.) the simple inclusion of these covariates may result in little or no benefit, providing no genotype-specific information. The inclusion of additional information, at the genotype-in-environment level, such as information on physiological traits or features captured with high-throughput phenotyping platforms, have shown improvements in predictive ability of up to twofold.

In this study, we developed a prediction method that uses environmental data to project the amount of environmental stimuli processed by genotypes during the most influential windows of time during the growing season, allowing the inclusion of these as surrogates of secondary traits. For this, phenotypic (27 traits) and genomic data from the Strawberry Breeding Program of the University of Florida were considered. In addition, climatic data for 50 weather covariates was available.

Two prediction models were compared: (i) M1, the conventional GS model based on main effects and interactions with environments; and (ii) M2 the proposed model based on processed environmental stimuli. Four prediction scenarios were considered for model validation: CV2 (tested genotypes/observed environments); CV1 (untested genotypes/observed environments); CV0 (tested genotypes/unobserved environments); and CV00 (untested/unobserved).

Results showed that across traits, environments, and prediction scenarios, the M2 model outperformed M1 by around threefold. Under CV2, M2 returned an average within environments predictive ability of 0.879 while for M1 was 0.297; the corresponding results under CV1, CV0 and CV00 were: (0.841, 0.288), (0.961, 0.274) and (0.643, 0.185).

These results show that there is a wide scope for improving predictive ability with the inclusion of predicted values of those projections based on processed environmental stimuli.

Breeding Opportunities

Collaboratively designed genomic tools maximize both genetic gain and economic efficiency

Breeding Bioinformatics and Related Tools

POSTER **1200**

Klara L. Verbyla[1], **John Buchanan**[1], **R.G. Tait Jr.**[2]

1 Center for Aquaculture Technologies, San Diego, CA
2 Neogen Genomics, Lincoln, NE

Genomics can deliver great benefits to agricultural breeding programs, including efficient management of diversity and inbreeding, accurate parentage assignment, optimal mating designs, improved breeding value prediction, selection decisions, and breeding strategies. Using the appropriate platform for the population of interest is critical. It is often expected that optimal results require a customized tool with a higher level of initial investment and larger ongoing costs. However, it is possible to keep costs reasonable with optimal outcomes through the creation or use of a collaboratively designed universal genotyping platform.

Collaborative genotyping platform are designed using diverse populations to ensure the core market set has broad utility alongside markers that capture specific population characteristics. Many industry parties can benefit through using such platforms, creating a sample volume to keep costs reasonable and enabling results and outcomes that are easily compared and evaluated. As the platform is updated and improved, the benefit flows to all users.

A successful example of this approach in the livestock domain is widely used Illumina Bovine BeadChip arrays which were collaboratively developed with partners across the USA, Europe and Australia. The platforms support many genomic applications, across both the dairy and beef industries, where widespread use create high demand keeping the price per sample low. More recent extensions of this concept have been deployed as GeneSeek Genomic Profiler arrays which leverage the continuously developing knowledge base about the genomic structure of a species.

This approach is also proving extremely beneficial for aquaculture species, as exemplified by the creation of a collaborative genotyping platform for L. vannamei shrimp designed using samples from eleven populations. This talk will discuss the design of the array, validating the array's results, how to achieve maximal benefit from the array, and the economic impacts of creating and using such genotyping platforms for aquaculture species.

Potato genomic tools use case at the International Potato Center Breeding Program

Breeding Bioinformatics and Related Tools

POSTER **1202**

Moctar Kante[1], **Hannele Lindqvist-Kreuze**[1]

1 International Potato Center, Lima, Peru

The international Potato Center (CIP) potato breeding program is committed to ensure and reinforce a sustainable crop production through providing clones that are adapted to farmers' cultivation conditions. Efforts to optimize the breeding pipeline are implemented and expected to increase the selection accuracy and the reduce the breeding cycle. We present here several low-cost genotyping method implementations that contribute increase the rate of genetic gain by either reducing the cycle length and operational costs or increasing the selection accuracy. A DArT mid-density targeted amplicon (DArTag) marker set have been developed with a total of around 2500 markers covering potato 12 chromosomes, and preliminary results when applied in CIP genomic selection breeding program show satisfactory prediction accuracy for late blight resistance. Trait markers were developed to characterize the breeding material mainly for disease resistance traits (late blight, PVY) which are often complex to accurately phenotype. A set of 21 quality control markers allows an identity confirmation across and within breeding stages and contribute to increase the selection accuracy. The development and use of these genomic tools in potato breeding is expected to change substantially the way potato breeding is done at CIP, and to allow a fast delivery of accurate breeding outputs to partners and producers.

Harnessing the power of plant genetics to mitigate climate change

Breeding Bioinformatics and Related Tools

POSTER **1203**

Julie Law

Salk Institute for Biological Studies, La Jolla, CA

Despite ongoing efforts at reducing carbon emissions, the levels of CO2 in the earth's atmosphere are continuing to rise at an alarming rate. The climate impacts of these increases are more evident than ever, stressing the urgency for developing new technologies to address the problem on a global scale. As plants are already key players in the earth carbon cycle, Salk's Harnessing Plants Initiative aims to leverage plant biology to generate plants that sequester more carbon. To achieve this goal, we are focused engineering traits that slow the degradation of plant-based carbon in soils, including larger root systems, deeper roots, and increased levels of naturally occurring root bio-polymers, like suberin. While these traits will ultimately need to be deployed in row and cover crops to gain the acreages needed for achieving a global impact on carbon sequestration, the initial trait discovery pipelines rely on plant models, including Arabidopsis. Here we will discuss our progress in leveraging gene regulatory networks to engineer increased suberin levels in roots and our ongoing efforts to translate these findings to select row and cover crops.

Building agricultural research capacity to achieve sustainable food production in Africa

Breeding Bioinformatics and Related Tools

POSTER **1204**

Dr Carol N. Ibe[1,2]

1 JR Biotek Foundation, USA
2 John Innes Centre, Norwich, United Kingdom

Modern genomics can accelerate crop and livestock improvement in African nations, but the chronic lack of technical capacity is a major limitation. Studies have shown that despite growth, the number of scientists in Africa is still comparatively low to many other continents (Kariuki and Kay, 2017). Yet these researchers have a better understanding of the local agricultural landscape and needs, as well as how farmers can harvest the benefits of these innovations.

At JR Biotek Foundation, we seek to tackle research capacity building deficits in African nations by leveraging regional and international partnerships to advance the development of high-impact food security-related projects. We design, develop, and deliver high-quality specialized molecular biology, genomics and bioinformatics training to upskill and enable African early-career agricultural researchers to answer their own research questions. To date, we have provided high-quality technical training to 250+ researchers via our in-person 'Reach & Teach Science in Africa' plant molecular biolab Workshops, and virtually to thousands of others, in over 25 African nations. Our grassroot train-the-trainer approach has created a powerful and vibrant community of Africa-based researchers who are training and empowering others at their local institutions. This has significantly expanded our reach and contributed to a better management of limited resources.

Here, we advocate for adequate investments in scientific research capacity building in African nations. Over the last six years, we have found that most successful long-term collaborative projects have a practical and strategic plan for capacity building. Such sustainable and impactful collaborations often involve African partners in key decision-making processes. Afterall, they are best embedded in their national and institutional contexts and understand, how the politics and power dynamics may be navigated. Therefore, to achieve sustainable agriculture and food security in Africa through the application of modern genomics and other biotechnologies, we must prioritize capacity building and engage local partners in a way that does not perpetuate power asymmetries.

CRISPr Mediated Science

CRISPR-mediated depletion of abundant rRNA molecules from host and bacteria improves sensitivity in microbial and metagenomic RNA sequencing

CRISPr Mediated Science

POSTER 700

Jon Armstrong[1], **Keith Brown**[1], **Azeem Siddique**[1], **Sridhar Ranganathan**[1], **Yvain Desplait**[1], **Dhanya Ramachandran**[1], **Gaia Suckow**[1], **Faizan Khalid**[1], **Sonal Choudhary**[1]

1 Jumpcode Genomics, San Diego, CA

Abundant ribosomal RNA from host organelles and bacterial species inhibit the sensitivity of RNA-sequencing to detect and characterize low expressed transcripts when performing shotgun RNA-Seq. This is particularly true when processing samples with small quantities of total RNA available. Post library removal of abundant cDNA library molecules is achieved through the in-vitro use of the CRISPR-CAS system. In this poster we present the design, manufacture and application of the CRISPRclean PLUS™ technology applied to both bacterial transcriptomics and meta-transcriptomic sample types. Data in this presentation includes performance characteristics on mock microbial communities, clinical stool samples, over one dozen soil bacteria and a comprehensive analysis of microinche gene expression differences from S. Aureus biofilms. Both long read and short read sequencing data are included.

Sequencing the functional genome of plant species enhanced by the in-vitro use of CRISPR-Cas9 to remove repetitive elements from lentil and wheat genomes

CRISPr Mediated Science

POSTER 701

Keith Brown[1], **Jon Armstrong**[1], **Azeem Siddique**[1], **Sridhar Ranganathan**[1]

1 Jumpcode Genomics, San Diego, CA

Hybrid capture technologies have been used for decades to sequence unique protein coding regions across plant and animal species. The flaws in this approach; allelic dropout, representation bias and limited specificity are well documented. There is a need to expand beyond protein coding elements to other functional elements such as promoters, enhancers, UTRs, ATAC-seq accessible sites, CpG islands as well as introns while maintaining the cost benefits of genome wide targeted sequencing. Here we present CRISPRclean™ Functional Genome sequencing to remove repetitive elements from complex plant genomes including wheat and lentil. Over 600,000 single guide RNA / CAS 9 ribonucleoprotein (RNP) complexes are applied to target repetitive elements (Transposable Elements and simple repeats) that make up the vast majority of these genomes. By removing these abundant repetitive elements before sequencing, genotyping yields are dramatically increased while costs are significantly reduced.

CRISPr
Mediated Science

POSTER 702

Genome editing to produce monosex and sterile fish for aquaculture

John Buchanan, **Takeshi Umazume**, **Xavier Lauth**

Center for Aquaculture Technologies, San Diego, CA
jbuchanan@aquatechcenter.com

The ability to produce sterile progeny from broodstock for aquaculture has significant benefits to culture productivity and environmental sustainability. We describe the development of strategies to generate, breed and mass-produce infertile fish. Our solutions rely on precise genetic modifications to create broodstock lines that can be incorporated into breeding programs. These approaches were validated in tilapia but are transferrable to multiple species of fish. We expect that adoption of these technologies will result in broad economic and environmental benefits for aquaculture.

Our strategy for mass producing sterile fish is designed to produce monosex, sterile populations in culture. In addition to the benefit of sterile fish, this allows the benefit of sexually dimorphic performance traits in culture. We first investigated gene mutations in two evolutionarily conserved pathways, one governing sex differentiation and the other sexual competency. We created edits in genes necessary for spermiogenesis and steroid hormone synthesis causing male sterility and masculinization, respectively. Double gene edit combinations for these genes produced all-male sterile populations. Likewise, we created variants in genes whose inactivation caused females to develop atrophic ovaries arrested at a previtellogenic stage or string-like ovaries lacking oocytes. We further disrupted genes causing genetic males to sex reverse into females. Double gene edit combinations for these genes produced all-female, sterile populations.

Propagation of the double KO broodstock lines was achieved via germ cell transplantation from a juvenile edited donor into a germ cell free wild-type recipient embryo. In the resulting recipients, the induced edits had no effect as the genes targeted are not expressed in germ cells. With this approach, we generated fertile broodstock that successfully mas- produced sterile, monosex populations.

CRISPr
Mediated Science

POSTER 703

The role of upstream and downstream cis-regulatory elements and their interactions in the regulation of CLAVATA3 in tomato and Arabidopsis

Danielle Ciren[1], **Zachary Lippman**[1]

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mutations in cis-regulatory sequences are important contributors to crop domestication and improvement. Gene regulation is often controlled by combinations of cis-regulatory elements upstream and downstream of genes, within introns and UTRs, and at sites far away. These various elements likely interact genetically and/or physically to fine-tune expression temporally and spatially. CLAVATA3 (CLV3) is a dosage-dependent signaling peptide gene in plants that negatively regulates meristem size and flower organ number. Previously, CRISPR-Cas9 multiplex targeting of the promoter of tomato CLV3 (SlCLV3) created an allelic series with a continuum of variation for fruit locule number. In order to explore the contribution of downstream regulatory elements, we used the same approach targeting the proximal 3' region of SlCLV3. Unlike the promoter targeting, only a few of these alleles resulted in subtle changes in locule number. We are currently studying the consequences of combining upstream and downstream mutations by targeting specific promoter regions in the downstream alleles, and vice-versa. In related work, we are addressing the challenge of comparing the cis-regulatory elements of conserved genes between different plant families, where there is often high sequence divergence. Again taking advantage of CRISPR-Cas9, we engineered dozens of individual and combined promoter and 3' alleles in Arabidopsis CLV3 (AtCLVC3). Notably, unlike in tomato, the AtCLV3 promoter was highly buffered to large sequence perturbations, while deletion of specific regions downstream resulted in weak increases in carpel number. Promoter-downstream combinations may have synergistic effects on carpel number. We are also currently exploring potential epistatic interactions among our alleles and other genes involved in meristem size regulation. Within and between species functional dissections of cis-regulatory elements will help unravel the complex and combinatorial nature of gene regulation in plant growth and development, as well as aid in the engineering of trait variation for crop improvement.

Improved fluorescent CRISPR fusion proteins allow for enrichment of edited cells by fluorescence-activated cell sorting

CRISPr Mediated Science

POSTER 704

Steve E. Glenn, **Michael A. Collingwood, Christopher A. Vakulskas, Sarah F. Beaudoin, Nicole M. Bode, Mark A. Behlk**

1 Integrated DNA Technologies, Coralville, IA

It is often desirable to visualize CRISPR enzymes in live cells as a means to verify efficient delivery of CRISPR reagents. The most common method to accomplish this goal with protein cargo is to make an in-frame fusion to fluorescent proteins like GFP or RFP. Here we present an improved set of CRISPR reagents that have been carefully fine-tuned to allow visualization in live cells while still retain near-wildtype levels of on-target activity. We demonstrate that fluorescent CRISPR enzyme fusion proteins and fluorescently labeled tracrRNAs can be used in combination with fluorescence-activated cell sorting to achieve levels of editing greater than with unlabeled CRISPR enzymes. We expect that these reagents will be a valuable resource in research.

Differential evolution of cis-regulation in deeply conserved developmental programs

CRISPr Mediated Science

POSTER 705

Amy Lanctot[1,2], **Anat Hendelman**[2], **Sophia Zebell**[1,2], **Zachary Lippman**[1,2]

1 Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Cis-regulatory control of gene expression is fundamental for the diversity of molecular pathways that underlie multicellular life. Particularly in plants, whose growth continues throughout their lives and responds to their environment, cis-regulation allows for the fine-tuning of developmental pathways essential for fitness. One of these pathways is flower formation, a key component of reproductive growth. Given the importance of seed and fruit crops, a comprehensive understanding of this developmental program is essential for precision agriculture. One pair of genes, LFY and UFO, is necessary and sufficient for specifying floral identity. These genes' functions are highly conserved within flowering plants, but perhaps surprisingly their expression patterns are not. In the Solanaceae, UFO expression coincides with and drives flowering while LFY is constitutively expressed, but in the Brassicaceae LFY expression coincides with and drives flowering while UFO is constitutively expressed. Our question is how flower formation, mediated by evolutionarily conserved genes, is nonetheless subject to family-specific cis-regulation that causes opposite temporal expression. To investigate this question, we are using a novel sequential alignment approach, Conservatory, to identify non-coding regions of deep sequence conservation between the Brassicaceae and Solanaceae. Conserved non-coding sequences often regulate conserved developmental programs, and differences within these conserved sequences can cause qualitative and quantitative changes in phenotype between species. Using CRISPR genome editing, we are disrupting conserved promoter sequences of LFY and UFO in Arabidopsis and tomato, two representative species of our model families. We are analyzing how disrupting these sequences affect both the distinct temporal expression patterns of LFY and UFO in these species and their shared developmental function in flower formation. This comparative approach can readily be expanded to other systems and other target loci. Such work will yield insight into how conserved non-coding sequences regulate conserved developmental pathways across large evolutionary timescales and will provide a novel tool in the arsenal for quantitative trait engineering.

CRISPr
Mediated Science

POSTER 707

Functional characterization of CRISPR/Cas9 NANOS3 knockout bovine testes using single-cell RNA sequencing

Maci L. Mueller[1], **Alba V. Ledesma**[1], **Joseph R. Owen**[2], **Sadie L. Hennig**[1], **Brett R. McNabb**[3], **Pablo J. Ross**[1], **Alison L. Van Eenennaam**[1]

1 Department of Animal Science, University of California, Davis, Davis, CA
2 Department of Obstetrics, Gynecology & Reproductive Science, University of California, San Diego, San Diego, CA
3 School of Veterinary Medicine, University of California, Davis, Davis, CA

NANOS3 is expressed early in development by primordial germ cells (PGCs) and has an essential role in protecting migrating PGCs from apoptosis. Its expression is known to be critical for germline development of both sexes in several organisms, but the role has not yet been characterized for males of agriculturally important species, such as bulls. This study generated NANOS3 knockout (KO) cattle using the CRISPR/Cas9 system and employed single cell RNA-sequencing (scRNA-seq) analysis to characterize the testicular germ and somatic cell populations of NANOS3 KO, compared to wildtype (wt), bovine testis at two developmental timepoints, fetal (n = 2 KO and 2 wt) and perinatal (n = 1 KO and 1 wt). A gene knockout approach using co-injection of two selected guide RNA (gRNA)/Cas9 ribonucleoprotein complexes into bovine zygotes (6 hours post insemination) was optimized to achieve a high NANOS3 KO rate in developing embryos (89%, n = 74/83, 7 replicates). A KO was defined as 0% wt, or non-mutated, NANOS3 alleles, based on PCR and Sanger Sequence analysis of the target region. Subsequent embryo transfers of NANOS3 KO blastocysts resulted in a 31% (n = 8/26) 35-day pregnancy rate. For scRNA-seq analysis, single cells were isolated from whole and cross-sections of bovine testes obtained from fetal (90 days post conception (dpc)) and perinatal (283 dpc) cattle, respectively. Libraries were prepared using a split-pool combinatorial barcoding kit (Parse Biosciences, Seattle, WA) and sequenced on an Illumina NovaSeq 6000 instrument (150 base paired-end). Additionally, a full term NANOS3 KO bull calf (#838) was born. At sexual maturity, this bull (#838) was found to have normal libido, an anatomically normal reproductive tract, and microscope evaluation of an ejaculate obtained via electroejaculation revealed seminal plasma only with no spermatozoa present. These results provide insights into bovine germline development and suggest the potential of NANOS3 KO bulls to serve as hosts for germline complementation to enable donor-derived gamete production. Additionally, this study demonstrates the efficiency and utility of using gene editing to knockout, or disrupt, targeted genes important to cell fate decisions to fully elucidate their role and function in developmental pathways.

CRISPr
Mediated Science

POSTER 708

Optimizing a high-throughput gene editing pipeline for crop improvement

Nikolaos Tsakirpaloglou, **Sudip Biswas, Tia Dunbar, Aya Bridgeland, Endang Septiningsih, Michael J. Thomson**

Crop Genome Editing Laboratory, Department of Soil and Crop Sciences, Texas A&M University and Texas A&M AgriLife Research, College Station, TX

Gene editing using CRISPR/Cas technology promises to greatly accelerate crop improvement. However, several bottlenecks, such as difficulties with tissue culture and regeneration of elite breeding lines, need to be addressed prior to realizing the full potential of this approach. The Texas A&M AgriLife Research Crop Genome Editing Lab (CGEL) leads several research projects to address these challenges and optimize a high-throughput gene editing pipeline for crop improvement. Initial research activities in the lab were funded internally with support by the Texas A&M Presidential X-Grant Initiative and Texas A&M AgriLife Research for gene editing across major crops. Currently, efforts focus on exploring novel delivery approaches to help bypass the lengthy tissue culture and regeneration process, including using carbon nanotube-based delivery for in planta transformation in rice, sorghum, cowpea, and peanut funded by a USDA NIFA grant. Here, we discuss how the application of single-wall carbon nanotubes (SWCNTs) was used to deliver genome editing components in monocot and dicot crop species (e.g., germinating rice seeds and cowpea leaves, respectively) and the systemic detection of edits at the target regions of phytoene desaturase (PDS) gene for the respective species. Our aim is to provide breeding and research groups with a rapid gene editing pipeline to test candidate genes in their programs, with the ultimate goal of developing nutritious, high yielding, stress-tolerant crops for the future.

CRISPr
Mediated Science

POSTER 709

An increased activity LbCas12a mutant (LbCas12a Ultra) enables robust genome editing in plants using ribonucleoprotein delivery

Liyang Zhang[1], **Yingxiao Zhang**[2], **Yanhao Cheng**[2], **Steve Glenn**[1], **Michael A. Collingwood**[1], **Nicole M. Bode**[1], **Sarah Beaudoin**[1], **Yiping Qi**[2], **Christopher A. Vakulskas**[1]

1 Integrated DNA Technologies, Coralville, IA
2 Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD

CRISPR-Cas9 system has revolutionized the field of precise genome engineering and gene therapy. Alternative systems to the Streptococcus pyogenes Cas9 (SpCas9) with divergent properties have been mined out of bacteria and archaeal genomes at a rapid pace. However, while many novel systems display RNA-guided targeting activity in bacteria and certain mammalian system, few are sufficiently active for efficient gene editing in plant species particularly when using the ribonucleotprotein (RNP) delivery format. Protein engineering of CRISPR systems is frequently required to unveil their true potential. Using Cas12a from Lachnospiraceae bacterium (LbCas12a) which appears best suited for use in plant species, we present a rapid and comprehensive workflow to identify beneficial point mutations with enhanced activity. By coupling a positive-selection DNA cleavage assay in E. coli with a single saturation mutagenesis library, we successfully generated a sequence-to-activity map of LbCas12a throughout the entire gene sequence. This high-resolution activity map revealed hundreds of specific point mutations that enabled better DNA cleavage by LbCas12a in E. coli. We subsequently characterized two dozen novel point mutations, and demonstrated their improved editing efficiency in human cells when delivered as an RNP complex. The optimal variant with stacked point mutations, i.e. IDT LbCas12a-Ultra, outperformed the existing D156R variant in human cells, and elevated the editing efficiency over wild-type protein by ~X-fold in (Yiping?) when delivered as RNP. The enhanced DNA targeting meditated Ultra mutations substantially improved the editing efficiency of LbCas12a-based Adenine base editor (LbCas12a-ABE) as RNP in human cells. Overall, we envision that LbCas12a-Ultra and associated base editor proteins will have broad utility for plant genome engineering. Moreover, our high-throughput engineering workflow of LbCas12a presented here may pave the way to systematically evolve novel CRISPR systems with enhanced functionality.

CRISPr
Mediated Science

POSTER 710

Regulation of dosage-dependent gene expression and phenotypes in cis and trans

Sophia G. Zebell[1,2], **Anat Hendelman**[2], **Simon Sretnovic**[3], **Yiping Qi**[3,4], **Zachary B. Lippman**[1,2]

1 Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
3 Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD
4 Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD

Improving the predictability and tunability of gene expression is crucial to engineering quantitative phenotypes. However, the complex relationship between cis regulatory sequence variation, trans factor binding, and phenotypic output is only beginning to be understood. To address this relationship, we have established tomato inflorescence patterning as a highly quantitative assay for expression of the dose-dependent developmental regulator ENHANCER OF JOINTLESS 2 (EJ2). Through CRISPR-Cas9 nuclease and base-editing of evolutionarily conserved sequences in the EJ2 promoter that overlap with open chromatin, we have characterized sequences and nucleotides essential for controlling EJ2 expression. Alleles that perturb these sequences give rise to a robust quantitative spectrum of inflorescence branching phenotypes in a sensitized mutant background. Motif enrichment and gene expression analysis identified candidate transcription factor regulators, which bind the regulatory region of interest in preliminary assays. We propose to utilize these transcription factors in heterologous DNA-binding and transcription activation assays, as well as in vitro kinetic DNA-binding assays with our panel of cis-regulatory alleles. By conducting molecular biology experiments with alleles first generated in planta, we correlate DNA binding data with quantitative whole organism phenotyping at a large scale. Through this work, we are establishing mechanistic links between DNA-binding, transcription regulation, and quantitative phenotypic variation.

CRISPr
Mediated Science

POSTER 711

Uncovering quantitative epistasis in tomato stem cell control

Lyndsey Aguirre[1], **David McCandlish**[2], **Dolores Cummings**[3], **Sam Hutton**[3], **Zachary B. Lippman**[1,2,4]

1 Cold Spring Harbor Laboratory, School of Biological Sciences, Cold Spring Harbor, NY
2 Cold Spring Harbor Laboratory, Quantitative Biology, Cold Spring Harbor, NY
3 University of Florida, Plant Breeding, Wimauma, FL
4 Howard Hughes Medical Institute (HHMI), Cold Spring Harbor, NY

Introducing genetic variation during crop improvement often results in both beneficial and non-beneficial epistatic interactions that influence phenotypic outcomes. Epistasis can be defined as a statistical interaction between two or more genetic mutations with quantitative phenotypes that leads to a deviation from strictly additive effects. However, the study of epistasis has classically been limited to pairwise interactions of genes using null mutations. This has led to a lack of nuance in understanding the quantitative aspects of epistatic interactions. The genetic program regulating development of the shoot apical meristem (SAM) is an excellent model system in which to interrogate quantitative epistasis. Cell proliferation in the SAM is regulated by a negative feedback loop between the genes CLAVATA3 (CLV3) and WUSCHEL (WUS) that is highly sensitive to dosage. Furthermore, quantitative variation in meristem size has been generated via CRISPR/Cas9 multiplex targeting of the promoter region of CLV3 in the model crop Solanum lycopersicum (tomato). Additionally, SlCLE9, a close paralog of SlCLV3, is actively upregulated to buffer stem cell homeostasis in a slclv3 mutant background. Here we utilize a suite of phenotypically diverse promoter alleles of SlCLV3 to elucidate the regulation of the dosage dependent epistatic interactions that exist between the intersecting SlCLV3-SlWUS and SlCLV3-SlCLE9 regulatory circuits. Using quantitative genetics and expression analyses, we show the full extent of the dose-dependent nature of active compensation between SlCLV3 and SlCLE9, as well as the epistatic interactions that occur between SlCLV3 promoter alleles and a weak gain-of-function domestication allele of SlWUS. Ultimately, dissection of the interlocus responses during active compensation and other regulatory epistatic interactions has provided the opportunity to understand the quantitative essence of a robust and deeply conserved developmental program. Moreover, these approaches have led to a new understanding of the polygenic, dose-dependent nature of epistatic interactions and potentially their impact on evolution and domestication.

CRISPr
Mediated Science

POSTER 712

Functional analysis and characterization of the hornless phenotype in cattle

Alison L. Van Eenennaam[1], **Sadie L. Hennig**[1], **Josephine F. Trott**[1], **Bret R. McNabb**[2], **Joseph R. Owen**[3], **Amy E. Young**[4], **Maci L. Mueller**[1], **Jason C. Lin**[1], **James D. Murray**[1,2]

1 Department of Animal Science, University of California, Davis, CA
2 Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, CA
3 Department of Obstetrics, Gynecology and Reproductive Science, University of California, San Diego, CA
4 Center for Equine Health, School of Veterinary Medicine, University of California, Davis, CA

We have used gene editing (GnEd) to explore the hornless (polled) phenotype in cattle. Horns can be dangerous, and their physical removal is undertaken to protect animals and their handlers. Dehorning is increasingly associated with animal welfare concerns. A naturally occurring dominant genetic mutation (Pc allele) comprised of a 212 bp duplicated DNA sequence replacing a 10-bp sequence at the POLLED locus, results in the hornless phenotype. This Pc allele is intergenic, and has been hypothesized to possibly affect regulation of nearby long non-coding RNAs. We hypothesized that either the 10 bp sequence missing in the Pc allele, or alternatively the expression of a long non-coding RNA (lincRNA#1) sequence results in the polled phenotype. GnEd of bovine embryos with CRISPR-Cas9 and dual guide RNAs was used to generate targeted, homozygous deletions at both loci independently. The resulting animals were identical to controls, signifying that the absence of these two genomic regions does not result in the polled phenotype. Additionally, we undertook multiyear genomic and phenotypic analyses of six offspring of a dairy bull GnEd to be homozygous for the Pc allele. The offspring were heterozygous and polled as expected, and did not differ in their growth, health or development, nor the nutritional composition of their meat and milk, when evaluated against contemporary comparator controls. We further modelled how GnEd could be used to introgress the Pc allele in the US dairy cattle population by GnEd of elite artificial insemination bulls. Collectively, this research shows how the advent of targeted GnEd tools opens the way for hypothesis testing of putative gene function, and the rapid introgression of useful alleles into livestock breeding programs. However, the promise of this technology is far from certain. In contrast to GnEd crops, genomic alterations introduced by GnEd in food animals including targeted deletions and intraspecies allele introgressions (e.g. Pc) that could have been produced using conventional breeding, are currently considered unapproved new animal drugs by the US Food and Drug Administration. Drug approval is a lengthy and expensive process, limiting access of US researchers and small companies to employ GnEd in livestock improvement programs.

Ecology, Environment, Evolution and Economics

Agroecology as the best platform for the sustainable food systems

Kakha Nadiradze[1]

1 Association for Farmers Rights Defense (AFRD), Tbilisi, Georgia

Farmers and other value chain actors in Georgia innovate using the best practices that improve their adaptive capacity to climate change and to the increasing scarcity of natural resources based on the Agroecology philosophy. Many Farmers Members of the Association for Farmers Rights Defense, AFRD applying the agro-ecological or 'climate-smart agriculture' concepts and approaches. This means they are taking into account the ecological characteristics of the landscape, the opportunities to improve (agro-) biodiversity and the need to reduce greenhouse gases, in order to enhance the resilience of the Foos systems.

Ancient maize adaptation to novel environments

Ecology, Environment, Evolution and Economics

POSTER 502

Miguel Vallebueno-Estrada[1,2], **Bruce Benz**[3], **Michael Blake**[4], **Kelly Swarts**[1,5]

1 Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria
2 Vienna International Postdoctoral Program, Vienna BioCenter, Vienna, Austria
3 Texas Wesleyan University, Fort Worth, TX
4 The University of British Columbia, Vancouver and Kelowna, British Columbia, Canada
5 Max Perutz Labs, Department of Structural and Computational Biology, University of Vienna, Vienna, Austria

All organisms must contend with rapidly changing environments in the face of climate change in order to ensure the survival of the population. Domesticated plants, with a 10,000 year history of adapting to new environments, provide an excellent model for understanding genetic responses to changing climate as well as domestication genetics. Maize (Zea mays ssp. mays), a globally critical crop plant with extensive genomic resources, was domesticated from the tropical American grass teosinte (Zea mays spp. parviglumis). Although maize demographics during domestication and improvement were not straightforward, populations fluctuated as people moved maize into increasingly novel environments, resulting in a wide variety of adaptations to different environments in a short period of time. Maize has a complex demographic history and archaeological samples represent a unique opportunity to directly sample ancient diversity and to evaluate allelic combinations not present in extant samples in the context of changing environments.

Identification of genetic mechanisms to redesign crop production systems for sustainability

Ecology, Environment, Evolution and Economics

POSTER 503

John McKay

Colorado State University, Fort Collins, CO

Current US cropping systems were not designed to minimize greenhouse gas emissions nor to maximize Carbon sequestration. An example is corn, produced in the Us on 90 million acres annually, where the crop is used domestically for feed, fuel and food, as well as sold as an export. Like other commodity crops, corn production is considered a public good and the cost of the current crop production systems is massively subsidized by the US government. Atmospheric CO2 concentrations, minimizing greenhouse gas emissions and maximizing Carbon sequestration in cropping systems can also be considered as public goods, for current and future generations. In addition, many corporations and governments have acknowledged the urgent need for reductions in GHG emissions, as well as the role of annual cropping systems in reducing emissions and sequestering C. This move towards carbon negative cropping systems requires changes in management and genetics. The genotypes currently used in corn production are bred to be produced the high rainfall regions of the US, in fertilized fields with excess Nitrogen. Lower Nitrogen requirements and deeper root systems are two key changes that can make annual crop production carbon negative. I will discuss the biological levers that can be used to make crops carbon negative, and the crop traits that are needed to achieve this. I will also discuss ongoing research to rapidly identify genes controlling the key traits in corn.

Epigenetics Tools and Biology

Identification of parental allele-specific methylation in pigs through targeted sequence DNA methylation enrichment methods

Jean-Noël Hubert[1], **Nathalie Iannuccelli**[1], **Cédric Cabau**[2], **Rémy-Félix Serre**[3], **Céline Vandecasteele**[3], **Eva Jacomet**[1], **Sylvie Brenot**[4], **Yvon Billon**[4], **William Hebrard**[4], **Cécile Donnadieu**[3], **Julie Demars**[1]

1 GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France
2 Sigenae, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France
3 GeT-PlaGe, Genotoul, INRAE, Castanet-Tolosan, France
4 GenESI, Saint-Pierre d'Amilly, France

Genomic imprinting is a particularly attractive example of epigenetic regulation, since in the same cell one of the two parental alleles is stably repressed by epigenetic modifications, whereas the other allele is maintained in an active state. This allele-specific regulation is entirely dependent on whether the gene is inherited from the mother or from the father. Imprinted genes are typically located in clusters of 3-12 genes that are spread over 20 kb-3.7 Mb of DNA, although examples of single imprinted genes do exist. While imprinted genes, genomic imprinting mechanisms and their roles in growth, development and behaviour are well characterized in human and mice, a global overview is still missing in other organisms such as livestock.

We built an in silico pig imprintome via a homologous comparison of all the genes known to be imprinted in both human and mice. This led to a set of 165 genomic regions ranging from 458 bp to 2,337,711 bp, including 218 genes, yielding a total panel of 23,346,943 bp. Two methods of DNA methylation enrichment, through capture hybridization followed by sequencing, were tested to detect allele-specific epigenetic modifications. The two underlying technologies differ from each other in various aspects like the reaction for detecting methylation, the relative position of the capture step and the probe type. In addition, pigs from reciprocal crosses were used as primary resource to identify parental allele-specific methylation.

We firstly checked quality control criteria including mapping quality, off-target percentage, CpG conversion and CpG methylation. Both technologies showed advantages and drawbacks. The capture was heterogeneous while including highly CpG-rich sequence for one technology, whereas the second one yielded a more homogeneous coverage of target regions with a medium density of CpG. We identified differentially methylated regions depending on the parental origin, such as imprinting control regions, which are key elements of imprinted clusters regulating the parental origin-specific expression of genes. We designed a molecular tool targeting the pig imprintome suitable for the study of genomic imprinting mechanisms and their contribution to complex traits.

Epigenetics Tools and Biology

POSTER 601

Development of a quantitative CUT&RUN platform for scaled epigenomics studies in agricultural research

Andrea L. Johnstone[1], **Keith E. Maier**[1], **Ellen Weinzapfel**[1], **Irina K. Popova**[1], **Bryan J. Venters**[1], **Katherine Novitzky**[1], **Danielle N. Maryanski**[1], **Keli L. Rodriguez**[1], **Matthew R. Marunde**[1], **Carolina P. Lin**[1], **Ashton Gladney**[1], **Rachel Watson**[1], **Zachary B. Gillespie**[1], **Augustus Adeleke**[1], **Saarang Gopinath**[1], **Leslie Lewis**[1], **Hannah Willis**[1], **Lu Sun**[1], **Marcus A. Cheek**[1], **Matthew J. Meiners**[1], **James R. Bone**[1], **Martis W. Cowles**[1], **Zu-Wen Sun**[1], **Michael-Christopher Keogh**[1]

1 EpiCypher Inc., Durham, NC

Mapping the genomic localization of transcriptional signaling proteins and epigenetic modifications is critical to understand plant development and phenotypic plasticity. ChIP-seq has historically been the leading approach for mapping chromatin elements. However, it is plagued by a variety of shortcomings, including a large number of required cells, poor signal-to-noise, low sample throughput, and high cost due to deep sequencing requirements. Cleavage Under Targets and Release Using Nuclease (CUT&RUN) is the next-generation epigenomic mapping approach that delivers high signal-to-noise mapping data using a fraction of the required cells and sequencing depth compared to ChIP-seq. We describe the development of a robust, quantitative automation-compatible CUT&RUN workflow to enable scaled genomics studies for agricultural research. A key component of these workflows is the development of spike-in controls comprised of DNA-barcoded recombinant nucleosomes, which represent the physiological targets in CUT&RUN. We developed the use of these spike-in nucleosomes as controls to quantify on- vs. off-target recovery of widely used histone PTM antibodies, showing that most antibodies are not fit-for-purpose in epigenomics research. We also describe a novel approach to develop and validate antibodies for best performance in CUT&RUN. Applying these high-quality reagents and defined controls into an automated CUT&RUN workflow provides a quantitative readout of experimental success, ultimately enabling scaled and low-cost workflows with reliable results. Our quantitative CUT&RUN approach commoditizes the use of epigenomics for agricultural research, enabling insights into epigenetic regulatory mechanisms underlying plant development and environmental responses.

Epigenetics Tools and Biology

POSTER 602

Differences in activity and stability drive variable transposable element accumulation in tropical and temperate maize

Shujun Ou[1], **Yinjie Qiu**[2], **Arun S. Seetharam**[1], **Nancy Manchanda**[1], **Jonathan I. Gent**[3], **Sarah N. Anderson**[4], **Matthew B. Hufford**[1,*], **Candice N. Hirsch**[2,*]

1 Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA
2 Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN
3 Department of Plant Biology, University of Georgia, Athens, GA
4 Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA
*Corresponding authors (mhufford@iastate.edu for MBH and cnhirsch@umn.edu for CNH)

Eukaryotic genome size varies extensively within and between species. Much of the interspecific variation has been attributed to transposable elements (TEs). However, the extent and cause of genome size variation within species are less known. Here we analyzed the TE content of 26 high-quality maize genomes and revealed an excess of 42.4 Mb of TE sequences per genome in tropical maize relative to temperate maize. A small number of TE families, mainly LTR retrotransposons, drive these differences. Evidence from the methylome, transcriptome, LTR age distribution, and LTR insertional polymorphism revealed that 64.8% of the variability was contributed by LTR families that were young and more expressed in tropical maize, and 18.5% was driven by LTR families that were removed or lost in temperate maize. This study demonstrates the important role of TEs in driving genome size variation within species via amplification and purging of TEs.

Epigenetics Tools and Biology

POSTER 603

Transgenerational epigenetic inheritance and immunity priming in Marek's disease resistance

Yanghua He[1], **Robert Taylor**[2], **Chris Ashwell**[2], **Huanmin Zhang**[3], **Hans Cheng**[3], **Jiuzhou Song**[4]

1 University of Hawaii, Honolulu, HI
2 West Virginia University, Morgantown, WV
3 United States Department of Agriculture (USDA) Agricultural Research Service (ARS) Avian Disease & Oncology Laboratory (ADOL), East Lansing, MI
4 University of Maryland, College Park, MD

Viruses as obligate intracellular pathogens threaten humans, poultry health, and welfare. Disease outcomes are a complex determination shaped by polygenetic effects, epigenetics, immunity, and environmental factors. Besides genetics, epigenetics influences host innate and adaptive immune responses. However, we don't know whether epigenetic alterations be transferred to germinal cells that can cross the barriers of sexual reproduction and thus perpetuate new phenotypes? These issues are related to the "Lamarckian" question of whether we inherit environmentally induced traits. Marek's Disease (MD) is a T cell lymphomatous in many kinds of birds. We used well-defined chicken inbred lines on Marek's disease study and their ascendants composed of several recombinant congenic strains. We used novel statistical methods to evaluate the similarity of the DNA methylation spectrums and found solid evidence to illustrate transgenerational epigenetic inheritance. The results from the study demonstrated that DNA methylation patterns could be inherited over 23 generations and underly unique immune pathways. The information further advanced our knowledge on transgenerational epigenetic inheritance in disease resistance.

Epigenetics Tools and Biology

POSTER 604

Elevated transposable element copy number is associated with reduced fitness in maize

Michelle Stitzer

Cornell University, Ithaca, NY

The maize genomic sequence presents an incredible abundance and diversity of transposable elements (TEs). It is often noted that 85% of the maize genome consists of TEs, and TE alleles have been linked to major phenotypic changes in plant architecture, flowering time, and day length adaptation. Across maize individuals, the proportion of the genome derived from TEs is broadly consistent, yet there is extensive variation in the abundance of TE families and the position of individual TE copies. Only about 20% of all TE copies are found at the same position across maize individuals. Despite this great variation, it is unclear the degree to which these differences in TE content affect plant phenotypes and fitness. We project TE copy number inferred from whole genome assemblies to genotyped recombinant inbred lines, and associate TE copy number to phenotypes. After correcting for parental ancestry, we find that total TE copy number is negatively associated with fitness-related traits like yield. In organisms with smaller genomes, like yeast and Drosophila, there is strong experimental evidence for the deleterious impact of TEs on fitness, even when TEs do not directly disrupt genes. Yet, for larger genomes like maize with many more TEs, it appears these fitness costs are not as extreme, and limited to only some TE families. Altogether, while the maize genome can tolerate a large TE load, TEs continue to contribute to quantitative variation in agriculturally relevant phenotypes.

Epigenetics Tools and Biology

POSTER 605

Decoding phenotypic plasticity as an epigenetically programmed response in plants

Xiaodong Yang, Hardik Kundariya, Robersy Sanchez, Ha Eun Jeh, Isaac Dopp, Sally A, Mackenzie

Departments of Biology and Plant Science, The Pennsylvania State University

Plants undergo epigenetic changes in response to environment, some with transgenerational effects on growth and resilience. This epigenetic reprogramming of plant phenotype is amenable to detailed study within the MSH1 system as a cross-species model. We have found that disruption of the plant-specific gene MSH1 can trigger at least four distinct epigenetic states that impact plant stress response, growth vigor and fitness. The most complex of these msh1-derived states displays heritable epigenetic growth vigor following crosses of msh1 to wildtype, with low frequency reversion/interconversion back to a nongenetic 'memory' state. Phenotypic plasticity in the four msh1-derived states is accompanied by nonrandom cytosine methylation repatterning of pivotal growth-versus-stress, chromatin remodeling, and RNA spliceosome gene networks. Central network hubs were confirmed to be RdDM targets genetically and by high-resolution methylome analysis. Our studies identified dcl2/3/4-sensitive intra-genic methylation sites within identifiable sequence motifs, supporting a model of epigenomic specificity and programmed response processes in plants. These outcomes have important implications for the influence and malleability of epigenetic systems in crop resilience.

Functional Genomics and Breeding

Gene regulatory networks shape developmental plasticity of root cell types under water extremes

Functional Genomics and Breeding

POSTER **900**

Alexander T. Borowsky[1], **Mauricio Reynoso**[1], **Germain Pauluzzi**[1], **Julia Bailey-Serres**[1]

1 Center for Plant Cell Biology, Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA

Understanding how roots modulate development and metabolism under varied irrigation or rainfall is crucial for development of climate resilient crops. However, root development and environmental response involves the complex orchestration of different genetic programs in different cell types. For example, in rice, the accelerated development of aerenchyma through programmed cell death of cortical cells in response to waterlogging and the formation of a water- and airtight barrier of suberin in the exodermis. We established a toolbox of tagged rice lines to profile translating mRNAs and chromatin accessibility within specific cell populations. We used these tools to generate multi-omic profiles of rice root cell types in a range of environments: plates in the lab, controlled greenhouse stress and recovery conditions, and outdoors in a paddy. Through integration of chromatin and mRNA data, we resolve regulatory networks of genes involved in key developmental and metabolic processes and identify enriched cis-regulatory elements associated with these gene sets. We also take a systems-level network approach to define hierarchies of transcriptional regulators. These functional genomic datasets allow us to understand physiological responses; for example, we identify networks involved in cell cycle attenuation under submergence; the alteration of metabolic priorities in the exodermis under drought favoring suberin biosynthesis over iron uptake and nitrogen assimilation; and alterations to hierarchal regulators of xylem development under drought. Collectively, this resource will facilitate genetic improvements in root systems for optimal climate resilience.

Phenotypic variability of porcine organoids derived from different segments of the pig intestine

Functional Genomics and Breeding

POSTER **901**

Fany Blanc, **Mayrone Mongellaz, Frédéric Pepke, Claudia Bevilacqua, Julie Rivière, Marthe Vilotte, Giorgia Egidy-Maskos, Elisabetta Giuffra**

Université Paris-Saclay, INRAE, AgroParisTech, Jouy-en-Josas, France

Emerging in vitro systems such as organoids are expected to bridge the gaps between cell, tissue and whole animal scale knowledge, with strong ethical benefits for animal sciences (3Rs). Organoids could be used in genotype-to-phenotype research to study how genetic variation affects key cellular processes underlying complex traits e.g. affecting health and disease resilience, but efforts are needed to assess their effective ability and reliability to reflect the phenotypes of the original tissue they are derived from. Our aim was to characterize the variability of pig intestinal organoids within different gut segments and/or animals vs. the original tissue. We generated organoids from the small intestine (duodenum, jejunum, ileum) and colon from four Large White commercial pigs at slaughter age using a protocol that is convenient for commercial slaughter conditions from cryopreserved biopsies. We compared gut organoids cultured in 3D and 2D to the original tissues in terms of morphology, number and growth rate, cell composition and barrier tissue integrity. We profiled the innate immune response of organoids in 2D by measuring cytokine secretion after in vitro stimulation by various PAMPS (LPS, flagellin, Poly(I:C)), pro-inflammatory cytokines (IL-1β, IL-6 and TNFα) and heat killed bacteria (S. typhimurium and E. coli). Using principal component analyses and linear mixed statistical models, we highlighted the specificities due to both the gut regions and the animals on the measured phenotypes. We are gaining insights into this comparison using transcriptomics to decipher to what extend the pig organoids have conserved the molecular identity of the different gut segments they are derived from.

Differential cold responses and temporal gene expression patterns across and within Brassica rapa morphotypes

Functional Genomics and Breeding

POSTER **902**

Madeline W. Oravec[1], **Angela Ricono**[1], **Kathleen Greenham**[1]

1 University of Minnesota, Twin Cities, Saint Paul, MN

Developing stress tolerant crops is vital in our changing climate, but to maintain yield, plants must balance stress response with growth and productivity. Engineering constitutive activation of stress response pathways is a common strategy for enhancing tolerance, but given the metabolic drag of continuous stress response, reduced plant growth and yield can be an undesirable byproduct of this approach. Plants coordinate processes through circadian and diel regulation of gene expression and signaling. Cold stress response relies upon this regulation to balance response and growth. Additionally, some plants undergo changes through cold acclimation at moderate temperatures to bolster cold stress response and tolerance. Temporal reprogramming of cold acclimation responsive genes could enhance cold tolerance without compensating growth and yield. Brassica rapa is a close relative of Arabidopsis and is a morphologically diverse crop species, including sarson, Chinese cabbage, pak choi, turnip, and oilseed varieties (morphotypes). We characterized cold stress responses of diverse genotypes of B. rapa and found differential responses across and within morphotypes during and after cold acclimation and a freeze stress event. Using a cold acclimation time-course transcriptomic dataset, we identified variation in the temporal gene expression patterns of cold-responsive and physiologically relevant genes across genotypes, under control conditions and in response to cold acclimation. For example, cold sensitive B. rapa R500 (yellow sarson) and cold tolerant HN53 (Chinese cabbage) displayed unique patterns of variation in gene expression of a key circadian regulated cold-responsive gene, COR27, in response to cold acclimation conditions. Differential responses during acclimation likely contribute to the observed differential freeze stress responses. By coupling variation in gene expression patterns with differential stress responses, we leveraged within species variation to detect acclimation responses associated with enhanced cold stress tolerance. These findings highlight differential cold responses across and within B. rapa morphotypes and progress our understanding of diel regulation of gene expression and signaling during cold acclimation. Additionally, results provide candidate genetic engineering targets for temporal manipulation to increase cold tolerance while maintaining growth and productivity.

Spatiotemporal analysis of root responses to vegetational shade at the single cell level

Functional Genomics and Breeding

POSTER **903**

Daniele Rosado[1], **Ullas V. Pedmale**[1]

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Light is fundamental for plant development, as it provides energy for photosynthesis and information about the environment. During domestication, humans bred plants that grow well around the globe, including areas with varying day-lengths or solar irradiance. However, as arable land areas shrink as consequence of climate change, producers are forced to grow plants under suboptimal light conditions. For instance, to maximize land use, producers often grow plants under high density. However, since plants perceive the changes in light intensity and spectral composition when sun light is filtered though the dense foliage, this approach might result in yield losses if applied to shade-intolerant crops. When shaded, these plants activate a series of morphological responses to compete for sun light, including rapid elongation of stems, branches, and leaf stalks. This costly adaptation results in reduced resilience to environmental stresses, root growth and, consequently, yield. Thus, understanding how plants respond to suboptimal light is paramount for increasing food security. Indeed, great effort has been made in the past decades to understand how shoots perceive and respond to shade; however, little is known of the mechanisms controlling root growth under such conditions. In a recent work, we analyzed the transcriptome of roots from shaded Arabidopsis and tomato seedlings and found a conserved mechanism in which plants activate defense and stress programs to restrict root growth. However, manipulation of such programs might result in increased susceptibility to other environmental pressures. Furthermore, the whole root transcriptome couldn't provide enough resolution to uncover how the shade signal perceived by the shoots is transduced along the roots; nor how each of the different cell-types within the roots respond to this stimulus. Here, we analyzed the transcriptome of roots exposed to shade under high temporal and spatial resolution using single-cell transcriptomics. We've compared 11 cell-types of roots grown under controlled or shade conditions in a time-course. We've uncovered how the different cell-types in this complex organ respond to shade and propose new targets for manipulation with the aim of decreasing the negative effects of shade-avoidance in the roots, without disturbing defense and stress responses in other physiological contexts.

A first-generation porcine scImmuneAtlas: scRNAseq-based deep comparative annotation of porcine immune tissue cell types

Functional Genomics and Breeding

POSTER **904**

Christopher K. Tuggle[1], **Lance N. Daharsh**[1], **Sathesh K. Sivasankaran**[2], **Kristen A. Byrne**[2], **Juber Herrera-Uribe**[1], **Crystal L. Loving**[2]

1 Department of Animal Science, Iowa State University, Ames, IA
2 United States Department of Agriculture (USDA) Agricultural Research Service (ARS), National Animal Disease Center, Food Safety and Enteric Pathogens Research Unit, Ames, IA

An improved understanding of the porcine immune system will benefit both translational biomedical disease research and food production. We performed scRNA-seq on bone marrow, lymph node, spleen, and thymus from healthy adult pigs and sequenced a total of 65,782 cells. After QC, 50,559 cells and 18,673 genes were used for downstream analysis. Each tissue was individually mapped using non-linear dimensional reduction and distinct clusters were found and analyzed using IKAP, ROGUE, pair-wise differential gene expression testing, and random forest models. Using a combination of model defined and canonical immune cell markers, we were able to annotate each cluster as part of a diverse set of immune cell types identified. We used our previously published porcine PBMC scRNA-seq dataset to compare peripheral blood against tissue specific cell types. We were also able to confirm our cell type annotations across shared cell types. We further validated these annotations using publicly available human scRNA-seq datasets specific for each corresponding porcine immune tissue. We compared expression profiles between similarly annotated cell types between human and pig, and identified cells with expression profiles that were unique to pig immune tissues. Finally, we integrated all of the immune tissue data to compare all cell types across the four tissues, and were able to identify shared and tissue specific cell types. Our study of immune tissues will be an important resource for improved annotation of porcine immune genes and cell types and will also help to inform human translational biomedical research using pigs as a biomedical model.

Leveraging synthetic promoters to control gene expression in plants

Functional Genomics and Breeding

POSTER **905**

Anna Yaschenko[1], **Chengsong Zhao**[1], **Margot Ruffieux**[1], **Jose Alonso**[1], **Anna Stepanova**[1]

1North Carolina State University, Raleigh, NC

Understanding gene expression regulation is central to the development of biotechnological solutions for several pressing agricultural problems. The ability to precisely regulate gene expression would allow for man-driven control of plant growth. The advent of synthetic biology methods has opened the door to programmable genetic constructs in plants that confer tunability and spatiotemporal regulation to gene expression. These tools can be applied in a variety of agricultural research projects to create a multiplex of customizable promoters that drive the expression of genes of interest. However, the relationship between promoter elements within the promoter of a gene and the level of that gene's expression is not well defined. To begin tackling this question, we are building synthetic promoters harboring up to ten transcription factor binding sites using GoldenBraid technology and cloning these promoters upstream of a reporter gene. This is done by inserting one to ten copies of a protospacer, a 23bp recognition sequence for the dCas9-based synthetic transcription factor, into a neutral promoter sequence that has no known transcription factor binding sites in plants. We assemble the reporter gene and transiently co-express it along with a dCas9 activation system in Nicotiana benthamiana to test the effects of protospacer position, orientation, angular phase shift, copy number, and spacing on reporter expression. Preliminary results show that constructs with protospacers in either the sense or antisense orientation confer comparable levels of gene expression, an increase in protospacer copy number boosts reporter activity, and that there may be a decrease in expression as the distance between the protospacer and the core promoter increases. This work is expected to shed light on the rules of nature dictating promoter architecture that, in turn, determine gene expression levels, ultimately paving the way to the creation of tunable expression systems. We will apply this knowledge to build stimulus-inducible constructs from promoter elements of hormone-inducible genes to confer hormone-regulated expression to genes of interest.

Gene Expression Tools, Techniques and Studies

A simple, customizable approach to selectively remove abundant unwanted RNAs and improve the sensitivity of transcript detection in any species

Lynne Apone[1], **Deyra N. Rodríguez**[1], **Bradley W. Langhorst**[1], **Kaylinnette Pinet**[1], **Gautam Naishadham**[1], **Jian Sun**[1], **Keerthana Krishnan**[1], **Fiona J. Stewart**[1], **Eileen T. Dimalanta**[1]

1 New England Biolabs, Ipswich, MA

The presence of highly abundant transcripts, with minimal biological interest in total RNA samples, creates a challenge to whole-transcriptome sequencing. These highly expressed transcripts dominate readouts, obscuring the detection of more informative lower abundance transcripts. Here, we present a simple, customizable approach to enrich for RNAs of interest by eliminating unwanted RNAs. This method is based on hybridization of probes to the targeted RNA and subsequent enzymatic degradation of the selected RNAs. The probe sequences confer RNA removal specificity and can be designed to deplete unwanted RNA from any organism. We used this method to design a set of probes to efficiently remove rRNA from a broad range of flowering plants including corn, rice and maze. In addition, we developed a simple, user-friendly web tool that enables custom depletion of any RNAs of interest without the aid of a bioinformatician. We used this web tool and depletion method to remove rRNA from total RNA of various species, including the pig Sus domesticus and mosquito Aedes aegypti. Additionally, we used this approach to supplement an existing anti rRNA probe set to improve depletion efficiency and expand the usefulness of the existing probe set.

Using strand-specific RNA sequencing we measured depletion efficiency and transcript expression. We achieved high depletion efficiency (up to 99%) for all targeted RNAs across different species, while maintaining transcript abundance of non-targeted RNA. This translated into an enrichment of RNAs of interest and an increased depth of sequencing coverage.

The method described here is a simple and reliable solution to improve sensitivity in RNA-Seq studies. The design tool offers flexibility and control over the probe design process facilitating a customized and economical RNA sequencing experience. Importantly, the depletion and library construction method are amenable to high throughput sample preparation and robotic automation.

Single-cell transcript mapping in a monocot species using Molecular Cartography

Gene Expression Tools, Techniques and Studies

POSTER 1101

Philip Benfey[1], **Mingyuan Zhu**[1]

1 Department of Biology, Duke University, Durham, NC

Root development is a dynamic process in which distinct cell groups are established in a defined spatiotemporal manner. Identities of individual cells can be characterized by their unique gene expression profiles. Recent advances in plant single cell genomics promote the understanding of developmental mechanisms. However, our current knowledge of root development is based mainly on the dicot model plant, Arabidopsis. Little is known about transcriptional programs that underlie root cell type specification in monocot crops. Here, we harmonize single-cell RNA sequencing (scRNA-seq) data from over 49,000 rice root cells to construct a rice root Atlas. We are able to identify putative clusters for most cell types in our Atlas data based on published reference expression profiles. To validate cell type identity, we identify marker genes that should represent eight major cell types in rice roots. To map the local gene activity, we use Molecular Cartography, an optimized multiplexed Fluorescence in situ hybridization (FISH) technology, to simultaneously explore the in-situ gene-expression profiles of up to 50 genes. We present the cellular localization of marker RNAs in sections of rice roots. Our goal is to use multiplexed FISH data to validate both the enrichment and specificity of markers at single-molecule level. In addition, the spatial gradients of marker gene expression may reveal potential developmental trajectories of rice root cells. Overall, our study demonstrates the application of spatial transcriptomics in crop plants. The integration of single-cell and spatial transcriptomics leads to high-confidence annotation of specific cell types. The spatial expression patterns of marker genes at both the cellular and tissue level provide insights into the mechanisms underlying cell type specification.

Mapping cellular divergence in cereals using single-cell and single-nuclei RNA sequencing and Molecular Cartography

Gene Expression Tools, Techniques and Studies

POSTER 1102

Kenneth Birnbaum[1], **Bruno Guillotin**[1]

1 Center for Genomics and Systems Biology, New York University, New York, NY

Single-cell RNA-seq and spatial transcriptomics methods open the opportunity to compare evolutionary changes in gene expression among species. Here, we apply both single-cell and single-nuclei RNA-seq to maize, sorghum, and Setaria roots. A comparison of single-cell and single-nucleus RNA-seq shows high correlation in multiple species, with complementary strengths and weaknesses in the two profiling approaches. We combined both approaches for all three species, generating a pan-genome expression dataset of highly conserved and evolutionary labile transcripts genome wide. We used Molecular Cartography, a high throughput spatial transcriptomics platform, to map the expression profiles of key developmental genes derived from our combined single-cell and single-nuclei RNAseq techniques in maize. The high-resolution spatial gene expression map of maize obtained through spatial transcriptomics allowed us to construct a detailed map of homologous cell identities across the three taxa using all the data. In a sensitive analysis of adaptation following genome duplication, a comparative transcriptome analysis within cell types showed strong signals of dosage compensation between surviving homeologs in the two maize genomes relative to unduplicated sorghum genes. To detect changes in cellular divergence within this taxonomic clade, we developed new techniques to examine changes in gene expression in orthologous cell types across maize, sorghum, and outgroup Setaria. The root cap columella cells showed the most rapid change, particularly in maize. Genes involved in mucilage production appear to have been recruited from cortex/epidermis in sorghum into the root cap in maize. The results show the power of comparative single-cell profiling to rapidly map conserved cellular markers, analyze genome evolution, and associate gene expression programs to agronomically important traits.

Identifying meristem-type specific regulatory circuits in complex inflorescences of barley using transcriptomics at single cell resolution

Gene Expression Tools, Techniques and Studies

POSTER 1103

Edgar Demesa-Arevalo[1], **Róisín Ní Chuaig**[1], **Isaia Vardanega**[1], **Leonie Schrage**[1], **Hajer Youssif**[1], **Jan-Niklas Seiffert**[1], **Tobias Lautwein**[2], **Karl Köhrer**[2], **Maria von Korff Schmising**[3], **Rüdiger Simon**[3]

1 Institute for Developmental Genetics and Cluster of Excellence on Plant Sciences, Heinrich Heine University, Düsseldorf, North Rhine-Westphalia, Germany
2 Biological and Medical Research Center (BMFZ), Genomics and Transcriptomics Laboratory (GTL), Medical Faculty, Heinrich Heine University, Düsseldorf, North Rhine-Westphalia, Germany
3 Institute for Plant Genetics, Heinrich Heine University and Cluster of Excellence on Plant Sciences, Heinrich Heine University, Düsseldorf, North Rhine-Westphalia, Germany

Novel transcriptomics technologies with single-cell resolution allow the study of lesser established model organisms. Such studies are particularly useful in monocot models such as barley and other temperate cereals. In these monocots, the inflorescences are, compared with those of Arabidopsis, considerably more complex and generated by several distinct meristems. These meristem types are the inflorescence meristem (IM), which is the shoot apical meristem (SAM) after floral induction, the triple spikelet meristem (TSM) generated at the flanks of the IM which further develops into the central spikelet meristem (CSM), and two lateral spikelet meristems (LSM). The spikelet meristems give rise to the floral meristem (FM), that generate all floral organs. However, our knowledge of the gene regulatory networks governing the developmental diversity between these meristem types and their developmental trajectories is very limited and mostly focused on a small set of genes that were analysed due to available mutant phenotypes.

The growing spikes of monocots offer a gradient of developmental stages, with the potential to capture all meristem types and their intermediate developmental states (from IM to floret organs). Aiming to better understand these processes, we standardized protocols to isolate protoplasts from barley spikes compatible with two single-cell sequencing platforms based on microfluidics or micro-well plates. These experiments allowed us to sequence 6241 cells arrayed in 17 clusters containing genes known to regulating spike development. Due to the limited annotation of gene specific/cell-type specific clusters in barley meristems, we used Molecular Cartography to visualize the gene expression patterns of our candidate genes and clusters at cellular resolution. This novel technology allowed us to simultaneously capture the expression profiles of 100 genes selected from our single-cell transcriptomics datasets. Obtaining such detailed spatiotemporal dimension for our gene expression data provided us with important insights into the gene regulatory networks governing spike formation and helped us select potent candidate genes for further functional characterization.

Exploring chicken early transcripts landscape with Iso-Seq

Gene Expression Tools, Techniques and Studies

POSTER 1104

Sébastien Guizard[1*], **Katarzyna Miedzinska**[1], **Jacqueline Smith**[1], **Jonathan Smith**[1], **Richard Kuo**[1], **Davey Megan**[1], **Mick Watson**[1]

1 The Roslin Institute, Easter Bush Campus, Midlothian, United Kingdom
*sguizard@ed.ac.uk

The EU H2020 Gene-Switch project is a collaborative endeavour that will produce new genome information to enable the characterization of genetic and epigenetic determinants of complex traits in the two monogastric species (chicken and pig) that are the primary sources of meat worldwide.

The characterization of genome functional elements is being made through a wide diversity of analysis methods (RNASeq, miRNA-Seq, ATAC-Seq, WGBS, ChIP-Seq, IsoSeq, Promoter Hi-C), on seven tissues at three developmental stages (pig: 30, 70 days post-fertilization, new born and chicken: E8, E15, hatched).

Iso-Seq, with circular cDNA sequencing, generates accurate full-length transcripts (<1% error). It allows direct identification of gene intron-exon structure and thus better isoform detection. This method has growing utility and has been used to re-annotate genomes but no standard pipeline has yet been released.

We are thus developing an nf-core isoseq pipeline, which enables insight into the chicken early transcriptomic landscape. The pipeline generates HIFI sequences, maps transcripts to the genome and reduces gene model redundancy with TAMA collapse. Thanks to the nextflow language and containerization technologies, it can be easily run on a wide range of cluster configurations.

We processed the Iso-Seq data for each chicken sample to reveal the transcriptome for a given tissue at a given time point. Comparison of annotations gives a first look at transcriptional landscapes across tissues and the dynamic over time development.

“This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the grant agreement n° 817998”.

Unleashing the power of single cell technologies: combining single-nuclei RNA-seq and Molecular Cartography in soybean

Gene Expression Tools, Techniques and Studies

POSTER 1105

Marc Libault[1], **Sandra Thibivilliers**[1], **Sutton Tennant**[1], **Sergio Alan Cervantes Perez**[1], **Samik Bhattacharya**[2], **Frank Reinecke**[2], **Jasper Kläver**[2]

1 University of Nebraska-Lincoln, Lincoln, NE
2 Resolve Biosciences, Monheim am Rhein, Germany

The sustainable production of crops necessary to meet demand requires the development of innovative new strategies. Genomic engineering represents an attractive solution to develop the next generation of high-performing crops. However, the implementation of such strategies requires a deep and high-resolution understanding of the role and activity of crop genes. Considering that plants are composed of various cells and cell types sharing the same genetic information, plant cells differentially use their genomic information to establish cell-type-specific transcriptional programs and to gain unique biological functions.

To precisely characterize the activity of each gene in a cell-type , we and others reported the use of single cell/nucleus RNA sequencing (scRNA-seq/sNucRNA-seq) and single nucleus Assay for Transposase Accessible Chromatin sequencing (sNucATAC-seq) technologies; mostly on the model plant Arabidopsis thaliana. However, even though these technologies prove useful in characterizing cell type-specific transcriptomes, the functional annotation of the identified clusters remains challenging; especially when considering non-model plant species.

In the present study, we demonstrated the use of Molecular Cartography alongside sNucATAC-seq to add a spatiotemporal dimension to our Soybean root and nodule single cell expression profiles. This technology allowed us to visualize the expression of 100 preselected genes at cellular resolution. Given our results, we expect single cell/nucleus technologies paired with high resolution spatial analysis to provide a new understanding of plant gene activity and function. Such detailed knowledge will be essential in establishing meaningful genomic engineering strategies to improve crop performance.

The wheat pan-transcriptome

Gene Expression Tools, Techniques and Studies

POSTER 1108

Manuel Spannagl[2], **Benjamen White**[1], **Thomas Lux**[2], **Rachel Rusholme-Pilcher**[1], **Sabrina J. Ward**[1], **Gemy Kaithakottil**[1], **David Swarbreck**[1], **Anthony Hall**[1], **10+ Wheat Genome Project**

1 Earlham Institute, Norwich Research Park, Norwich, United Kingdom
2 Plant Genome and Systems Biology (PGSB), Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany

Wheat is the most widely cultivated crop in the world with over 215 million hectares grown annually. To meet the demands of a growing global population, breeders face the challenge of increasing wheat production by around 60% within the next 40 years. The 10+ Wheat Genomes Project recently sequenced and assembled an additional 15 wheat cultivars to develop our understanding of genetic diversity and selection within the pan-genome of wheat (Walkowiak et al, Nature 2020). In this work we present a wheat pan-transcriptome with de novo gene predictions and differential expression analysis for 14 wheat cultivars over five different tissues and whole seedlings sampled at dusk/dawn. Analysis of these de novo gene annotations, available in Ensembl Plants release 52, facilitated the discovery of novel genes specific to particular cultivars, including genes absent from the existing Chinese Spring reference, and copy number variations including tandem arrays in high-resolution. Through our gene expression analysis, we improve the definition of the core and dispensable genomes, highlight tissue-specific genes, and reveal changes in bias of sub-genome homeolog expression between cultivars.

This work provides both a novel resource for the wider wheat community and reveals several important findings about the transcriptional diversity in wheat beyond a single reference genome.

Validation of maize ear and root single cell data by spatial transcriptomics using Molecular Cartography

Gene Expression
Tools, Techniques
and Studies

POSTER **1109**

David Jackson[1], **Ken Birnbaum**[2], **Jack Xu**[1], **Bruno Guillotin**[2]

1 University of Nebraska-Lincoln, Lincoln, NE
2 Resolve Biosciences, Monheim am Rhein, Germany

The sustainable production of crops necessary to meet demand requires the development of innovative new strategies. Genomic engineering represents an attractive solution to develop the next generation of high-performing crops. However, the implementation of such strategies requires a deep and high-resolution understanding of the role and activity of crop genes. Considering that plants are composed of various cells and cell types sharing the same genetic information, plant cells differentially use their genomic information to establish cell-type-specific transcriptional programs and to gain unique biological functions.

To precisely characterize the activity of each gene in a cell-type , we and others reported the use of single cell/nucleus RNA sequencing (scRNA-seq/sNucRNA-seq) and single nucleus Assay for Transposase Accessible Chromatin sequencing (sNucATAC-seq) technologies; mostly on the model plant Arabidopsis thaliana. However, even though these technologies prove useful in characterizing cell type-specific transcriptomes, the functional annotation of the identified clusters remains challenging; especially when considering non-model plant species.

In the present study, we demonstrated the use of Molecular Cartography alongside sNucATAC-seq to add a spatiotemporal dimension to our Soybean root and nodule single cell expression profiles. This technology allowed us to visualize the expression of 100 preselected genes at cellular resolution. Given our results, we expect single cell/nucleus technologies paired with high resolution spatial analysis to provide a new understanding of plant gene activity and function. Such detailed knowledge will be essential in establishing meaningful genomic engineering strategies to improve crop performance.

Genome Annotation, AI and Microbiome Science

Genome Annotation, AI and Microbiome Science

POSTER 101

ChronoGauge: telling the time in crops using AI

Joshua Colmer[1], **Connor Reynolds**[1], **Hannah Rees**[1], **Antony Dodd**[2], **Anthony Hall**[1]

1 Earlham Institute, Norwich Research Park, Norwich, United Kingdom
2 John Innes Centre, Norwich Research Park, Norwich, United Kingdom

ChronoGauge is an open-source AI-based approach to telling the circadian time of a crop by using an artificial neural network and a set of rhythmically-expressed biomarker genes.

The circadian clock is an endogenous cycle entrained by external cues such as light and temperature that enables plants to synchronise their behaviour and processes with the environment. To predict the endogenous time of a plant, we first identify genes that are rhythmically expressed and cluster them by their phases. Using a customised sequential feature selection method, we optimise the rhythmic genes to use in our neural network by minimising our model's error between the actual times and predicted times.

We have trained ChronoGauge on labelled, publicly available circadian Arabidopsis thaliana RNA-Seq datasets and tested it on additional circadian datasets as well as datasets that include clock mutants and temperature variations. By training an ensemble of models using multiple biomarker feature sets, we identify in which mutant and varied environmental datasets certain biomarker sets are effective and ineffective predictors – enabling biological inference regarding the effects of clock mutations and environmental changes on a plant's internal timing, gene expression and regulation.

In addition to applying ChronoGauge to a variety of Arabidopsis datasets, we have tested the tool on internal wheat, public Brassica, and public soybean circadian datasets as well as dusk and dawn samples from a diverse range of wheat cultivars. ChronoGauge's accurate predictions on Arabidopsis, wheat, Brassica, and soybean demonstrate the wide applications available for optimising the timing of agricultural processes.

In the future, we plan to apply ChronoGauge to transcriptome data from the 1001 Genomes Project to investigate circadian clock variation and phenotypes in Arabidopsis thaliana accessions. In addition to this, we will continue to apply ChronoGauge to different Brassica species as well as wheat cultivars – generating circadian clock biomarkers for each of these crops.

Genome Annotation, AI and Microbiome Science

POSTER 102

Global assessment of translational regulation within bovine tissues

Anna K. Goldkamp[1], **Darren E. Hagen**[1]

1 Department of Animal and Food Sciences, Oklahoma State University, Stillwater, OK

Due to their crucial role in translation, tRNAs were once thought to be ubiquitously expressed in all tissues and species. However, recent reports suggest elevated methylation levels at tRNA gene clusters can result in transcriptional repression and act as an epigenetic mechanism to fine tune tRNA expression. In addition, differentially methylated regions (DMRs) have been associated with Large Offspring Syndrome (LOS) in cattle. Previously, we utilized high-throughput tRNA sequencing to investigate tRNA expression in skeletal muscle and liver tissue of day 105 artificial insemination-conceived (Control-AI), ART-conceived with a normal body weight (ART-Normal), and ART-conceived bovine fetuses (ART-LOS). Through this work, we identified alterations in tRNA expression due to in vitro fertilization and overgrowth. In addition, we identified tissue specificity of these anticodon: codon interactions, suggesting that there is biased tRNA expression to potentially regulate protein synthesis.

One of the greatest factors in determining translation fidelity and efficiency is codon usage bias. Sources of sequence variation, like codon usage, often drive differences in elongation rate and gene expression. Highly expressed genes are thought to be codon-biased to support efficient translation, in which the encoded codons correspond to highly abundant tRNAs. To investigate the dynamics of translational regulation within bovine tissues, ribosome profiling was used to capture transcripts being actively translated by the ribosome. Following nuclease digestion, we performed high throughput sequencing of ribosome bound fragments in order to identify actively translated regions at a single-nucleotide resolution. Using RIBOseq and RNAseq datasets, we identified a total of 307 genes with differential translation efficiencies across muscle, kidney, and liver tissues. GO enrichment analysis revealed that these translationally regulated genes were related to tissue specific molecular function. Further, synonymous SNPs were once considered to be silent due to the degeneracy of the genetic code. However, synonymous SNPs may disturb protein abundance and function through alterations in translational efficiency. Specifically, codons pairing to high-abundance tRNAs could be translated more quickly compared to low-abundance tRNAs. Therefore, we have also performed tRNA sequencing in order to investigate the relationship between tRNA expression and translational stalling events due to synonymous mutations.

The wheat pan-transcriptome: identifying consensus gene networks and developing de novo annotations

Genome Annotation, AI and Microbiome Science

POSTER 103

Anthony Hall[1], **Rachel Rusholme-Pilcher**[1], **Thomas Lux**[2], **Benjamin White**[1], **Gemy Kaithakottil**[1], **David Swarbreck**[1], **Manuel Spannagl**[2], **10+ Wheat Genome Project**

1 Earlham Institute, Norwich Research Park, Norwich, United Kingdom
2 Plant Genome and Systems Biology (PGSB), Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany

Wheat is the most widely cultivated crop in the world with over 215 million hectares grown annually. To meet the demands of a growing global population, breeders face the challenge of increasing wheat production by around 60% within the next 40 years. The international wheat community, as part of the 10+ Wheat Genomes Project, has recently sequenced and assembled an additional 15 wheat cultivars to develop a pan-genome of wheat, quantify available diversity and identify genomic regions under selection. The pan-transcriptome initiative builds on this work, conducting de novo annotation and differential expression analysis on these diverse assemblies. These de novo annotations, available soon on Ensembl plants, represent the most complete and uniform structural annotations to date for wheat. They have enabled the computation of a comprehensive set of orthogroups, which has facilitated the discovery of genes novel to specific cultivars, high-resolution tandem arrays and genes absent from the existing Chinese Spring reference. Detailed co-expression analysis of these varieties has revealed a core set of consensus network modules. These modules demonstrate patterns of expression that are conserved across both cultivars and tissue types. Analysis of the genes within these modules has revealed triads that are split between divergent expression modules and genes that show cultivar-specific expression profiles. Through interaction with the wider wheat community, presence/absence variation, changes in gene models and patterns of expression will be associated with local adaptations and cultivar-specific traits giving valuable insight into the diversity of wheat beyond a single reference genome.

Workflows to support high-throughput functional annotation of genomes

Genome Annotation, AI and Microbiome Science

POSTER 105

Amanda Cooksey[1], **Surya Saha**[2], **Monica Polechau**[3], **Fiona McCarthy**[4]

1 CyVerse, University of Arizona, Tucson, AZ
2 Boyce Thompson Institute, Ithaca, NY
3 National Agricultural Library, Agricultural Research Service, United States Department of Agriculture (USDA), Beltsville, MD
4 School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ

Large-scale, co-ordinated efforts are underway to identify functional elements within animal genomes through such effort as the Functional Annotation of Animal Genomes and the i5k projects. However, identification of functional elements is only the first step for annotation; genetic elements must also be linked with biological functions. For example, transcriptome studies rely on Gene Ontology (GO) and pathway information to develop a model of perturbed gene expression, and gene products that have no GO or pathway annotations are dropped from this functional modeling. Likewise, comparative studies may identify gene changes between lines, breeds, or species but if the function of the gene is not known, the impact of gene changes is not able to be fully understood. Improved genome assemblies and gene identification is rapidly enabling the identification of novel genes in a broad range of species, including noncoding RNA genes that play key roles in regulating biological processes. While established workflows exist to provide rapid, genome-scale functional annotations for protein coding genes, there is no equivalent approach to predict functions for noncoding genes. We provide an assessment of current functional annotation quality and discuss preliminary work developing novel workflows for functionally annotating ncRNAs identified through large-scale sequencing.

Conserved and rewired metabolic modules revealed by cross-species co-expression

Genome Annotation, AI and Microbiome Science

POSTER 106

Michael Passalacqua[1], **Hamsini Suresh**[1], **Megan Crow**[1], **Jesse Gillis**[2]

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2 University of Toronto, Toronto, Ontario, Canada

Plant metabolism produces compounds with an extensive range of functions, many of which are critical in defense and stress tolerance. By identifying metabolic circuits conserved across species, we can elucidate essential systems, separating them from niche metabolic pathways. In order to evaluate conservation of a metabolic module, we turn to RNA-sequencing data, which captures both sequence and regulatory changes in function. Typically, metabolism is split into two broad categories – primary metabolism genes essential for growth, and secondary metabolism genes that perform specialized functions. Primary metabolism genes are highly conserved, as their functions are similar across all plant cells. In contrast, secondary metabolism genes are highly variable, as species specific pressures shape their evolution. Utilizing over 10,000 publicly available RNA-sequencing datasets from 16 plant species, we constructed robust co-expression networks specific to each species. Leveraging abundant GO annotations in 10 of these species, we identified deeply conserved metabolic modules whose co-expression partners are remarkably retained even across phylogenomic distances of hundreds of millions of years. Furthermore, we also revealed modules being rapidly rewired towards new functions, differentiating them from sets of genes that are not members of a shared regulatory module. We find that secondary metabolism modules are frequently repurposed for novel functions across species, with 65% of identified secondary metabolism modules experiencing major shifts in co-expression partners. When these rewired modules were further investigated, we found that those experiencing the biggest shifts are either specific to just a few genera (glucosinolate modules) or are related to pathogen defense (terpenoids). These findings support and extend previous work, which has revealed defense genes as an area of major evolutionary importance and has found that primary metabolism genes have very high sequence conservation. We identify both highly conserved and rapidly rewired metabolic modules across a broad swath of the plant kingdom, highlighting how the rewiring of modules, both temporally and in response to stress, can lead to functional shifts in the patterns of metabolic circuitry.

An atlas of tissue- and sex-specific regulatory elements in the equine genome

Genome Annotation, AI and Microbiome Science

POSTER 107

Jessica L. Petersen[1], **Alexa M. Barber**[1], **Nicole B. Kingsley**[2], **Sichong C. Peng**[2], **Elena Giulotto**[3], **Carrie J. Finno**[2], **Rebecca R. Bellone**[2], **Ted Kalbfleisch**[4]

1 University of Nebraska, Lincoln, NE
2 University of California-Davis, Davis, CA
3 University of Pavia, Pavia, Italy
4 University of Kentucky, Lexington, KY

The equine Functional Annotation of Animal Genomes (FAANG) project has focused primarily on the analyses of 8 tissues of adult Thoroughbred mares (N=2) and stallions (N=2). The tissues studied include adipose, parietal cortex, heart, liver, lung, skeletal muscle, and ovary/testis, which represent the "core" tissues evaluated across FAANG species, and the laminae, which is of particular importance to the horse. Additionally, the community has "adopted" tissues that are of relevance for equine projects including over 40 tissues for transcriptome analysis and four tissues for histone mark evaluation (spleen, metacarpal 3, sesamoid, and skin) in the mares. Transcriptome data demonstrated the expected uniqueness of expression in the testis of the stallions, the ovaries of the mares, and in the brain (parietal cortex) of both sexes. ChIP-seq for histone modifications corresponding to enhancers (H3K4me1), promoters (H3K4me3), activators (H3K27ac), and repressors (H3K27me3) in the mares identified hundreds of thousands of peaks covering from 0.6% (H3K27me3 in brain and adipose) to 6.28% (H3K27me3 in skin) of the genome. Transcription factor binding motifs enriched in tissue-specific enhancers were also identified. ChIP analyses from the same tissues of the stallions revealed similar patterns of peak prevalence and tissue-specificity with H3K4me3 of the testis having a large proportion of peaks unique to that tissue. Notable differences were observed in comparing the ChIP data from mares and stallions; peak number, size, and genome coverage were greater in the stallions compared to that found in the mares. For instance, on average the stallion peaks were nearly 30% longer than those of the mares. It is hypothesized that differences in library preparation and sequencing (mares: 50bp single-end vs. stallions: 50bp paired-end) are confounding comparisons between sexes, requiring additional efforts to optimize analyses for robust comparisons between sexes and across tissues. In addition to these data, ATAC-seq, reduced representation bisulfite sequencing, and studies of centromere function have been conducted in the 8 prioritized tissues. Data also continue to be collected to evaluate CTCF-binding sites and isoform expression. Together, this represents an expanding resource available for identifying genome function and characterizing regulatory mutations in the horse, both between sexes and among tissues.

Identification of transcription factors controlling photosynthesis using machine learning approaches

Genome Annotation, AI and Microbiome Science

POSTER 109

Donat Wulf[1,2], **Janik Sielemann**[1,2], **Sanja K. Zenker**[1], **Wiebke Halpape**[1], **Fabian J. Krüger**[3], **Andrea Bräutigam**[1]

1 Computational Biology, Faculty of Biology, CeBiTec, Bielefeld University, Bielefeld, Germany
2 Graduate School DILS, Bielefeld Institute for Bioinformatics Infrastructure, Bielefeld University, Bielefeld, Germany
3 Technische Universität Berlin, Berlin Germany

Climate change challenges in agriculture require an increasing speed in adaptation to keep yields high under stress conditions. This requires a better understanding of gene regulatory networks (GRNs) in plants, for example the regulation of C3 and C4 photosynthesis to predict phenotypic changes upon changes in the GRN. To answer this question we use three computational approaches. To associate transcription factors (TFs) with traits we infer GRNs from expression data using random forest machine learning. Associations for six monocot crop plants were predicted and compared with Arabidopsis, Marchantia and Chlamydomonas. Typically, ML prediction of GRNs result in false positives of up to 80%. We developed enrichment analyses to overcome high error rates in the prediction of GRNs and enable identification of candidate TFs for traits. We hypothesized that GRNs are conserved and found a high degree of conservation of GRNs. The conservation follows their phylogenetic relationship. Conservation was validated for selected candidates by examining transcription factor binding sites across Arabidopsis, Marchantia and maize with cross-species DAP-seq. Conservation in at least three different species GRNs enables identification of high confidence TF candidates. Currently 36 candidates are pipelined for knockout, overexpression and RNA-seq. Data for one promising photosynthesis candidate will be available in February.

To better predict DNA::transcription factor interaction, we test which parameters beyond sequence influence DNA binding. Experimental binding site investigation by ampDAP-seq suggests that actual binding sites are on average 14-fold more specific than sequence motif alone predicts. To test the role of DNA shape for specific binding, we trained a random forest regressor on DNA shape. On average we almost double prediction accuracy and conclude DNA shape has a key role. Predictions for sequences not present in the genome are validated by EMSA (Sielemann et al. 2021, Nature Communications).

To leverage data on open chromatin regions, we developed new network inference approaches from scATAC-seq data in maize and compared results with RNA-seq GRNs. We are currently integrating binding data, RNA-seq GRNs and ATAC-seq GRNs to improve accuracy and prediction capability.
The combination of machine learning and experimental approaches provide insights in GRNs which identify candidate TFs controlling photosynthesis.

Porcine immune regulatory networks at single cell resolution using scATACseq

Genome Annotation, AI and Microbiome Science

POSTER 110

Pengxin Yang[1], **Lance Daharsh**[1], **Juber Herrera Uribe**[1], **Kristen Byrne**[2], **Crystal Loving**[2], **Christopher Tuggle**[1]

1 Iowa State University, Ames, IA
2 United States Department of Agriculture (USDA), Agricultural Research Service (ARS), National Animal Disease Center (NADC), Food Safety and Enteric Pathogens Research Unit, Ames, IA

Identifying the cell type specific epigenomic regulatory network of porcine immune cells will provide important functional annotation for improving pig disease resilience. We used scATACseq to profile the chromatin accessibility of PBMC cells collected from two FAANG founder pigs. 17,230 nuclei passed QC and 110,444 common accessible peaks of fragments were found (CellRanger ATAC). We identified 36 clusters of cells, using a shared nearest neighbor clustering algorithm (Seurat/Signac). Regulatory regions potentially controlling cluster-specific gene expression were identified by measuring differentially accessible peaks (DAPs) among clusters. Predicted gene activity profiles were created by assigning fragments to nearby genes (<2000 bp from TSS). Each cluster was classified in 2 ways: (1) by evaluating the predicted gene activity and the chromatin accessibility of differentially accessible cis-elements close to genes with known expression in specific immune cell types, and (2) by integrating gene activity with our published PBMC scRNAseq data. These methods provided highly concordant cell type annotations.

Using these cell-type labels from integration with scRNAseq, we found and annotated 14,092 cell type-specific DAPs. These results predict many novel regulatory elements associated with specific cell type expression patterns. For example, we identified a cis-element within the B-cell marker gene PAX5 that is only differentially accessible in B cells. Further, GO analysis of genes with DAPs of each cell type identified significantly enriched GO terms highly related to immune functions of the corresponding cell type.

To identify modules of interacting genomic elements predicted to control nearby genes, cis-acting Co-accessibility networks (CCAN) using Cicero was constructed for each cell type. For example, we found 4257 co-accessible modules in a CCAN analysis of chromosome 1. Transcription factor (TF) binding motif enrichment analysis on the DAPs of each cell type was performed using HOMER, and, combined with the CCAN analysis, will predict TF regulating genes through these newly identified regulatory element networks.

Our results demonstrate the power of scATAC to identify and annotate cell type-specific regulatory elements and predict TF binding to CCAN elements in the porcine immune system. Putative target genes of such identified TFs can be explored to help elucidate mechanisms regulating immune gene expression.

Genome Annotation, AI and Microbiome Science

POSTER 112

Fish microbiomes 101: disentangling the rules governing marine fish mucosal microbiomes across 101 species and its implications for probiotic discovery in aquaculture

Jeremiah Minich[1], **Andreas Härer**[2], **Joseph Vechinski**[2], **Ben Frable**[2], **Zachary Skelton**[2], **Emily Kunselman**[2], **Todd Michael**[1], **Eric E. Allen**[2], **Rob Knight**[3]

1 The Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA
2 Scripps Institution of Oceanography, University of California, San Diego, La Jolla CA
3 Department of Pediatrics, University of California, San Diego, La Jolla CA

Aquaculture is the fastest growing agriculture sector with over 400 species of fish being reared. Along with sparse genomics tools a major challenge in fish production is high hatchery mortality rates (especially in marine fish) which is partially driven by microbial dysbiosis. Most probiotics used in aquaculture are from terrestrial livestock systems and therefore may not be the most effective strains. Thus, we designed an experiment to understand the extent to which specific biological and environmental factors influence the fish microbiome including if microbiomes were more similar in genetically similar fish. We collected 101 species of marine fishes from Southern California spanning 22 orders, 55 families, and 83 genera and representing ~25% of local marine fish diversity and compared microbiomes (gill, skin, midgut, and hindgut) using alpha, beta, and gamma diversity along with establishing a novel method to estimate microbial biomass. Body site location was the most significant driver in microbial community composition which suggests that probiotics should be developed for body sites independently. Host genetic similarity was associated with microbiome similarity in the gill, skin, and hindgut, but not midgut ($P<0.05$, Mantel test). For neritic fishes, gill microbial biomass was negatively correlated with distance from shore. Fishes with high gill microbial biomass densities were associated with microbes originating from the marine sediment rather than sea water despite that overall sea water contributed more microbes than marine sediment to the fish mucus for all four body sites tested ($P<0.05$, Mann-Whitney). Fish are the most diverse and widely distributed vertebrates, yet least studied in microbial ecology thus we performed a meta-analysis against other vertebrate classes (birds=216 species, mammals = 208 species, reptiles = 52 species, amphibians = 20 species). Mammals had the highest gamma diversity when controlling for species number while fishes have the highest percent of unique microbes (92% unique to fishes). The majority of microbial diversity in fishes exists in body sites other than the hindgut, principally the midgut, gill, and skin and that when included collectively may be 5.5 times higher than hindgut. These body sites of fish may represent a novel source of Bacillus and Lactobacillus probiotics.

Genome Annotation, AI and Microbiome Science

POSTER 113

The functional annotation of the ovine reference genome

Brenda M. Murdoch[1], **Kimberly M. Davenport**[1], **Mazdak Salavati**[2], **Emily Clark**[2], **Alan Archibald**[2], **Stephen N. White**[3], **Michelle R. Mousel**[4], **Kim C. Worley**[5], **Suraj Bhattarai**[6], **Stephanie D. McKay**[6], **Shangqian Xie**[1], **Alex Caulton**[7], **Shannon Clarke**[7], **Rudiger Brauning**[7], **Tracy Hadfield**[8], **Benjamin Rosen**[9], **Timothy P.L. Smith**[10], **Noelle E. Cockett**[8]

1 Department of Animal, Veterinary, and Food Science, University of Idaho, Moscow, ID
2 The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Scotland
3 Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA
4 United States Department of Agriculture (USDA), Agricultural Research Service (ARS), Animal Disease Research Unit (ADRU), Pullman, WA
5 Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX
6 Department of Animal and Veterinary Science, University of Vermont, Burlington, VT
7 AgResearch
8 Utah State University, Logan, UT
9 USDA, ARS, Animal Genomics and Improvement Laboratory (AGIL), Beltsville, MD
10 USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE

The annotation of regulatory elements in the genome is critical for understanding complex phenotypic traits related to production traits and health. The Ovine Functional Annotation of Animal Genomes (FAANG) Project aims to characterize transcriptional regulatory elements across the sheep reference genome in a large collection of tissues to facilitate a better understanding of the mechanisms influencing these traits. Functional assays, including evaluation of RNA transcript abundance (RNA-seq), determination of transcription start sites by cap analysis of gene expression (CAGE), localization of expression-related histone modifications by chromatin immunoprecipitation with sequencing (ChIP-seq), identification of "open" euchromatin by assay for transposase-accessible chromatin with sequencing (ATAC-seq), determination of DNA modification status by whole genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS) were performed on a subset of tissues from the animal used for the reference genome. Chromatin states depicting promoters, enhancers (active, poised, and repressed), and accessible chromatin across the genome were defined with ChromHMM and compared across tissues. Active promoter and enhancer states were found to reside in open chromatin regions associated with hypomethylated regions, whereas poised and repressed enhancers were not commonly observed in open chromatin and tended to be associated with hypermethylated sites. The combined data provide a high-resolution annotation of the expressed and regulatory regions of the ovine genome and represent a valuable resource to facilitate a deeper understanding of how gene-regulation control influences complex traits in this globally important livestock species.

Genotyping Tools and Techniques

Quality analysis of 1,536 unique dual barcodes for agrigenomics applications

Marcel Brun[1], **Matthew Stull**[1], **Joshua Hill**[1], **Emma Schulze**[1], **Richard P. Metz**[1], **Charles D. Johnson**[1]

1 Texas A&M AgriLife Research Genomics and Bioinformatics Service, College Station, TX

Agrigenomics is a growing field of research focused on the challenge of feeding the world population by understanding the genomes of crop and livestock. New approaches such as low coverage sequencing with imputation has dramatically reduced the cost of agricultural focused genotyping and accelerated the need for massive scale sample processing. A solution to large multiplex sequencing was the use of combinatorial dual index barcodes, but index misassignment make this approach not recommendable. As new unique dual barcodes (UDI) have come on line, barcode quality is still an important issue, in particular for low coverage sequencing: demultiplexing removes reads based on mismatches in the barcode sequence, hence lower quality barcodes result in lower sequencing yields, affecting downstream analysis. In low coverage sequencing lower yielding samples may receive insufficient coverage, resulting possibly on the loss of the sample in the analysis stage.

In this study we analyzed base calling error and sequencing performance of 1,536 UDI, which provides solution to large number of samples and high-quality multiplexing, to determine if there exists a relationship between barcode sequence and base quality. To this end, we evaluated the barcodes performance across several Illumina NovaSeq runs. We measured the Phred quality score and the demultiplexing errors per base. As a result, we detected a subset of barcodes that displayed higher error rate, usually related to low C content in the sequences, and being from T bases being read as G. These finding were confirmed by five additional runs, where the high error barcodes provided some of the lowest yields consistently. Additionally, Phred quality scores for i5 barcodes show a considerable drop on the first 2 bases, which is consistent with a higher mismatch rate. Correction on the quality performance of these barcodes would result on a more consistent yield across hundreds or thousands of samples, which would result on better downstream analysis. We showed that changing our mismatches from 1 to 0 resulted in a 10X reduction in index hopping.

Supplementation of the AgriSeq Canine SNP Parentage and ID Panel with additional ISAG and gender determination markers

Genotyping Tools and Techniques

POSTER 1003

Angela Burrell[1], **Krishna Reddy Gujjula**[1], **Haktan Suren**[1], **Rick Conrad**[1]

1 Thermo Fisher Scientific, Austin, TX

Parentage testing and genomics-assisted breeding are critical aspects of successful veterinary 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.

The AgriSeqTM Canine SNP Parentage and ID Panel, released in 2019, is an amplicon-based next-generation sequencing panel for parentage determination in dogs. In 2020 ISAG finalized the list of 233 recommended markers for canine parentage determination. The final recommendation contained five autosomal and three sex chromosome markers not present on the original AgriSeq panel. The AgriSeqTM Canine SNP Parentage and ID Panel was quickly updated to include not only the 8 missing markers, but 5 additional gender determination markers to ensure robust and repeatable gender determination results. The final panel contains 392 SNPs, including 4 markers targeting both the X and Y chromosome and 4 markers targeting the Y chromosome exclusively, and 2 deletions.

Utilizing the AgriSeqTM HTS Library Kit, a high-throughput targeted amplification and re-sequencing workflow, performance was validated on a panel of >100 samples. Libraries were sequenced on the Ion S5™ using an Ion 540™ chip with genotyping calling generated using the Torrent Variant Caller (TVC) plugin. Mean call rates, the percentage of markers on a panel generating a genotype call, was >98%. Concordance with orthogonal testing, including the Axiom Canine HD array and CE sequencing, was >99.4%. Gender determination accuracy was 100% for all samples tested and parentage determination was accurately assigned when testing the ISAG Comparison test samples.

The data demonstrates the utility of the AgriSeq targeted GBS approach for canine parentage and gender determination applications.

For Research Use Only. Not for use in diagnostic procedures.

Genomic DNA extraction from oil palm and coconut leaf samples for genotyping and identification with targeted GBS application

Genotyping Tools and Techniques

POSTER 1004

Kellsie Kice, **Claudio Carrasco***, **Rick Conrad**, **Srinivas Chadaram**

Thermo Fisher Scientific, Austin, TX

Current extraction methods used for obtaining genomic DNA from plant tissue do not produce DNA of high quality that can be routinely used in a high-throughput Genotype by Sequencing applications. The DNA yield is low and are too expensive for high throughput genotyping. Traditional methods for DNA extraction involve freezing and grinding of the sample tissue, which makes it difficult to use in a high throughput environment. Here we describe a new procedure to extract genomic DNA from leaf punches that is suitable for use in targeted genotype by sequencing applications. Our method involves 2-3 leaf punches (oil palm and coconut) that require a mechanical disruption by bead beating and then DNA extraction by processing the samples on a Kingfisher Flex with the MagMAX CORE Nucleic Acid Purification Kit and Plant RNA Isolation Aid. The new procedure yielded samples with a concentration range of 20-50 ng per µL of purified DNA per sample. Inhibition testing indicated that the extracted samples did not contain any measurable PCR inhibitors that can affect PCR amplification. The extracted DNA samples had high genotype call rates above 90% when genotyped with the AgriSeqTM Targeted Genotype by Sequencing workflow with custom designed SNP panels for Oil palm and Coconut. We conclude from the results that our method of DNA extraction yields genomic DNA that is cost-effective, suitable for genomic sequencing, and easily scalable to meet any High throughput requirement.

For Research Use Only. Not for use in diagnostic procedures.

AgriSeq™ for fungi surveillance in economically important crops

Genotyping Tools and Techniques

POSTER 1006

Srinivas Chadaram[1], **Krishna Reddy Gujjula**[1], **Stéphane Daly**[1], **Stefano Torriani**[2], **Lorenzo Borghi**[2], **Hakten Suren**[1]

1 Thermo Fisher Scientific, Austin, TX
2 Syngenta Crop Protection, Stein, Switzerland

Amplicon based targeted Genotyping-By-Sequencing (GBS) has enabled scientists to screen for known variants in a high throughput and cost-effective way. In this exploratory study, we extend the application of AgriSeqTM to fungi genetic mutation surveillance in economically important crops.

We designed an AgriSeqTM targeted sequencing panels for the detection and functional genotyping of three fungal pathogens - Zymoseptoria tritici, Pyrenophora teres, and Ramularia collo-cygni. These fungi affect economically important crops of Wheat and Barley and hence are of particular importance in agronomy. Utilizing AgriSeqTM HTS Library Kit, a high throughput targeted amplification and re-sequencing workflow, the panel's performance was tested on set of 83 diverse samples. The sample set included pure fungi strain, pure fungi strain with host, fungi strains diluted with water, DNA from symptomatic leaves and DNA from air probes. Libraries were sequenced on the Ion S5TM using an Ion 540TM chip with mutations/variants identified using the Torrent Variant Caller (TVC) plugin.

Results show that the panel is sensitive for samples with low concentration and is specific to the fungi target regions. The call rates for sample pure strain, pure strain with host, strains diluted with water, DNA from symptomatic leaves and DNA from air probes was 100%, 98.8%, 95.8%, 98.5% and 93.6% respectively and combined call rate of 97.8%. In addition to the known markers, novel mutations were also identified and reported.

The results demonstrate the capability of AgriSeqTM for variant identification in fungi.

For Research Use Only. Not for use in diagnostic procedures.

Paragon AgriType 2.0: optimization of a two-step multiplexed targeted NGS genotyping solution with high-throughput scalability

Genotyping Tools and Techniques

POSTER 1008

Jack Dong[1], **Julia Spencer**[1], **Lucie Lee**[1], **Michael Clark**[1], **Kalyani Patankar**[1], **Lily Liu**[1], **Julia Yoo**[1], **Jeffrey Liu**[1], **Yutao Fu**[1], **David Debruyne**[1]

1 Paragon Genomics, Inc., Hayward, CA

In order to meet our ever-growing global food demand, genome characterization proves crucial to optimize and render crop production systems more productive and sustainable, with Next Generation Sequencing becoming instrumental to advance agrigenomics research. Indeed, genotyping innovations drove further development of molecular breeding strategies, with key applications such as trait mapping, marker-assisted selection, or genetic engineering. But with existing challenges such as population expansion and climate change, high-throughput and cost-effective methods are required to maximize genotyping effectiveness in up to thousands of samples per day. Targeted solutions represent ideal candidates, by enabling sample multiplexing at an unprecedented scale while remaining highly informative. However, current workflows are still hindered by time-consuming steps and limited automation compatibility to fully empower mass-scale projects.

Here, we present a novel genotyping by sequencing solution, AgriType 2.0, constituting a simple and rapid library preparation workflow, with high-throughput capability (384-well plate) in mind. In summary, we adapted our targeted, multiplex PCR approach through the replacement of cleanup steps by a convenient intermediate dilution, as well as the implementation of a pooled final purification step, to not only improve turn-around time and ease of use, but also to facilitate automated processing. Experimental validation of our new method established its robust performance in several distinct crops, using panels targeting up to 7,000 markers, highlighting sequencing uniformity and call rate accuracy comparable to our CleanPlex technology. These results hereby confirm the successful streamlining of our NGS technology to effectively support large-scale agrigenomics studies aimed at crop improvement.

Genotyping of crops using optimized SNP arrays

Genotyping Tools and Techniques

POSTER 1009

Martin W. Ganal[1], **Joerg Plieske**[1], **Andreas Polley**[1], **Dagmar Kulosa**[1], **Thomas Gross**[1], **Heike Gnad**[1]

1 SGS Institut Fresenius GmbH TraitGenetics Section (SGS IF TG), Seeland OT Gatersleben, Germany

Genotyping for plant genetic studies, germplasm and variety characterization, and plant breeding requires a compromise between accuracy, resolution and costs per sample. In order to develop such genotyping arrays for major crop plants, we have assembled public SNP marker sets that (i) comprise of molecular markers that are of high technical quality, (ii) display a good level of polymorphism over a wide range of adapted and wild material for the respective crop based on hundreds of genotyped lines; (iii) are not in perfect linkage disequilibrium with each other; (iv) are evenly distributed along the chromosomes based on recombination; and are, if possible, linked to known genes for specific traits that are of breeding or scientific relevance. Especially, for plant breeding purpose such as Genomic Selection and other aspects of marker-assisted breeding, we have further reduced such genotyping arrays for small (a few thousand) marker sets that can be used in a cost-efficient way including imputation based on parental material genotyped with larger marker sets. Such optimized genotyping arrays for maize, soybean, wheat, barley, sunflower and Brassica are now being used widely within genetic research and accelerated plant breeding both for the academic and private sector.

Comparison of various kit-based DNA extraction methods to accurately and reliably detect genetically engineered soybean MON87705

Genotyping Tools and Techniques

POSTER 1010

Brenda Gee, **Katarina Gustin**

Zymo Research Corporation, Irvine, CA

There has been a growing interest in monitoring the presence of genetically modified organism (GMO) crops in countries such as the US, Canada, and other grain-exporting countries. This increase in awareness has prompted the European Union and other grain-importing countries to recognize the need to regulate the tolerance and traceability of various, different grains that are grown and imported into these countries. Legislation in the EU, specifically, requires labelling and handling when 0.9% of genetically modified traits are detected within a sample. The EU Commission's choice for method of detection and quantification is real time PCR, and for successful analyses, high quality DNA (good yield and high purity) is needed. In this study, we are evaluating various kit-based DNA purification methods on GMO certified soybean seed MON87705 and comparing them against the conventional CTAB method to determine the utility and practicality of these kit-based methods and the reliability when compared to the more commonly used CTAB method.

GBS solution for polyploidy: evolving next generation sequencing for production agriculture

Genotyping Tools and Techniques

POSTER 1011

Krishna Reddy Gujjula[1], **Haktan Suren**[1], **Chris Willis**[1], **Stéphane Daly**[2], **Maite Lelievre**[2], **Claudio Carrasco**[1]

1 Thermo Fisher Scientific, Austin, TX
2 Thermo Fisher Scientific, Lissieu, France

AgriSeqTM Targeted Genotyping by Sequencing is a valuable tool for high throughput, cost effective SNP detection and genotyping for Marker Assisted Selection or Breeding (MAS/MAB) in agriculture. AgriSeqTM is primarily being used for genotyping diploid organisms, leaving a segment of the plant genotyping, i.e., polyploids. Genotyping polyploids is more challenging compared to diploid organisms due to the presence of more than two sets of chromosomes and requires different analysis strategies.

We developed a statistical method to genotype polyploids using AgriSeqTM. To evaluate the performance of the method we designed two AgriSeqTM panels for auto-tetraploid potato and auto-hexaploid chrysanthemum. Utilizing the AgriSeqTM HTS Library Kit, each panel was used to generate sequencing data from 95 DNA samples on the Ion S5TM sequencing system.

The results show that the mean genotype call rate of markers across the samples for potato and chrysanthemum was 98.5% and 93% respectively. We evaluated the concordance for two cases: 1) Same sequencing data input; AgriSeqTM genotypes compared with fitpoly genotypes, 2) AgriSeqTM genotypes compared with genotypes from a microarray technology. For case 1, concordance was 99% and 96.5% and for case 2, concordance was 97.3% and 92.4% for potato and chrysanthemum respectively. In case 2, if we assume that the allele dosages differing by one are same then the concordance is > 99.5%.

The results demonstrate the performance of the genotyping method for polyploids to be used with AgriSeqTM.

For Research Use Only. Not for use in diagnostic procedures.

Analysis of various kit-based methods for DNA extraction from genetically engineered corn MON87460

Genotyping Tools and Techniques

POSTER 1012

Katarina Gustin, **Brenda Gee**

Zymo Research Corporation, Irvine, CA

The prevalence of genetically modified organisms in agriculture has prompted regulations regarding GMO content in seeds. Conventional extraction methods like CTAB are the standard for DNA extraction from genetically engineered seeds; however, these methods can be time consuming, inconsistent, and result in impure DNA. Seeds, including GE corn MON87460, tend to be tough-to-lyse and to extract high quality DNA for downstream quantitative PCR analysis. Kit based DNA extraction method is an attractive alternative over traditional organic based extraction methods due to their speed, ease of use and the practicality to detect low level of genetically engineered events in grain samples. Here we evaluated various kit-based methods to compare the DNA extraction yield, purity, limit of detection and their practicality over the CTAB methods. In this study we are using MON87460 corn, utilizing the limit of detection parameters that have been established by the European Grain Commission's guidelines for GE threshold of 0.09% MON87460 event per sample. Results are compared by their ability to detect 0.9% MON87460 with specificity and minimal inhibition.

Dual-barcode feasibility in the AgriSeq targeted genotyping-by-sequencing workflow

Genotyping Tools and Techniques

POSTER 1013

Jennifer Lavoie[1], **Angela Burrell**[2]

1 Thermo Fisher Scientific, Austin, TX
2 Thermo Fisher Scientific, Waltham, MA

Targeted Genotyping by Sequencing (GBS) is a robust and cost-effective method for marker-assisted breeding and selection. Targeted GBS provides a scalable workflow, allowing for thousands of markers to be analyzed in up to thousands of samples per day.

The Applied Biosystems™ AgriSeq™ targeted GBS solution allows for up to 1536 uniquely barcoded samples to be processed in a single sequencing run. The current workflow utilizes a single-barcode approach, with 1536 barcodes available in 16 96-well plates. With amplicon coverage of 100X and using an Ion 540™ chip averaging 70M reads, up to 500 markers can be analyzed in 1536 samples on a single chip. Using an Ion 550™ chip averaging 115M reads, up to 750 markers can be analyzed in 1536 samples on a single chip. As GBS use expands, demand increases for higher throughput, faster turnaround, and increased multiplexing capability in GBS workflows.

Increasing available barcodes could increase the sample throughput of the AgriSeq™ workflow. In the current single barcode workflow, increasing available barcodes means increasing the number of 96-well plates to manage, which is undesirable. As an alternate solution for increasing throughput and barcode availability, we are exploring the use of dual barcodes. With this method, each sample would receive a combination of barcodes, allowing for more samples to be uniquely barcoded, which increases multiplexing capability and cost effectiveness without adding 96-well barcode plates. Here we report the feasibility of implementing the use of dual barcodes in the AgriSeq™ workflow and its implications for marker performance and coverage requirements.

Advances in DNA microarray and next generation sequencing tools for agrigenomics

Genotyping Tools and Techniques

POSTER 1014

Kahlil Lawless[1]

1 Illumina Inc., Foster City, CA

Next Generation Sequencing (NGS) has been the backbone of Agrigenomics research for more than a decade, and high-throughput microarray platforms have enabled genotyping at a cost and scale required to apply this information to breeding programs. Illumina continues to update and improve both technology platforms through R&D investments, manufacturing and supply improvements, and deployment new products and features that reduce costs and increase ease of use, opening the door to new discoveries and applied use cases into new areas such as environmental biomonitoring, innovative crop and soil treatments, veterinary health, conservation, and food chain management.

Kahlil Lawless, Global Agrigenomics Segment Manager at Illumina Inc, will share recent advances and improvements, as well as examples from the field where these technologies are being utilized successfully in research programs and applied testing laboratories.

Beef: Is it really what's for dinner? Onsite meat species determination assay

Genotyping Tools and Techniques

POSTER 1015

Seth Lewin, **Colin Brewer, Heather Piscatelli, Abe Oommen**

MatMaCorp, Lincoln, NE

Consumers purchasing processed meat must rely on labeling to make dietary decisions. Intentional mislabeling or unintentional contamination of meat products during processing can pose a problem for consumers with specific dietary concerns such as people adhering to religion-specified diets, those with specific allergies, or other medically specified diets. A label alone cannot guarantee consumer protection unless a certification process documenting a clean supply chain is in place, as seen with Halal and Kosher designations. Therefore, the need exists for a rapid, onsite DNA "barcoding" test to identify the presence of unwanted species so that producers may certify a premium product.

To demonstrate the use of such a test in practice, MatMaCorp's Meat Identification C-SAND (Combined Sequence Amplification and Nucleotide Detection) Assay was used to identify beef samples processed at the University of Nebraska's animal science facility that might contain pork. Briefly, a batch of pure beef was processed on clean equipment followed batches by beef contaminated with an increasing amount of pork. A final batch of beef was processed on the same equipment without cleaning to pick up any residual pork remaining on the equipment. A sample from each batch was collected for DNA isolation using MatMaCorp's StickE Column Tissue DNA Isolation kit, and MatMaCorp's MagicTip Tissue DNA Isolation kit. DNA eluted from each was used as template in PCR and C-SAND Assays to identify which samples contained pork. The ease and portability of MatMaCorp's products allows the user to conduct the assay onsite, within 2 hours. Producers now can monitor product purity at near real-time and can confidently label the product as such.

A high-throughput Applied Biosystems™ Axiom™ Bovine Genotyping array with 100,000 markers optimized for dairy evaluation

Genotyping Tools and Techniques

POSTER 1016

Ali Pirani[1], **Dorothy Oliver**[1], **Christofer Bertani**[1], **Mohini Patil**[1]

1 Thermo Fisher Scientific, Inc., Santa Clara, CA

While the world population increases at an unprecedented rate, meeting the growing food needs continues to be a challenge. For more than a decade, the bovine dairy industry has employed the genetics of their cattle to improve production traits, such as milk yield and protein percentage. These methods have shown to be critical for the improvement of dairy cattle productivity. This breeding strategy is achieved by genotyping thousands of biallelic SNPs, interrogating loci well-distributed across the entire genome, potentially capturing all relevant quantitative trait loci (QTL). These genotypes are converted to genomic estimated breeding values (GEBVs) using a method called Genomic Selection (GS). The application requires the interrogation of a fixed set of markers rapidly over thousands of samples, so medium-density, 25,000 to 100,000 marker microarrays are an ideal fit.

For genotyping dairy cattle, Thermo Fisher Scientific provides numerous Applied Biosystems Axiom microarrays measuring around 65,000 markers. These arrays, such as the Axiom Bovine Genotyping v3 Array includes 44,000 markers recognized by the Council on Dairy Cattle Breeding (CDCB). Recently, the CDCB released a list of 80,000 markers used for genetic evaluation.

Thermo Fisher Scientific has developed a 100,000-marker microarray to interrogate all 80,000 CDCB relevant makers. In addition, this array includes markers for even genomic coverage, economically valuable traits-associated markers, sex-linked markers, microsatellite imputation markers, and parentage verification, such as the International Society for Animal Genetics (ISAG) 200 and ICAR 354 markers. This higher-density panel can also be useful in tracking undesirable genetic trends, such as inbreeding depression, to drive overall genetic improvement of dairy cattle in commercial breeding programs.

Optimization of automated NGS workflows for agrigenomics

Genotyping Tools and Techniques

POSTER 1017

Matthew Stull[1], **Marcel Brun**[1], **Joshua Hill**[1], **Emma Schulze**[1], **Richard P. Metz**[1], **Charles D. Johnson**[1]

1 Texas A&M AgriLife Research Genomics and Bioinformatics Service, College Station, TX

Agrigenomics is a growing field of research focused on feeding the world population by studying the genomes of crops and livestock. Next Generation Sequencing (NGS) is a key tool in agricultural focused genotyping, delivering high quality genomic information at ever-decreasing costs and increased throughput, through a combination of improved chemistry, hardware, and software. As sequencing prices decrease, more studies focus on hundreds or thousands of samples, and a larger proportion of the cost and effort moves from sequencing to library preparation. To decrease the cost and time for large NGS studies we must library preparation costs without compromising quality or consistency. The rise of low-coverage genotyping is an example of this.

Robotic liquid handlers can improve speed, accuracy and precision when compared to manual library preparation, but they also can enable miniaturization of volumes used for the process. 96-wells robots use similar plates as the ones used by manual pipetting, but for 384-well robots the change to 384-well plates results in reduction in reagents used per well. As volume decreases, any inaccuracy, either due to insufficient dead volumes, inherent precision, or evaporation, can lead to large effects on the resulting libraries.

In this work we describe our implementation of miniaturization for DNASeq, RNASeq, and 16S protocols. Our DNA-seq and 16S protocols are currently optimized to 1/10 reaction ratios in 384 well plates, by calibrating reaction ratios and diluting barcodes. As a result of our changes in reagents and elution volumes, and careful calibration, we obtained libraries with consistent yields and size, and high success rate, above 99% for high quality starting material. RNA-Seq protocols are being transitioned as we increase scale from 2/5 reaction volume on the 96 well Sciclone to a 1/5 384 well Sciclone.

In summary, with the adapted protocols optimized for the 384-well workstation, large agrigenomics projects with hundreds or thousands of samples can be prepped in less time, with higher precision and reduced costs.

Reference Genome Assembly Techniques and Initiatives

Living fossils: the Wollemi pine genome and insights into genomic and epigenomic evolution

Reference Genome Assembly Techniques and Initiatives

POSTER **800**

Cristiane de Santis Alves[1,2] **Olivia Mendivil Ramos**[1], **Melissa Kramer**[1], **Srividya Ramakrishnan**[3], **Veronica Sondervan**[4], **Gil Eshel**[4], **Kranthi Varala**[6], **Sergios-Orestis Kolokotronis**[7], **Samantha Frangos**[4], **Cecilia Zumajo-Cardona**[5], **Damon P. Little**[5], **Manpreet Katari**[4], **Mark Richardson**[9], **Larry Croft**[8], **Hannah McPherson**[10], **Maurizio Rossetto**[10], **Sara Goodwin**[1], **Barbara Ambrose**[5], **Dennis Stevenson**[5], **Gloria Coruzzi**[4], **Michael C. Schatz**[3], **Richard McCombie**[1], **Rob Martienssen**[1,2]

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2 Howard Hughes Medical Institute, Chevy Chase, MD
3 Department of Computer Science, Johns Hopkins University, Baltimore, MD
4 Center for Genomics and Systems Biology, New York University, New York, NY
5 New York Botanical Garden, Bronx, NY
6 Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN
7 School of Public Health, SUNY Downstate Health Sciences University, Brooklyn, NY
8 Deakin Genomics Research and Discovery Facility, Deakin University, Victoria, Australia
9 School of Life and Environmental Sciences, Deakin University, Victoria, Australia
10 Australian Institute of Botanical Science, Royal Botanic Gardens and Domain Trust, Sydney, Australia

Gymnosperms are important evolutionary forerunners to flowering plants, due to the emergence of pollen and seeds. Wollemia nobilis, a member of the Araucariaceae family, is a critically endangered gymnosperm that appeared in the Cretaceous period and was thought to be extinct but has remained morphologically unchanged until today. With the advent of long read sequencing to improve the assembly of large genomes with high repetitive content like the gymnosperms, The New York Plant Genome Consortium has generated a reference-quality genome of the living fossil Wollemia nobilis. The optimized de novo assembly pipeline we developed has produced a highly contiguous and accurate genome sequence of this large genome (~12Gb, N50 ~9Mb). We used RNA-seq and small RNA-seq from different tissues to comprehensively annotate ~48,000 genes, ~62,000 transposable elements, and ~500,000 small RNAs. We found that 75% of the sRNA content is composed of 24 nucleotide siRNAs, and at least 277 microRNAs, of which 38 are conserved across other plant species. We also identified nine putative DICER-LIKE and 23 putative ARGONAUTE genes, half of which fall into the AGO4,6,8,9 family and are preferentially expressed in ovules. To help address the evolution of epigenetic and RNAi pathways, we leveraged the raw signal from the Nanopore long-read data to profile the base modifications in the leaf genome using multiple methods including Nanopolish, Megalodon, DeepSignal-Plant. We have generated whole-genome bisulfite data from ovules and pollen and found similar levels of methylation in both tissues, around 78%, 65%, and 2.6% in CG, CHG, and CHH contexts respectively. Most of the methylation occurs in regions annotated as transposable elements. Prominently, we found that some LTR elements from the Copia but mainly Gypsy superfamilies are being processed as siRNAs. This suggests that DNA methylation could be part of a similar RNA-dependent DNA Methylation pathway to help keep transposable elements silenced. More analyses are taking place to aid in unraveling regulatory pathways, our efforts to integrate the data will allow a better understanding of the genomic resilience to environmental changes over geological timescales.

Chromosome-scale genome assembly of the medicinal plant Teucrium marum (Lamiaceae)

Reference Genome Assembly Techniques and Initiatives

POSTER **801**

Sefa Ayten[1], **John Hamilton**[2], **Brieanne Vaillancourt**[2], **C. Robin Buell**[1,2]

1 The University of Georgia, Institute of Plant Breeding, Genetics, and Genomics, Athens, GA
2 The University of Georgia, Center for Applied Genetic Technologies, Athens, GA

A wide number of specialized metabolites produced by medicinal plants are valuable sources for treatment of health disorders. Improvements in genome sequencing technologies and reduced sequencing costs have resulted in a substantial increase in the number of published genomes in recent years, result in accelerating the discovery of new involved genes in plant natural product biosynthesis and better understanding of evolutionary developments of specialized metabolism. As one of the largest genera of Lamiaceae (mint family), Teucrium species, commonly known as germanders, are renowned for their medicinal properties and aromatic compounds. Teucrium marum is used in the folk medicine for its antibacterial, antimicrobial, anti-inflammatory, antioxidant, and antipyretic activities. It also has euphoria effects on cats and produces specialized secondary metabolites, such as the plant defensive agents dolichodial, teucrein, and dolicholactone. Here, we present a high quality de novo genome assembly of T. marum, the first assembly of a Teucrium genus, using Oxford Nanopore Technologies long reads and high-throughput chromosome conformation capture reads to construct chromosome-scale pseudomolecules, and annotation of genome assembly. T. marum is a highly heterozygous diploid and we generated a 610.3Mb assembly, spanning 15 pseudomolecules with a 32.9Mb N50 scaffold size. Genome assembly and annotation of this species will provide a resource for elucidation of some candidate genes and deep understanding of some metabolic pathways which is responsible for biosynthesis of specialized metabolites, such as the insect repellent dolichodial.

Use of TELL-Seq technology for de novo assembly of microbial genomes

Reference Genome Assembly Techniques and Initiatives

POSTER **802**

Robin Bombardi[1], **Ana Granat**[1], **Tiffany Truong**[1], **Erich Jaeger**[1], **Yong Wang**[2], **Tom Chen**[3], **Feng Chen**[1]

1 Illumina, Inc, Foster City, CA
2 Universal Sequencing Technology, Canton, MA
3 Universal Sequencing Technology, Carlsbad, CA

BACKGROUND: Microbial genomes exhibit varying sequence composition, often containing repetitive or low-complexity regions that can complicate de novo genome assembly. Short reads characterize the genomes efficiently when reference genomes are available, but when sequencing genomes for the first time some regions can be difficult, leading to gaps in otherwise contiguous assemblies. Using clonally barcoded beads, TELL-Seq™ technology allows reads originating from the same long parent template molecule to be associated informatically, thereby bridging intervening gaps in sequence. Using TELL-Seq™ technology to obtain microbial de novo genetic information provides a fresh start for pathogen surveillance in agriculture application areas.

METHODS: Eight microbial species with genomic GC content ranging from 28% to 69% were chosen for TELL-Seq™ library preparation. Genomic DNA was extracted using the MagAttract kit (Qiagen) according to the manufacturer's instructions, yielding gDNA fragments with an average length of ~30kb. 0.5ng of gDNA input was used for each prep and bead input into PCR was proportional to genome size, which ranged from 1.6 to 5.4 Mb. All eight libraries were sequenced in a single MiSeq run and reads were downsampled to achieve a mean coverage of 100X for each genome. Sequence data was processed and assembled with the TELL-Read and TELL-Link software (Universal Sequencing Technology), respectively.

RESULTS: All library products showed the expected fragment size distribution regardless of GC content. A total of 1-3M read pairs was needed to achieve 100X coverage of each microbial genome. TELL-Read analysis indicated input parent template molecule lengths of 15-28kb. >96% of each genome was captured and TELL-Link assembly gave very large contigs and NG50 values that approached the size of the largest chromosome present in each genome, and misassembly was limited in most organisms.

CONCLUSIONS: TELL-Seq™ technology enables high-quality assembly of microbial genomes from short read sequences using sub-nanogram inputs of genomic DNA. Optimizing gDNA extractions to increase input fragment lengths and removal of fragments <20kb is expected to further improve performance. Inaccurate input quantitation can contribute to decreased performance, as was observed for B. cereus, and highly repetitive genomes such as that of B. pertussis may be more prone to misassembly.

Cutting-edge genomics and engineering insects for sustainable food, feed, and pharma

Reference Genome Assembly Techniques and Initiatives

POSTER **803**

Aaron T. Dossey[1,2], **Clay Chu**[1,2]

1 All Things Bugs LLC, Midwest City, OK
2 Invertebrate Studies Institute (non-profit 501c3), Oklahoma City, OK

Farm raised insects offer a substantial, let nearly entirely overlooked, solutions for a more sustainable agriculture system and healthier planet. High quality genomes and genetic engineering technologies can expand their value immensely. Recent horrific reports from IPCC, IPBES, UN and others on climate change and biodiversity loss, along with increasing human population, demonstrate the urgent critical need to reduce human consumption from earth and its ecosphere. Already 70% of agricultural land and 30% of the land on earth is used for livestock. Diversification of our food supply is critical for food security, and insects are a very promising solution. They utilize less energy, feed, land and water than other livestock and contribute less to climate change and pollution. While insects, such as crickets, mealworms, black soldier flies, and others are naturally highly efficient and nutritious, high quality genome assemblies and genetic engineering via CRISPR and other techniques can even further expand their utility for clean, sustainable food, pharma and other valuable applications. Our company, in collaboration with USDA ARS, with funding from DARPA, is pioneering this new field. We have generated high quality chromosomally scaffolded annotated reference genomes of three of the world's most farmed insect species: House cricket (Acheta domesticus), Banded Cricket (Gryllodes sigillatus) and Yellow Mealworm (Tenebrio molitor). The cricket genomes will be the first for those species. Additionally, we have generated genetic engineering tools utilizing CRISPR/Cas9 toward improving insects as food crops, and for non-food applications including vaccines, bioactive peptides, and antimicrobials. Discoveries from cricket and mealworm genomes will be discussed as well as genetic engineering work developing a white-eye knock-out, EGFP knock-in biomarker system. We are also developing proprietary strains expressing genes for Vitamin A biosynthesis, higher quality more digestible proteins, lower food allergen risk, and resistance to a major densovirus pathogen at cricket farms. Overall, this presentation will illustrate our genomic and genetic engineering work and provide an overview of our company's broader mission of developing insects as an industrial food and bioresource.

Establishing two Physalis species as new Solanaceae systems enables genetic reevaluation of the inflated calyx syndrome

Reference Genome Assembly Techniques and Initiatives

POSTER **804**

Jia He[1,2], **Michael Alonge**[3], **Srividya Ramakrishnan**[3], **Matthias Benoit**[1,2], **Sebastian Soyk**[1],†, **Nathan T. Reem**[4], **Anat Hendelman**[1], **Joyce Van Eck**[4], **Michael C. Schatz**[1,3,5], **Zachary B. Lippman**[1,2]

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2 Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
3 Department of Computer Science, Johns Hopkins University, Baltimore, MD
4 Boyce Thompson Institute, Ithaca, NY
5 Department of Biology, Johns Hopkins University, Baltimore, MD

The Solanaceae family is home to many orphan crops and wild relatives, which, with high quality genomes and efficient genome editing, offer opportunities to address a wide range of species-specific and comparative biological questions. Particularly promising are species in the genus Physalis, which include diploid orphan berry crops that are rapid cycling and easy to grow. We previously established genome editing in P. grisea (groundcherry) and improved productivity traits by targeting orthologs of tomato domestication genes. However, P. grisea and its relatives are most studied for the Inflated Calyx Syndrome (ICS), a spectacular morphological novelty where sepals grow exceptionally large to encapsulate fruits in a papery husk. To advance Physalis as a new Solanaceae system and begin a careful dissection of ICS, we established chromosomal-scale genomes for P. grisea and its close relative P. pruinosa. In addition to facilitating rapid dissection of color variation in floral nectar guides, these genomes provided the foundation to evaluate contributions from a suite of candidate MADS-box genes in ICS. Surprisingly, targeted mutagenesis of 12 MADS-box genes, including two previously supported regulators, had no effect on ICS. To identify new regulators, we used forward genetics and found huskless (hu), a mutation in an AP2-like transcription factor that eliminates the inflated calyx by merging sepals and petals into a single organ. Our resources and findings cement Physalis as a powerful new genetic system, and force a reset in the search for mechanisms underlying the evolution of ICS.

Impact of nanopore sequencing innovations on comprehensive genomic and genetic understanding of crops

Reference Genome Assembly Techniques and Initiatives

POSTER **805**

René Hogers[1,*], **Alexander H.J. Wittenberg**[1], **Ilse VrijenHoek**[1], **Erwin Datema**[1], **Esther Verstege**[1], **Sevgin Demirci**[1], **Harrie Schneiders**[1], **Raymond Hulzink**[1], **Antoine Janssen**[1], **Wilco Ligterink**[1], **Nathalie van Orsouw**[1], **Dick Roelofs**[1]

1 Keygene NV, Wageningen, Gelderland, The Netherlands

KeyGene® is the go-to plant research company for technological innovation for crop improvement. The availability of high-quality, contiguous reference genomes is essential for effective marker development and gene discovery, associated with important crop breeding traits. Therefore, technology development within KeyGene allows its shareholders and customers to significantly enhance crop improvements. The early adoption of long read sequencing technologies (e.g. Oxford Nanopore Technologies) by KeyGene has become an essential driver of innovation. For instance, improvements on the Oxford Nanopore Technologies platform and the development of proprietary genome assembly software enabled KeyGene to generate genome sequences of unprecedented quality and contiguity for a variety of crops. Still, whole genome sequencing of large populations is not cost effective, especially when studying large and complex genomes. Therefore, KeyGene has developed innovations to reduce genome complexity in a random or targeted manner. Because these applications are long read sequencing-based, they can comprehensively detect all variation, from single nucleotide polymorphisms to structural variation (including copy number variation) in the context of genetic linkage. This opens the possibility to perform more precise haplotype-based molecular breeding contributing to crop improvement. Moreover, variation in DNA modifications, proven to be associated with important agricultural traits such as fruit ripening, can be instantly revealed using Oxford Nanopore Technologies sequencing. To reach the highest quality possible on this platform, we include latest Q20+ sequencing chemistries, flow cells with R10.4 pores and plant optimized basecallers. An overview of the latest results and ongoing developments will be presented. This will include results of plant DNA optimized basecall training, advances in genome assembly software as well as technology to enable haplotype-based genotyping and DNA modification detection.

The highly rearranged reference genome of hexaploid oat cv 'Sang' provides insights into the unique health properties of oats

Reference Genome Assembly Techniques and Initiatives

POSTER **806**

Nadia Kamal[1], **Nikos Tsardakas Renhuldt**[2], **Johan Bentzer**[2], **Heidrun Gundlach**[1], **Georg Haberer**[1], **Angela Juhasz**[12], **Thomas Lux**[1], **Utpal Bose**[12,13], **Jason Tye-Din**[14,15], **Daniel Lang**[1], **Nico van Gessel9**, **Ralf Reski**[9], **Yong-Bi Fu**[17], **Peter Spégel**[4], **Alf Ceplitis**[6], **Axel Himmelbach**[8], **Amanda J. Waters**[20], **Wubishet A. Bekele**[16], **Michelle L. Colgrave**[12], **Mats Hansson**[3], **Nils Stein**[8,10], **Klaus Mayer F.X.**[1,19], **Eric N. Jellen**[18], **Peter J. Maughan**[18], **Nicholas A. Tinker**[16], **Martin Mascher**[8,11], **Olof Olsson**[2,5], **Manuel Spannagl**[1], **Nick Sirijovski**[2,5,7]

1 Helmholtz Zentrum München, German Research Center for Environmental Health, Plant Genome and Systems Biology, Neuherberg, Germany.
2 Lund University, ScanOats, Department of Chemistry, Pure and Applied Biochemistry, Lund, Sweden
3 Lund University, Department of Biology, Molecular Cell Biology, Lund, Sweden
4 Lund University, Department of Chemistry, Centre for Analysis and Synthesis, Lund, Sweden
5 CropTailor AB, c/o Lund University, Department of Chemistry, Division of Pure and Applied Biochemistry, Lund, Sweden
6 Lantmännen, Plant Breeding, Svalöv, Sweden
7 Oatly AB, Food Science, Lund, Sweden.
8 Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Domestication Genomics, Seeland, Germany
9 University of Freiburg Plant Biotechnology, Faculty of Biology, Freiburg, Germany
10 Center of integrated Breeding Research (CiBreed), Department of Crop Sciences, Georg-August-University, Göttingen, Germany
11 German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany
12 Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Edith Cowan University, School of Science, Joondalup WA, Australia
13 Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, QLD, Australia
14 Walter and Eliza Hall Institute of Medical Research, Immunology Division, Parkville, VIC, Australia
15 The Royal Melbourne Hospital, Gastroenterology, Parkville, VIC, Australia
16 Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, Canada
17 Plant Gene Resources of Canada, Agriculture and Agri-Food Canada, Saskatoon, Canada
18 Brigham Young University, Plant & Wildlife Sciences Department, Utah, USA
19 School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
20 PepsiCo R&D, St. Paul, MN, USA

Helmholtz Center Munich, Neuherberg, Germany

Oat is becoming an increasingly important global crop due to its unique health properties and low carbon footprint. We present a fully annotated high-quality reference genome assembly of allohexaploid common oat (Avena sativa), cultivar "Sang" and provide an in-depth investigation of the genomic factors underlying the unparalleled qualities of oats in human health and nutrition. Compared to other cereals, the oat prolamin gene families exhibit qualitative and quantitative differences in the genomic and proteomic basis for the expression of peptides that can trigger celiac disease, supporting the safety of oats in a gluten-free diet. Decoding the large cellulose synthase gene superfamily of oat provides a genetic foundation underlying the biosynthesis of oat beta-glucans and the health claims associated with blood cholesterol and post-meal glycemic responses.

By comparative structural analysis of hexaploid, tetraploid and diploid Avena species, we identified large-scale chromosome rearrangements in the oat genome and traced them back in the polyploidization history. This provided evidence of hitherto hidden breeding barriers in oat and offers an explanation for previous difficulties experienced in introgression breeding in oat. We analysed gene and repeat content as well as expression abundance across the three subgenomes of hexaploid oat and tested observed differences in relation to the structural rearrangements. We also demonstrate the usefulness of the reference genome by the single-gene-mapping of an agronomic trait involved in epicuticular wax production.

The newly established reference genome for hexaploid oat, together with its extensive annotations and genomic resources, provides the basis for a better understanding of basic oat biology. It will facilitate more targeted breeding of oats and sets the foundation for on-going oat pan-genome projects.

Lessons learned while developing genomic resources and tools for a hemipteran pest of maize

Reference Genome Assembly Techniques and Initiatives

POSTER **807**

Marcé Lorenzen[1], **William Klobasa**[1], **Dorith Rotenberg**[1], **Anna Whitfield**[1]

1 North Carolina State University, Raleigh, NC

The ability to utilize state-of-the-art genomic tools can open the door to a wide range of new pest control solutions, including the use of gene drive-based technologies. However, there may be considerable barriers to deploying these tools in a new species. For example, despite the power of CRISPR/Cas9-based genome editing, published reports of confirmed genome editing in hemipteran species are rare. Here we will describe the addition of a new hemipteran species to the list, the corn planthopper, Peregrinus maidis, and offer tips and tricks we wish we had known at the outset of the project. Like many agricultural pests, P. maidis lacked a sequenced genome, therefore the first step – and roadblock – was in the development of genomic resources. While generating a genome assembly for an insect with an estimated 750-Mb genome may sound like a straightforward task, we quickly discovered the task was far more challenging than anticipated. We will describe the steps that ultimately enabled the production of high-quality, nearly endosymbiont-free HiFi reads for this species, as well as production of Hi-C data to support chromosome-level genome assembly. We will also describe how we used such genomic and transcriptomic resources to reach our initial goal, that of developing methods for genetically modifying this important agricultural pest. Importantly, these are merely the first steps towards a much boarder goal: bringing game-changing genomic tools to bear on understanding and controlling important agricultural pests.

Rootless hair: an underappreciated, noninvasive sample type for genomic analyses in animal husbandry contexts

Reference Genome Assembly Techniques and Initiatives

POSTER **809**

Varsha Rao[1], **Colin Naughton**[1], **Chris Troll**[1], **Kelly Harkins**[1], **Beth Shapiro**[2], **Richard Green**[1,3]

1 Claret Bioscience LLC, Santa Cruz, CA
2 Ecology & Evolutionary Biology Department, University of California, Santa Cruz, Santa Cruz, CA
3 Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA

Next-generation sequencing (NGS) technology has revolutionized the way we explore the living world. However, in many scenarios, including in animal husbandry contexts, it may be difficult to obtain biological samples for routine genetic analyses. The use of rootless hair has only recently emerged in forensic genetics as a viable sample type for human genome-level analyses, including genotyping, because a single strand contains only picograms of highly fragmented DNA (< 50 bp). These characteristics have historically hindered conventional PCR-based testing as well as standard NGS library preparation methods.

Here we demonstrate the utility of a single-stranded NGS library preparation approach developed by Claret Bioscience that enables conversion of ultra-short DNA for whole genome sequencing. Using ancient DNA extraction methods coupled with the SRSLY PicoPlus NGS library preparation kit, we generated sequence-ready libraries in under 3 hours from as little as 100 pg of DNA from rootless hair/feathers. The samples were collected directly (e.g. cut or sheared) or indirectly (e.g. shed) from a variety of living animals including horse, sheep, chicken, alpaca, gibbon, dog, cat, as well as extinct animals like the wooly rhino and mammoth.

By mapping the sequencing reads to the respective reference genomes we show that SRSLY generated high quality DNA data, even from 50,000+ year old samples. The fragment length distribution of all samples showed the same mean length and ~10bp periodicity as observed in modern human hair samples. Even with low coverage (libraries sequenced to <4M reads), data from rootless hair was used to reconstruct complete mitochondrial genomes at high coverage, which can be used in evolutionary analysis, in determination of mtDNA derived phenotypical features, and more. Molecular complexity estimates show that each tested hair strand contained tens or hundreds of millions of unique DNA fragments, enabling a wide range of whole genome analyses, like SNP profiling. This novel sample type and genomic approach opens new possibilities for non-invasive wildlife sampling, livestock breed evaluation and conservation, ecological analyses, and other agricultural applications.

Automated telomere-to-telomere genome assembly with HiFi and Nanopore reads

Reference Genome Assembly Techniques and Initiatives

POSTER **810**

Mikko Rautiainen[1], **Sergey Nurk**[1], **Brian Walenz**[1], **Sergey Koren**[1], **Adam Phillippy**[1]

1 Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Recent developments in sequencing technology have greatly increased the accuracy and contiguity of genome assemblies. New PacBio HiFi and Oxford Nanopore (ONT) based sequencing reads have been shown to be capable of assembling all chromosomes of a human genome from end to end, as proven by the recent Telomere-to-Telomere project which used plentiful amounts of manual effort from experts around the world to fully assemble the first complete human genome. Here we present a hybrid assembly pipeline called Verkko, which uses a combination of HiFi and ONT reads to provide accurate, highly contiguous assemblies in an automated manner. On the pseudohaploid CHM13 cell line, Verkko fully and correctly assembles most chromosomes from telomere to telomere. On diploid human genomes Verkko provides highly contiguous phased assemblies which can further be combined with trio, Hi-C or StrandSeq data to output chromosome scale phased haplotypes. The assemblies created by Verkko are highly accurate with error rates on the order of one per million base pairs. As far as we are aware Verkko is the first assembler to provide automated telomere to telomere assemblies of entire human chromosomes.

Phased, chromosome-scale genome assemblies of tetraploid potato reveals a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity

Reference Genome Assembly Techniques and Initiatives

POSTER **811**

Genevieve Hoopes[1], **Noelle L. Anglin**[2], **Christian W. B. Bachem**[3], **Jeffrey B. Endelman**[4], **Laura M. Shannon**[5], **Martina V. Strömvik**[6], **Helen H. Tai**[7], **Björn Usadel**[8], **C. Robin Buell**[9], **Richard Finkers**[10], **Potato Pangenome Consortium**

1 J.R. Simplot Company, Boise ID
2 United States Department of Agriculture (USDA) Agricultural Research Service (ARS), Aberdeen, ID
3 Wageningen University & Research, Wageningen, The Netherlands
4 University of Wisconsin-Madison, Madison, WI
5 University of Minnesota, St. Paul, MN
6 McGill University, Sainte-Anne-de-Bellevue, Quebec, Canada
7 Agriculture and Agri-Food Canada, Fredericton Research and Development Centre, Fredericton, New Brunswick, Canada
8 Forschungszentrum Jülich, Jülich, Germany
9 University of Georgia, Athens, GA
10 GenNovation, Wageningen, The Netherlands

Cultivated potato is a clonally propagated autotetraploid species with a highly heterogeneous genome. Phased assemblies of six cultivars, including two chromosome-scale phased genome assemblies, revealed extensive allelic diversity, which included altered coding and transcript sequences, preferential allele expression, and structural variation, collectively resulting in a highly complex transcriptome and predicted proteome distributed across the homologous chromosomes. Wild species contribute to the extensive allelic diversity in tetraploid cultivars, demonstrating ancestral introgressions predating modern breeding efforts. As a clonally propagated autotetraploid that undergoes limited meiosis, dysfunctional and deleterious alleles are not purged in tetraploid potato. Nearly a quarter of the loci bore mutations predicted to have a high negative impact on protein function, complicating breeder's efforts to reduce genetic load. The StCDF1 locus controls maturity and analysis of six tetraploid genomes revealed 12 allelic variants correlated with maturity in a dosage dependent manner. Knowledge of the complexity of the tetraploid potato genome, with its rampant structural variation and embedded deleterious and dysfunctional alleles, will be key not only to implementing precision breeding of tetraploid cultivars, but also to the construction of homozygous, diploid potato germplasm containing favorable alleles to capitalize on heterosis in F1 hybrids.

Reference Genome Assembly Techniques and Initiatives

POSTER 812

HetTrek: improving assembly accuracy and contiguity of highly heterozygous genomes

Katharine M. Jenike[1,*], **T. Rhyker Ranallo-Benavidez**[2*], **Srividya Ramakrishnan**[3], **Michael Alonge**[3], **Sam Kovaka**[3], **Matthias Benoit**[4,5], **Zachary Lippman**[4,5], **Michael C. Schatz**[1,3,4]

1 Johns Hopkins School of Medicine, Genetic Medicine, Baltimore, MD
2 University of North Carolina, Bioinformatics and Genomics, Charlotte, NC
3 Johns Hopkins University, Computer Science, Baltimore, MD
4 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
5 Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

De novo assembly is a challenging problem due to sequencing errors, genomic repeats, and heterozygous variants. In highly heterozygous genomes, portions of haplotypes may be constructed separately in the assembly resulting in false regional duplications that interfere with downstream analysis. While a number of methods exist to overcome these challenges, such as sequencing inbred model organisms, performing gamete binning, trio analysis, or utilizing Hi-C data, these methods are resource-intensive and not always possible. Producing high quality assemblies of non-model organisms without trio data, gamete data, or Hi-C data is thus extremely challenging with existing methods.

Here we present HetTrek, a method that removes heterozygosity and corrects sequencing errors from single-molecule sequencing reads, even in repetitive plant genomes. Reads are edited based on the k-mer counts in overlapping reads. The output is a set of enhanced reads with greatly reduced error and heterozygosity that can be used by existing assembly pipelines for a highly contiguous and accurate de novo assembly. Importantly, HetTrek outputs the original and corrected heterozygous SNPs and indels which can be used for future analysis.

We employ HetTrek on a PacBio HiFi dataset of a highly heterozygous, clonally propagated, orphan crop Solanum muricatum (pepino). This reduced both the error and heterozygosity rates and ultimately produced a more contiguous and accurate assembly compared to a pseudo-haploid assembly with the raw reads. The improved assembly could enable genome editing and other functional assays in pepino. Furthermore, we demonstrate the utility of the HetTrek polishing feature on genome assemblies of both Wollemia and Gnetum sequenced with Oxford Nanopore long reads and Illumina short reads.

*Equal contributions

Reference Genome Assembly Techniques and Initiatives

POSTER 813

So you think you finished a genome?

Sergey Koren[1], **Arang Rhie**[1], **Sergey Nurk**[1], **Brian P. Walenz**[1], **Adam M. Phillippy**[1]

1 Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Recent advancements in technology and algorithms have enabled the completion of a human genome by the Telomere-to-Telomere consortium (T2T)(1). Key to this success was a conservative assembly strategy followed by validation and curation (2). While a handful of issues remain, analyses indicate the new assembly is a better reference than GRCh38 (3). Automated tools have also claimed to fully resolve human chromosomes and agriculturally important genomes, such as rice (4,5). However, with this great power comes a great responsibility to ensure that modern genome assemblies are not only continuous but accurate. We evaluated recent assemblies using both reference-based and reference-free methods (6,7,8), and found large variability between assemblers. On CHM13, sequenced with PacBio HiFi, HiCanu (9) had the fewest structural errors and 100% of its reconstructed centromeric arrays were accurate. Despite being more continuous, hifiasm (10) had more errors and only 50% of its centromeric arrays were accurate. For Oxford Nanopore ultra-long data of the same genome, Flye (11) was most accurate but with no complete centromeric arrays. Canu, using uncorrected ONT data, reconstructed the most centromeric arrays but only 30% were accurate. In a recently published rice assembly (5), we identified potential mis-assemblies including collapsed large-scale repeats. We hope these evaluations will serve as a guide to ensure not only high continuity but also high accuracy in the T2T era of genomics.

1. Nurk S, et al. “The complete sequence of a human genome.” biorXiv (2021).
2. Mc Cartney AM, et al. “Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies.” biorXiv (2021).
3. Aganezov S, et al. “A complete reference genome improves analysis of human genetic variation.” biorXiv (2021).
4. Awad M, et al. “GALA: gap-free chromosome-scale assembly with long reads.” biorXiv (2020).
5. Song JM, et al. “Two gap-free reference genomes and a global view of the centromere architecture in rice.” Mol, Plant (2021).
6. Gurevich A, et al. “QUAST: quality assessment tool for genome assemblies,” Bioinformatics (2013).
7. Mikheenko A, et al. “TandemTools: mapping long reads and assessing/improving assembly quality in extra-long tandem repeats.” Bioinformatics. (2020).
8. Formenti G, et al. “Merfin: improved variant filtering and polishing via k-mer validation.” biorXiv (2021).
9. Nurk S, et al. “HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads.” Genome Res. (2020).
10. Cheng H, et al. “Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.” Nat. Meth. (2021).
11. Kolmogorov M, et al., “Assembly of Long Error-Prone Reads Using Repeat Graphs”, Nat. Biotech. (2019).

Graph-based pangenome analysis of a haplotype-resolved genome assembly

Reference Genome Assembly Techniques and Initiatives

POSTER **814**

Bart Nijland[1], **Lidija Berke**[1], **Peter Van Dam**[1], **Stephan Ferreira**[1], **Eef Jonkheer**[2], **Sandra Smit**[2]

1 Genetwister Technologies B.V., Wageningen, The Netherlands
2 Wageningen University, Bioinformatics group, Wageningen, The Netherlands

Haplotype resolved assemblies provide the most accurate information about genetic content of an individual. However, such assemblies are difficult to generate and require novel analysis approaches that are still in their infancy. Here we report on different approaches for the assembly and analysis of a haplotype-resolved genome assembly of the most important ornamental nightshade, petunia.
Using the PacBio HiFi long-read sequencing technology, we first generated a conventional haploidized assembly of a highly heterozygous petunia hybrid. The assembly was 1.8 Gb in size, with the L50 statistic of 7 and the N50 statistic over 120 Mb. We further scaffolded the assembly into 7 chromosomes using Phase Genomics' Hi-C. To our knowledge, this represents the highest-quality petunia genome assembly and which is comparable to the quality of the reference assemblies of the related crop species tomato and potato. To obtain a haplotype-resolved genome assembly, we followed two parallel approaches. The first, trio-binning, is considered the gold standard and uses short-read data from the two parents to separate and assemble the two haplotypes. By contrast, the second approach based on Hi-C only used information from the analyzed petunia hybrid. As these approaches are typically not applied concurrently, we are taking full advantage of this experimental setup. We are currently comparing these approaches, in particular to identify misassemblies and haplotype switching, which will provide invaluable insights into haplotyping of complex heterozygous hybrid genomes.

Tools for genome analyses generally do not support haplotype-resolved assemblies. We are addressing this challenge by exploiting the graph-based pangenomics toolkit PanTools. Using the haplotype-resolved genome mode of PanTools, we will investigate structural differences as well as gene presence/absence variation between haplotypes. This innovative pangenome approach for analyses of haplotype-resolved assemblies completes the final step in the experimental workflow from assembling the genome of a single individual to generating insights into genetic variation.

Systems Biology

Systems Biology

POSTER 400

The Agricultural Genome to Phenome Initiative (AG2PI)

Jack C. Dekkers[1], **Jennifer Clarke**[2], **Carolyn J. Lawrence-Dill**[1], **David Ertl**[3], **Eric Lyons**[4], **Brenda M. Murdoch**[5], **Christopher K. Tuggle**[1], **Patrick S. Schnable**[1]

1 Iowa State University, Ames, IA
2 University of Nebraska-Lincoln, Lincoln, NE
3 Iowa Corn Growers Association, Johnston, IA
4 University of Arizona, Tucson, AZ
5 University of Idaho, Moscow, ID

To achieve sustainable genetic improvements of agricultural species, the expertise of a broad community of researchers must be engaged from both the crop and livestock communities and beyond. Genome to Phenome (G2P) research also requires integrative disciplines such as statistics and the data and engineering sciences. The objective of the Agricultural Genome to Phenome Initiative (AG2PI) is to assemble and prepare a transdisciplinary community to identify G2P research priorities and conduct G2P research. The genetics and genomics of crops and livestock have more in common than they do differences. However, those conducting the research generally collaborate with others working on closely related species rather than those working in dissimilar systems. Sharing information between the plant and animal kingdoms can eliminate redundancy, streamline the research process and reduce associated costs, and perhaps most importantly result in synergistic advances in both analytical techniques and biological understanding, thus improving the entire research pipeline. To accomplish this, AG2PI seeks to engage a broad and diverse community of researchers from academia and industry through Field Days, Conferences, Surveys, Training workshops, and Seed grants. See ag2pi.org for more information. In this poster, we provide an overview of the AG2PI project and opportunities for your involvement and collaboration. AG2PI is funded by USDA NIFA grants #2020-70412-32615 and #2021-70412-35233.

Systems Biology

POSTER 401

Equine FAANG Data Portal: interoperable and reusable derived datasets

Ted Kalbfleisch[1], **Douglas Antczak**[2], **Caitlin Castaneda**[3], **Brian W. Davis**[4], **Sian Ann Durward-Akhurst**[5], **Terje Raudsepp**[3], **Teruaki Tozaki**[6], **Molly E. McCue**[5], **Rebecca R. Bellone**[7], **Carrie J. Finno**[8], **Jessica L. Petersen**[9]

1 Department of Veterinary Science, University of Kentucky, Lexington, KY
2 Cornell University, Ithaca, NY
3 Texas A&M University, College Station, TX
4 Texas A&M University Interdisciplinary Program in Genetics & Genomics, College Station, TX
5 University of Minnesota, St. Paul, MN
6 Laboratory of Racing Chemistry, Tochigi, Japan
7 Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California-Davis, Davis, CA
8 University of California-Davis, Davis, CA
9 University of Nebraska-Lincoln, Lincoln, NE

The raw and derived data generated by the FAANG project or other externally funded work is valuable and has great utility. These data are currently submitted to data coordination centers such as the NCBI Sequence Read Archive, the European Nucleotide Archive, or project specific data coordination centers such as the FAANG Data Portal. These raw, and some derived, datasets are available for use under data use agreements which typically allow for use with few restrictions. Currently, both raw and derived data need to be downloaded in their entirety, and the raw data need to be mapped for any secondary analysis. Both these activities can pose formidable barriers to use of the data. Here we describe a data portal for URL based access of derived datasets that can be readily loaded into common tools such as IGV, the UCSC genome browser, or even command line tools able to read indexed files via a URL datastream. This allows scientists to quickly view subsets of mapped data to answer questions such as, "Is this transcript expressed in this tissue," or "Is there a histone mark in a particular tissue proximal to a polymorphic site." The system described here provides access to data generated in the Equine FAANG project, but the plan is to extend this to other equine datasets contributed by the community, as well as other species, and is envisioned to serve as a nexus for not only data, but containerized workflows to support standardized analyses.

Integrative genomic studies of gene networks and dynamics in growth and stress responses in bioenergy plants, fungi, and bacteria

Systems Biology

POSTER 402

Ronan O'Malley[1], **Leo Baumgart**[1], **Sharon Greenblum**[1], **Yuko Yoshinaga**[1], **Cindy Chen**[1], **Yu Zhang**[1], **Maxim Koriabane**[1]

1 U.S. Department of Energy, Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

Phenotypic variations that are central to crop improvement result from either the emergence of novel gene functions and pathways, or through the modifications of gene expression dynamics within the conserved pathways. While both mechanisms can contribute to phenotypic innovations, recent studies suggest that the changes in gene regulation may be more common, particularly in shorter evolutionary time frames and domestication. While comparative transcriptomics along and gene network modeling have proven to be powerful approaches to characterize regulatory plasticity, most studies to date have relied on bulk tissue RNA-seq analysis and as such fail to capture the cell-type resolution required to clearly understand the composition and function of plant gene networks. Furthermore, as changes in gene regulation are a direct consequence of changes in transcription factor binding at promoters and enhancers, comprehensive genome-wide transcription factor binding site atlases are crucial for deciphering how pathways are rewired during domestication and adaptive selection. To address these questions, we are developing integrative approaches to leverage complementary technologies including single-cell transcriptomics and chromatin accessibility assays (single cell RNA-seq and ATAC-seq) along with ultra-high throughput transcription factor binding assay (DAP-seq and multiDAP), to understand how pathways and ultimately plant traits are shaped through alterations in gene expression. We will present our current progress in applying these technologies in important bioenergy plants, fungi and bacteria to develop integrative atlases of gene networks to understand how they control and shape important growth and stress response pathways at the levels of DNA sequence, cell-type and tissue regulation.

Using galaxy for large-scale projects: improvements in data acquisition, storage, and filtering

Systems Biology

POSTER 403

Alex Ostrovsky[1], **Delphine Lariviere**[2], **Cristobal Gallardo**[3], **Dave Bouvier**[2], **Anton Nekrutenko**[2], **Björn Grüning**[3], **Michael C. Schatz**[1]

1 Johns Hopkins University, Baltimore, MD
2 Penn State University, University Park, PA
3 University of Freiburg, Freiburg, Germany

Galaxy (galaxyproject.org) is an open, web-based platform for accessible, reproducible, and transparent computational research. Though originally intended for life science applications, it has expanded into many areas of computational research including cheminformatics, climate science, and image analysis. As a community based project, Galaxy is a platform to facilitate an individual's or a team's research by providing a graphical user interface and backend infrastructure to design, execute, customize, publish, and reproduce analyses.

With recent partnerships with the Vertebrate Genomes Project, the NHGRI AnVIL, and other major efforts, several enhancements have been made to Galaxy's ability to be utilized on large-scale projects, allowing parallelized efforts without the worry of overlap or exhausting computing resources. This poster will highlight Galaxy's new capabilities in organizing, retrieving, analyzing, and uploading data to private repositories or buckets, current efforts on the globalization and integration of public computational infrastructure from the EU and US, unification of data analysis tool ecosystem, and development of workflow language and its interplay with other pipeline systems. Further, we will demonstrate the new mechanisms that have been recently implemented to streamline large-scale data analysis across many datasets via multiple workflows for genome assembly and analysis, highlighting the integration of these new capabilities in ready-to-run pipelines, using the Vertebrate Genomes Project as a currently-implemented example.

Extend mixed models to multi-layer neural networks for genomic prediction including intermediate omics data

Systems Biology

POSTER 404

Tianjing Zhao[1,2*], **Jian Zeng**[3], **Hao Cheng**[1]

1 Department of Animal Science, University of California, Davis, Davis, CA
2 Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA
3 Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

*tjzhao@ucdavis.edu

Background With the growing amount and diversity of intermediate omics data complementary to genomics (e.g., DNA methylation, gene expression, and protein abundance), there is a need to develop methods to incorporate intermediate omics data into conventional genomic evaluation. The effects of genotypes on phenotypes can be mediated by multiple layers of omics features through mechanisms such as regulatory cascades from epigenome, to transcriptome, and to proteome, which forms a connected multi-layer network naturally.

Methods We developed a new method named NN-LMM to model the multiple layers of regulation from genotypes to intermediate omics features, then to phenotypes, by extending conventional linear mixed models ("LMM") to multi-layer artificial neural networks ("NN"). NN-LMM incorporates intermediate omics features by adding middle layers between genotypes and phenotypes. Linear mixed models (e.g., pedigree-based BLUP, GBLUP, Bayesian Alphabet, single-step GBLUP, or single-step Bayesian Alphabet) can be used to sample marker effects or genetic values on intermediate omics features, and nonlinear activation functions in neural networks are used to approximate the nonlinear relationships between intermediate omics features and phenotypes. Compared to other methods, NN-LMM provides a more flexible and robust framework to incorporate intermediate omics features. NN-LMM allows various patterns of missing omics data, for example, individuals can have different missing omics features, and the assumption of nonlinearity between intermediate omics features and phenotypes may be more biologically realistic. NN-LMM has been implemented in an open-source package called "JWAS".

Results In simulation analysis, NN-LMM had significantly higher prediction accuracy than the recently proposed single-step approach for genomic prediction with incomplete intermediate omics data. When the underlying relationship between intermediate omics features and phenotypes was nonlinear, using the nonlinear activation function in NN-LMM had significantly better performance than using the linear activation function. Moreover, using an informative prior for the marker effects in NN-LMM further increased the prediction accuracy, as compared to using an infinitesimal model.

Index

Index

P	Passalacqua	Michael	Cold Spring Harbor Laboratory	84
	Petersen	Jessica	University of Nebraska-Lincoln	85, 123
	Pirani	Ali	Thermo Fisher Scientific	103
Q	Qu	Jiayi	University of California, Davis	25
R	Rao	Varsha	Claret Bioscience	115
	Rautiainen	Mikko	National Institutes of Health	116
	Rosado	Daniele	Cold Spring Harbor Laboratory	67
	Rowan	Troy	University of Tennessee	21, 26, 29
S	Safonova	Yana	University of California San Diego	30
	Schatz	Michael	Johns Hopkins University	106,110,118,125
	Song	Jiuzhou	University of Maryland	60
	Spannagl	Manuel	Helmholtz Zentrum München	77, 82, 112
	Stitzer	Michelle	Cornell University	61
	Stull	Matthew	Genomics And Bioinformatics Service	91, 104
	Suren	Haktan	Thermo Fisher Scientific	27, 92, 94, 98
T	Tsakirpaloglou	Nikolaos	Texas A&M University and Texas A&M AgriLife Research	47
	Tuggle	Christopher	Iowa State University	24, 68, 87, 122
	Tuyishimire	Esdras	University of Missouri-Columbia	26
V	Vakulskas	Christopher	Integrated DNA Technologies	44, 48
	Vallebueno	Miguel	Gregor Mendel Institute	54
	Van Eenennaam	Alison	University of California, Davis	46, 51
	Verbyla	Klara	Center for Aquaculture Technologies	35
W	Wang	Guosong	Texas A&M University	31
	Wilson	Lena	North Carolina State University	28
	Wulf	Donat	Bielefeld University	86
Y	Yang	Pengxin	Iowa State University	87
	Yaschenko	Anna	North Carolina State University	69
Z	Zebell	Sophia	Howard Hughes Medical Institute, Cold Spring Harbor Laboratory	45, 49
	Zhao	Tianjing	Department of Animal Science, University of California Davis	126