PEQG22

Population, Evolutionary, and Quantitative Genetics Conference

June 7–10, 2022

ABSTRACT BOOK

GENETICS GSA G3 Genes | Genomes | Genetics

4 Evolutionary forces shape the genetic architecture of complex traits Ryan Hernandez UCSF

Understanding the genetic architecture of complex traits is a central challenge in human genetics. There often remains a large disparity between heritability estimates from family-based studies and large-scale genome-wide association studies (GWAS), which has been sensationalized as the "missing heritability problem". Among the possible explanations for this disparity are rare variants of large effect that are neither tagged by existing genotyping platforms nor well imputed from existing reference panels. In this talk, I will discuss how historical and recent evolutionary forces shape the distribution of genetic variants in the genome, as well as the population genetic conditions in which these evolutionary forces impact the genetic architecture of complex traits. I will also discuss how dynamic admixture processes shape patterns of genetic ancestry in Latiné populations, which may be contributing to a shifting genetic architecture of complex traits in these populations.

5 High-Resolution Exploration of Collateral Sensitivity using Molecularly Barcoded *S. Cerevisiae* *Sam Apodaca*[1], Kara Schmidlin[1], Daphne Newell[1,2], Kerry Geiler-Samerotte[1,2] 1) Arizona State University Biodesign Institute, Tempe, AZ; 2) Arizona State University School of Life Sciences, Tempe, AZ

A major challenge for modern medicine is the increasing prevalence of drug-resistant microbes. In parallel with efforts to discover candidates for new antimicrobials, research is increasingly focusing on how the efficacy of existing drugs can be extended. For example, collateral sensitivity is an evolutionary phenomenon in which the mutation that allows a microbe to resist one drug makes it more susceptible to a second drug. While it has been suggested that this could be a potential treatment strategy, it only becomes viable if the phenomenon can be predicted and reproduced reliably across multiple resistant mutants. To make such predictions, we believe that it is necessary to consider all resistance mutations that can occur in a given drug condition, not just the most successful one. To do this, we utilized a library of molecularly barcoded *S. cerevisiae* to study the evolution of resistance to two drugs, fluconazole and radicicol. This enabled us to observe ~300,000 lineages as they evolved in 12 different concentrations and combinations of the two drugs over 200 generations. Sequencing data was used to track barcode frequencies over time, allowing us to identify hundreds of resistant lineages as they arose and quantify differences in how resistance evolution occurs under different conditions. To explore the collateral fitness effects of these resistance mutations, adaptive lineages from the evolved populations were pooled together and their fitnesses were remeasured in each of the 12 drug conditions. We successfully identified many lineages demonstrating collateral sensitivity, in other words, lineages that possessed mutations that were highly adaptive in their home environment but deleterious in others. But these patterns of collateral sensitivity were not reliable across different mutants. Barcode frequencies revealed that adaptive lineages from the same home conditions did not have similar, predictable fitness when exposed to non-home environments. Finally, we isolated lineages of interest and used whole-genome sequencing to identify the specific mutations that conferred these fitness effects. Our barcode system has allowed us to expand on previous research into collateral sensitivity by greatly increasing the number of lineages that can be studied simultaneously, highlighting the importance of accounting for more than just the highest frequency mutations when making predictions about adaptation.

6 Shared Features of Complex Trait Architecture Explained by Underlying Selection *Yuval Simons*[1], Hakhamanesh Mostafavi[1], Jonathan Pritchard[1], Guy Sella[2] 1) Stanford University, Stanford, CA; 2) Columbia Universtiy, New York, New York

In recent years, genome-wide association studies (GWAS) have begun to systematically uncover the genetic architecture of complex traits. Complex traits seem to greatly differ in their genetic architecture, differing in the number of genome-wide significant hits discovered and the proportion of trait variance these hits explain. We sought to understand the evolutionary basis of such differences using a model of pleiotropic stabilizing selection. Using our model, we used the frequencies and z-scores of GWAS hits to infer the distribution of selection coefficients at trait-affecting loci and the number of trait-affecting loci, i.e. the target size. Using 96 continuous traits that have over 100 hits in the UK biobank, we discovered that variants underlying most traits have surprisingly similar distributions of selection coefficients. Intuitively, sufficiently polygenic traits sample selection coefficients from the same common pool of functional variation spread throughout the genome.

This common distribution of selection coefficients implies that differences in trait architectures between highly polygenic traits arise almost exclusively from differences in heritability and target size. There are order of magnitude differences in the heritability of the traits we consider due to different degrees of environmental contribution. Traits differ even more in target size: we infer that target sizes span over two orders of magnitude. However, once these two factors are accounted for, we show that the architecture of almost all of these 96 traits is essentially identical.

We explore the implications of this result. First, we show that the proportion of variants discovered in GWAS and the proportion of heritability they explain has near-identical dependency on study size for all traits, after normalizing study size by the heritability per trait-affecting site. Second, we show that the lack of portability of GWAS-derived polygenic risk scores is mainly driven by genetic drift and selection, and this reduction should be near-identical across traits. Interestingly, about 150 GWAS hits, mostly of low frequency and large effect size, stand out as outliers to our model and many of these hits affect genes well-known to influence trait biology. Thus, deviations from our model may offer clues to trait biology.

7 When should we expect adaptation via a highly polygenic response vs selective sweeps? *William Milligan*[1], Laura Hayward[3], Guy Sella[1,2] 1) Department of Biological Sciences, Columbia University, New York, NY; 2) Program for Mathematical Genomics, Columbia University, New York, NY; 3) Institute of Science and Technology Austria, Klosterneuburg, Austria

The genetic basis of adaptation and its predictability are poorly understood. Examples of adaptation where the genetic basis has been mapped implicate selective sweeps at a few large-effect loci, despite our expectation that polygenic adaptation, which occurs via allele frequency changes at many small-effect loci, should be ubiquitous. Theoretical modeling of adaptation has been limited to only a few extreme settings (e.g., highly polygenic or Mendelian), and we do not have the theoretical basis to understand how adaptation behaves under other conditions. Here, we address this by modeling adaptation after a sudden shift in fitness optimum of a quantitative trait under a variety of settings, where we vary 1) the shift size, 2) the trait's genetic complexity – ranging from the Mendelian to the highly polygenic extreme, and 3) the distribution of fitness effect, e.g., the proportion of large-effect alleles. We then describe both the phenotypic response to selection (changes in trait mean and variance over time) and the allelic response (e.g., which alleles contribute to adaptation). Importantly, we find sweep-like behavior is common when polygenicity is low, large effect alleles are common, and shift sizes are large relative to the heritable variance; otherwise, polygenic adaptative responses are the norm. We derive quantitative conditions on model parameters that predict different kinds of phenotypic and genetic responses. Our results help to bridge the gap between theory and known examples of adaptation, and inform efforts to identify the signals of adaptation in humans and other species. Additionally, the conditions for selective sweeps overlap with the conditions for parallel evolution, so our results also bear on understanding the predictability of adaptation.

8 Guaranteeing unbiasedness in selection tests based on polygenic scores *Jennifer Blanc*, Jeremy Berg University of Chicago, Chicago, IL

Population stratification is a well-studied problem in genome-wide association studies, leading to biases in the estimated strength of phenotypic association for individual genetic variants. In short, if environmental effects on the phenotype are correlated with ancestry gradients within a GWAS panel, any variant that is stratified along this ancestry gradient will receive a biased effect size estimate. While state of the art methods to correct for stratification are generally effective in reducing the number of significant false positive associations, even subtle biases in effect size estimates can accumulate across loci, leading to systematic biases in polygenic scores. In turn, these biases in the distribution of polygenic scores can lead to false positives in downstream analyses, such as tests for polygenic adaptation or other analyses of among group genetic differences. One approach is to be extremely aggressive in the use of fixed effects genetic PCs to control for stratification. However, there is currently no way to tell conclusively if confounding effects have been removed. A second approach is to conduct the GWAS in an evolutionarily diverged sample that is less likely to share population genetic structure with the test panel. This renders potential biases in the effect sizes irrelevant to the test, but comes at the cost of significantly reduced statistical power due to the poor portability of polygenic scores across samples of divergent ancestry. Using theory from population and statistical genetics, together with simulations, we show that even if GWAS and test panels do share genetic structure it is possible to guarantee the unbiasedness of polygenic selection tests without needing to achieve the much more difficult task of guaranteeing that the effect sizes are completely unbiased. We show that by analyzing GWAS and test panels jointly in a unified framework, we can leverage the observed overlap in population structure between the two samples so as to protect the GWAS from stratification biases along the relevant axis of shared structure. More generally, our results have implications beyond tests for selection as any analysis that attempts to quantify the covariance between polygenic scores and demographic or environmental variables is subject to the same type of stratification biases, and can therefore benefit from our framework.

9 Improving Phenotype Prediction by Learning Patterns of Sharing across Multiple Phenotypes *Fabio Morgante*[1,2], Gao Wang[3,4], Yuxin Zou[5], Abhishek Sarkar[4], Peter Carbonetto[4], Yang Li[2,4], Matthew Stephens[4,5] 1) Center for Human Genetics, Department of Genetics and Biochemistry, Clemson University, Greenwood, SC; 2) Section of Genetic

Medicine, Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Neurology, Columbia University, New York City, NY; 4) Department of Human Genetics, University of Chicago, Chicago, IL; 5) Department of Statistics, University of Chicago, Chicago, IL

Predicting phenotypes from genotypes is a fundamental task in quantitative genetics. While phenotype prediction was pioneered in agricultural breeding for selection purpose, it has recently become important in human genetics as well. In fact, predicting disease risk and other medically relevant phenotypes via Polygenic Risk Scores (PRS) is one of the main goals of precision medicine.

With technological advances, it is now possible to measure multiple phenotypes in large samples. Multiple phenotypes can share their genetic component; therefore, modeling these phenotypes jointly may improve prediction accuracy by leveraging the shared effects across such phenotypes. However, effects can be shared across phenotypes in a variety of ways, so computationally efficient statistical methods are needed that can accurately and flexibly capture patterns of effect sharing.

Here, we describe new Bayesian multivariate regression methods that, by using flexible priors, are able to model many different patterns of effect sharing and specificity across phenotypes. We evaluated our methods in their ability to predict gene expression in multiple tissues using simulations with different patterns of effect sharing across tissues and genomic heritabilities. The results suggest that these new methods can predict better than existing univariate and multivariate methods, while also being computationally efficient. We then sought to replicate those results on real data by analyzing the Genotype Tissue Expression (GTEx) project data. We showed that our methods improve prediction performance on average for all the tissues, especially for those groups of tissues where shared effects have been previously described.

While gene expression prediction was used as an application, our methods are general multivariate regression methods that can be used for any multi-phenotype applications, including PRS computation and breeding value prediction. Thus, our methods have the potential to provide improvements across fields and organisms.

10 Climate-driven natural selection across protein-coding and cis-regulatory genetic variation *Shannon Hateley*[1,2], Marlot Westera[1,3], Kristy Mualim[1,2], Moises Exposito-Alonso[1,2] 1) Carnegie Institution for Science; 2) Stanford University; 3) Utrecht University

How do genetic polymorphisms steer climate adaptation? While extensive theory and data in human genomics posit that complex adaptive traits are primarily driven by changes in gene expression networks, study of coding variants in the plant model *Arabidopsis thaliana* finds hundreds of natural loss-of-function alleles involved in adaptation. However, the relative contributions of regulatory and protein coding variation to fitness in *A. thaliana* in ecologically-relevant conditions is still unknown. To assess the impact of natural selection on genetic variants in *A. thaliana*, we utilized fitness measurements from ~500 diverse ecotypes grown under high and low precipitation at two common garden field sites within the species' native range and quantified short-term natural selection on ~11 million variants across the *A. thaliana* genome. We next annotated the naturally-selected and neutral variants employing multiple methods: 1) We overlaid protein coding effect predictions and cis-regulatory information from existing genome annotations and RNA-seq and DAP-seq datasets. 2) We developed a machine-learning model to query the effect of individual cis-regulatory variants and identify those most likely to have functional consequence on chromatin accessibility and gene expression. And 3) because study of epigenetic variation in *A. thaliana* has found DNA methylation variation contributes substantially to natural variation in adaptive traits, we utilized bisulfite sequencing datasets to annotate differentially methylated variants in gene bodies and across the genome. The strength of short-term natural selection over variants was studied across such annotations relative to carefully matched neutral alleles. Controlling for variant type abundance, frequency, and genotyping rate, we assessed the impact of selection on coding and cis-regulatory variants, alone and in context of differential methylation status, and performed enrichment analysis of variant types, functions, and gene ontologies. We find that nonsynonymous mutations, while less common overall, have an outsized contribution to fitness relative to regulatory mutations, and that the impact of methylation status is context dependent, contingent upon methylation type and genomic region. This work advances our understanding of the architecture and functional nature of adaptive genetic variation.

11 Genotype-by-Diet interactions unmask cryptic genetic variants that regulate lifespan in outbred *Drosophila* *Luisa Pallares*[1], Amanda Lea[2], Clair Han[3], Peter Andolfatto[4], Julien Ayroles[5] 1) Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany; 2) Vanderbilt University, Nashville, TN; 3) Janelia Research Campus, Ashburn, VA; 4)

Columbia University, New York, NY; 5) Princeton University, Princeton, NJ

Evolutionary theory suggests that lifespan-associated alleles should be purged from the gene pool, and yet decades of GWAS and model organism studies have shown they persist. Here, we explore one possible explanation: alleles that regulate lifespan are context dependent. This idea is core to the "evolutionary mismatch" hypothesis. It predicts that previously adaptive or neutral alleles in human populations have become mismatched to our current lifestyle underlying the high incidence of non-communicable diseases that impact lifespan today. However, the lack of statistical power to identify genotype-by-environment interactions at a genome-wide scale have limited our ability to test these hypotheses. To address this problem, we exposed thousands of outbred *Drosophila* to a standard and a high sugar diet. We then sequenced over 10,000 individuals and track genome-wide allele frequency changes over time, as these populations aged. We mapped thousands of lifespan-altering alleles whose frequency changed throughout the course of the population's lifetime, and remarkably, a third of these lifespan-associated alleles appear cryptic in standard diet but play an important role in high sugar conditions. The identification of such large number of SNPs allowed us for the first time to test key predictions of the evolutionary mismatch hypothesis at genome-wide level. Specifically, we find that alleles that are now detrimental (reducing lifespan) are most likely to be recently derived, have stronger effects on a high-sugar diet, and were positively selected during the evolutionary history of the fruit fly. Our observations provide a) strong evidence for the pervasive nature of cryptic genetic variation and the key role it plays in shaping phenotypic variation between individuals, and b) support the hypothesis that historically neutral or beneficial alleles can become detrimental in novel conditions.

12 Dominance genetic effects on complex traits in pigs, rats and mice are associated with trans-acting dominance gene expression effects *Leilei Cui*[1,2,3], Bin Yang[1], Shijun Xiao[1], Jun Gao[1], Amelie Baud[4], Jonathan Flint[5], Richard Mott[2], Lusheng Huang[1] 1) Jiangxi Agriculture University, Nanchang, China; 2) UCL Genetics Institute, London, UK; 3) Nanchang University, Nanchang, China; 4) European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK; 5) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, US

Dominance arises from the interaction between different alleles within the same genomic locus, which historically played a major role in the study of biological traits. However, today most genome-wide association studies (GWAS) generally ignore dominance effects. This might bias and hinder our understanding of the mechanism and architecture of complex traits, and potentially fail to detect dominant quantitative trait loci (QTLs). In this study, we surveyed the dominance effect contributions to genetic architecture across many organismal and gene expression traits across three outbred mammalian populations: the White Duroc x Erhualian pig F2 intercross stock (n=1,005), the rat heterogeneous stock (n=1,407) and the mouse heterogeneous stock (n=2,002). We estimated additive and dominance variance components for each trait based on the whole genome SNPs, then performed GWAS by fitting both additive and dominance effects at each SNP using our newly developed software ADDO. We further mapped gene expression QTLs (eQTLs) of different tissues in all populations, and also investigated the relationship between their inheritance modes and regulation types. On average dominance accounted for 11% (from 0% to 57%) phenotypic variances in pigs, about one-third of the additive contributions, and similar to that observed in rats and mice. Interestingly, traits related to hematology and immunology generally tend to have a high dominance variance. We identified 157, 66 and 60 novel loci showing significant complete dominance or overdominance effects in pigs, rats and mice. The proportions of QTL types across these three species are similar: 55.6% (additive): 16.1% (complete dominance): 28.3% (overdominance). Most strikingly, across all populations, cis-acting eQTLs are predominantly additive while trans-acting eQTLs are enriched for dominance effects. A search for genes causal for dominance physiological QTLs revealed these are unlikely to be physically linked to the QTLs but instead are associated via trans-acting dominance eQTLs. This study suggests that the dominance effect contribution to the genetic architecture of complex traits in mammals is both important and has a distinct mechanism from additive contributions.

13 Evolutionary dynamics in simulated gene regulatory networks *Anastasia Teterina*[1,2], Peter Ralph[1,3], Patrick Phillips[1] 1) Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA; 2) Severtsov Institute of Ecology and Evolution RAS, Moscow, Russia; 3) Department of Mathematics, University of Oregon, Eugene, OR, USA

Understanding how the genotype of an individual, through development, phenotype, and interaction with the environment, maps to its fitness is a fundamental question in evolutionary, quantitative, and systems biology. The structure of the gene regulatory networks (GRNs) that generate phenotypes can be shaped by both adaptive and non-adaptive forces. Many critical properties of GRNs, such as robustness against new mutations, genetic variability under stochastic and selective processes, evolvability of different topologies, have been studied using evolutionary simulations (Wagner

1996, 2008, 2012; Ibáñez-Marcelo & Alarcón 2014; Payne & Wagner 2014; Kioukis & Pavlidis 2018). To connect the evolutionary dynamics of GRNs with estimates and is possible to obtain empirically, we developed a simulation framework in SLiM3 (Haller & Messer 2016, 2019) with selection on gene expression levels in evolvable, dynamic gene regulatory networks encoded in realistically-sized genomes. Using the genomic data from such individual-based simulations, we estimated genetic diversity along the genome, conducted in silico molecular biological experiments such as estimation of the effects of knockout or overexpression of the gene on phenotypes, genetic variances, epistatic interactions, and described the structure and heterogeneity of evolved networks. We explored the correlations between genetic diversity, quantitative genetic measures, network centrality statistics, phenotypic effects of molecular manipulations on the genes, and strength of selection on genes. Then, we assessed how the evolutionary trajectories of genes depend on their roles in GRNs, and evaluated our ability to predict distributions of fitness effects and the trajectory of evolution based on network structure. Finally, we utilized neural networks to identify the key components of the GRNs that changed under directional and stabilizing selection.

14 ARG-based Association Mapping *Vivian Link*, Caoqi Fan, Charleston Chiang, Nicholas Mancuso, Michael "Doc" Edge University of Southern California, Los Angeles, CA

Understanding the genetic basis of complex phenotypes through genome-wide association (GWA) studies is a central pursuit of human medical genetics. GWA methods are widely and successfully used, but they face challenges. Many of these challenges are related to the fact that variants are tested for association with a phenotype independently, whereas in reality variants at different sites are correlated because of their shared evolutionary history. One way to explicitly model this shared history is through the ancestral recombination graph (ARG), which can be thought of as a series of local coalescent trees. Recent computational and methodological breakthroughs have made it feasible to estimate accurate ARGs from biobank samples. Here, we explore the potential of a generalized ARG-based approach to association mapping. In terms of a local tree in the ARG, standard GWAS can be understood as testing a branch on which a typed mutation occurred for association with the phenotype. We propose a framework in which an arbitrary number of branches can be tested for association simultaneously. Standard GWAS, Identity-By-Descent Mapping, and some methods for local heritability estimation can be viewed as special cases under our generalized framework. Our methods may be especially beneficial for finding associations for phenotypes with multiple causal alleles in the same locus (allelic heterogeneity), and the ARG may also efficiently capture untyped variation. Another potential benefit of our framework is that it can reduce the number of tests performed compared with GWAS, alleviating the multiple testing burden. We illustrate the methods using simulations under realistic demographies and further explore how other tasks related to association mapping, such as fine mapping, can be viewed in terms of the ARG. We expect that by framing association mapping in terms of the ARG, we can increase statistical power to detect associations, and that our investigations may provide intuition about the benefits of using the ARG in population genomic methods in general.

16 Range expansion promotes introgression upon secondary contact *Ailene MacPherson*[1,2,7], Silu Wang[2,3,4], Ryo Yamaguchi[5,6], Loren Rieseberg[2], Sarah (Sally) Otto[2] 1) Simon Fraser University; 2) University of British Columbia; 3) University of California Davis; 4) University of California Berkeley; 5) Tokyo Metropolitan University; 6) Hokkaido University; 7) University of Toronto

Population genomic analysis of hybrid zones is instrumental to our understanding of the evolution of reproductive isolation. Many temperate hybrid zones are formed by the secondary contact between two parental populations that have undergone post-glacial range expansion. In this work, we showed that explicitly accounting for historical parental isolation followed by range expansion prior to secondary contact is fundamental for explaining genetic and fitness patterns in these hybrid zones. Specifically, ancestral population expansion can result in allele surfing where neutral or slightly deleterious mutations drift to the high frequency at the expansion front. If these surfed deleterious alleles are recessive, they can contribute to substantial heterosis in hybrids produced at secondary contact, counteracting negative effects of Bateson-Dobzhansky-Muller incompatibilities (BDMIs) hence weakening reproductive isolation. When BDMIs are linked to such recessive deleterious alleles the fitness benefit of introgression at these loci can facilitate introgression at the BDMIs. The extent to which this occurs depends on the strength of selection against the linked deleterious alleles and the distribution of recombination across the chromosome. Finally, surfing of neutral loci can alter the expected pattern of population ancestry, thus accounting for historical population expansion is necessary to develop accurate null genomic models of secondary-contact hybrid zones.

17 An orthologous gene coevolution network provides insight into eukaryotic cellular and genomic structure and function *Jacob Steenwyk*[1], Megan Phillips[1], Feng Yang[2,3], Swapneeta Date[1], Todd Graham[1], Judith Berman[2], Chris Hitting-

er[4], Antonis Rokas[1] 1) Vanderbilt University, Nashville, TN; 2) Tel Aviv University, Ramat Aviv, Israel; 3) Tongji University School of Medicine, Shanghai, China; 4) University of Wisconsin-Madison, Madison, WI

Orthologous gene coevolution—which refers to gene pairs whose evolutionary rates covary across speciation events—is often observed among functionally related genes. We present a comprehensive gene coevolution network inferred from the examination of nearly three million orthologous gene pairs from 332 budding yeast species spanning ~400 million years of eukaryotic evolution. Modules within the network provide insight into cellular and genomic structure and function, such as genes functioning in distinct cellular compartments and DNA replication. Examination of the phenotypic impact of network perturbation across 14 environmental conditions using deletion mutant data from the baker's yeast *Saccharomyces cerevisiae* suggests that fitness in diverse environments is impacted by orthologous gene neighborhood and connectivity. By mapping the network onto the chromosomes of *S. cerevisiae* and the opportunistic human pathogen *Candida albicans*, which diverged ~235 million years ago, we discovered that coevolving orthologous genes are not clustered in either species; rather, they are most often located on different chromosomes or far apart on the same chromosome. The budding yeast coevolution network captures the hierarchy of eukaryotic cellular structure and function, provides a roadmap for genotype-to-phenotype discovery, and portrays the genome as an extensively linked ensemble of genes.

18 The genetic basis of tail-loss evolution in humans and apes *Bo Xia*[1,2], Weimin Zhang[2], Aleksandra Wudzinska[2], Emily Huang[2], Ran Brosh[2], Maayan Pour[1], Alexander Miller[4], Jeremy Dasen[4], Matthew Maurano[2], Sang Kim[5], Jef Boeke[2,3,6], Itai Yanai[1,3] 1) Institute for Computational Medicine, NYU Langone Health, New York, NY; 2) Institute for Systems Genetics, NYU Langone Health, New York, NY; 3) Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY; 4) Department of Neuroscience and Physiology, NYU Langone Health, New York, NY; 5) Department of Pathology, NYU Langone Health, New York, NY; 6) Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY

The loss of the tail is one of the main anatomical changes to have occurred along the evolutionary lineage leading to modern humans. This morphological reprogramming in the ancestral hominoids has been long considered to have accommodated a characteristic style of locomotion and contributed to the evolution of bipedalism in humans. Yet, the precise genetic mechanism that facilitated tail-loss evolution in hominoids remains unknown. Primate genome sequencing projects have made possible the identification of causal links between genotypic and phenotypic changes. In particular, a comparative genomics approach can screen for hominoid-specific genetic elements in genomic regions associated with genes known to be involved in controlling tail development. Here, we present evidence that tail-loss evolution was mediated by the insertion of an individual transposable element (TE) – *Alu* element – into the intron of the *TBXT* gene (also called *T* or *Brachyury*) of the hominoid ancestor genome.
Unlike traditional thought that TEs in the middle of the intron are of little biological impact, we hypothesized that this *Alu* element could pair with a neighboring ancestral *Alu* element encoded in the reverse genomic orientation and thus induce the formation of a stem-loop structure in the newly transcribed pre-mRNA. We demonstrated that such an interaction pair of TEs leads to a hominoid-specific alternative splicing event, thus affecting *TBXT* gene function and ultimately influencing tail development and evolution.
To study the physiological effect of this splicing event, we generated a mouse model that mimics the expression of human *TBXT* products by expressing both full-length and exon-skipped isoforms of the mouse *TBXT* orthologue. We found that mice with this genotype exhibit a complete absence of the tail or a shortened tail, supporting the notion that the exon-skipped transcript is sufficient to induce a tail-loss phenotype, albeit with incomplete penetrance. We further propose that selection for the loss of the tail along the hominoid lineage was associated with an adaptive cost of potential neural tube defect as an evolutionary trade-off, which may thus continue to affect human health today.

19 Agricultural adaptation of common waterhemp over the last two centuries *Julia M. Kreiner*[1,2], Sergio M. Lattore[3,4], Hernán Burbano[3,4], John R. Stinchcombe[2], Sarah P. Otto[5], Detlef Weigel[3], Stephen I. Wright[2] 1) Department of Botany, University of British Columbia; 2) Department of Ecology & Evolutionary Biology, University of Toronto; 3) Department of Molecular Biology, Max Planck Institute for Biology; 4) Division of Biosciences, University of College London; 5) Department of Zoology, University of British Columbia

North America has seen one of the greatest increases in agricultural land use over the last two centuries, and in the latter half of the 20th century, intensification of agricultural practices. Native plants that predate this transition and that persist in the face of extreme human-mediated disturbance present a remarkable opportunity to learn about the evolutionary consequences of contemporary land use and input regimes. We studied the extent and tempo of agricultural adaptation in a native plant now pervasive in agricultural habitats, common waterhemp (*Amaranthus tuberculatus*), through

sequencing paired samples from natural and agricultural environments from the present day and historical individuals spanning the last two centuries. Despite near panmixia among environments, numerous loci across the genome showed strong evidence of antagonistic selection. These loci, including alleles coding for resistance to herbicides, have increased in frequency by more than 20% in agricultural environments over the last ~150 years with nearly all such change occurring since the intensification of practices in the 1960s (s = 0.0267). Over the same period, we show a clear expansion of southwestern ancestry into the northeastern part of the range and particularly so in agricultural environments. Regions of the genome enriched for southwestern ancestry showed the strongest signals of agricultural selection, implying a strong link between demographic and selective contemporary change. The intensification of agriculture has thus had extensive evolutionary impacts on both genome-wide diversity and variation for fitness in this native plant, facilitating its success as a 21st-century agricultural weed.

20 Genetic basis of carotenoid coloration in birds. Malgorzata Gazda 1) Institut Pasteur, Paris, France; 2) CIBIO, Vairao, Portugal; 3) University of Porto, Porto, Portugal

My research aims to find the genetic bases of coloration in birds. I examined yellow, orange, and red carotenoid-based pigmentation, pivotal drivers of avian diversity, in order to understand how birds acquire color and how coloration differs between sexes. In order to establish an association between genotype and phenotype, I sequenced genomes of different canary breeds, identified candidate genes associated with difference in coloration and followed up with functional tests. Firstly, I investigated how birds acquire carotenoid colouration. I examined mechanisms of carotenoid uptake, taking advantage of the white recessive canary breed, which carries an autosomal recessive mutation causing white plumage. Biochemical analysis revealed a genetic defect in carotenoid uptake, then genomic analyses showed that the white recessive allele is caused by a splice-donor site mutation in the scavenger receptor B1 gene (SCARB1). This mutation leads to a loss of function of the gene, which we demonstrated through functional assays. Taken together, SCARB1 is an essential mediator of the expression of carotenoid-based coloration in birds and suggests a potential link between visual displays and lipid metabolism.
Secondly, I analysed the molecular mechanisms of sexual dichromatism in birds (differences in coloration between males and females). Using a unique model of dichromatic hybrid canary in combination with genomic and transcriptomic analyses, I showed that dichromatism in mosaic canaries is controlled by a small autosomal region that alters the expression of the BCO2 gene in males and females, but specifically in the integument. To put the results in a broader evolutionary context, I applied transcriptomic analyses to a continuum of sexual dichromatism in finches that showed the importance of BCO2 for sexual dichromatism in other species. I demonstrated how large differences in ornamental coloration between sexes can evolve from simple molecular mechanisms controlled by genes of major effect and have important implications for theories of honest signaling and sexual selection.
My work is a significant contribution towards a better understanding of the evolution of avian coloration.

21 Don’t put all your eggs in one basket: what stochastic modeling tells us about bet hedger evolution *Maya Weissman*, Yevgeniy Raynes, Daniel Weinreich Brown University, Providence, RI

When environmental change is unpredictable and rapid, risk reduction becomes vital to the long term success for a lineage. Bet hedging is one possible strategy for risk reduction in varying environments. Here, a lineage sacrifices its short term fitness in order to insulate itself against future risk. Bet hedging is defined as any strategy that lowers temporal fitness variance at the expense of mean fitness; strategies are classified as either conservative or diversified. Examples include bacterial persistence, within-clutch egg size variation in salmon, and delayed seed germination in annual plants. Previous theory has relied on deterministic models, and has found that bet hedging is adaptive only when the geometric mean fitness of the bet hedger is greater than that of the specialist wild-type. While a few models have explicitly incorporated stochasticity in the environmental regime, the role of demographic stochasticity, or genetic drift, has been entirely overlooked in models of bet hedger evolution. We utilize a novel stochastic framework to simultaneously capture the effects of stochasticity in the environment and in reproductive output. Consistent with deterministic expectation, we find that when the geometric mean fitness of the bet hedger is greater than that of the resident wild-type specialist, the bet hedger is positively selected. However, we find that even if the geometric mean fitness of the bet hedger is modestly less than that of the resident wild-type, bet hedging can still be beneficial. More specifically, bet hedging is favored at large population sizes but disfavored at small population sizes. This phenomenon, where the sign of selection is dependent on population size, is known as sign inversion. Sign inversion has been previously observed in modifiers of mutation rate, recombination, and others, but this research is the first finding that bet hedging also exhibits sign inversion. Both conservative and diversified bet hedging strategies exhibit sign inversion in these regimes. Bet hedging only exhibits sign inversion when both stochasticity in reproduction and environment are considered; the essential interaction between these sources of stochasticity drives the probability of fixation above the neutral expectation. Our results demonstrate

that bet hedging strategies can be adaptive in a larger range of environmental conditions than previously thought.

22 Deciphering the mystery of sorghum tannin domestication in Africa: coevolution among sorghum, human, and bird Yueyu Wu[1], Tingting Guo[2], Qi Mu[2], Jinyu Wang[2], Xin Li[2], Yun Wu[2], Bin Tian[1], Ming Li Wang[3], Guihua Bai[3], Ramasamy Perumal[1], Harold Trick[1], Scott Bean[3], Ismail Dweikat[4], Mitchell Tuinstra[5], Geoffrey Morris[1], Tesfaye Tesso[1], Jianming Yu[2], *Xianran Li*[2,3] 1) Kansas State University; 2) Iowa State University; 3) USDA-ARS; 4) University of Nebraska, Lincoln; 5) Purdue University

A crucial step of domesticating cereal crops is reducing bitter substances in the edible seeds. After domestication, condensed tannins, potent secondary metabolites inducing an unpleasant taste perception, are purged from seeds of major cereals, except sorghum. Half of sorghum cultivars (domesticated in Africa) have condensed tannings in grains, which presents an evolutionary mystery. With a serendipity field observation in Iowa, we parsed the harmonic relationships among the domesticate-triad in the evolution of African agroecosystems at the continental scale.

Started from the observation of sparrows' feed preference on sorghum seeds from a mapping population, we uncovered that condensed tannins deterred sparrows and identified two interacting genes (*Tannin1* and *Tannin2*) controlling for tannin presence. Geographic distributions of the functional polymorphisms suggested that smallholders in different African regions independently domesticated non-tannin sorghum for palatable food like other cereals.

Tannin sorghum cultivars are predominantly grown in East & South Africa, while non-tannin in West Africa. The parallel geographic distribution of the world's most abundant wild bird, *Quelea quelea*, to tannin sorghum distribution supported that *Q. quelea* is the key biotic factor driving this differential at the continental scale. Smallholders in East & South Africa selected tannins to fight against the severe herbivore threats.

Using condensed tannins to fight against biotic threats is a smart strategy, however, how African smallholders can take "bad taste" tannin sorghum as a staple food needed to be reconciled. As the bad taste is induced by bitter taster receptors (TAS2Rs), mutations in TAS2Rs may lead varied sensitivity for condensed tannins. Analyzing African genomes uncovered a parallel geographic distribution of a non-synonymous SNP in TAS2R5 to the tannin sorghum distribution. The ancestral non-taste allele is predominantly observed in East & South Africa, while the derived taste allele sensitive to tannins is observed mainly from West Africa.

Three geographic distributions of domesticator, domesticate, and environment depict harmonic interactions in African agroecosystems. In East & South Africa, smallholders (perceiving less bitterness from tannins because of non-taster TAS2R5) selected tannin to fight against the pest *Q. quelea*; in West Africa with mild threats from *Q. quelea*, smallholders (carrying taster TAS2R5) selected non-tannin sorghum for better food.

Reference: Allelochemicals targeted to balance competing selections in African agroecosystems (https://doi.org/10.1038/s41477-019-0563-0)

23 The Population Genetics of Convergent Adaptation in Maize and Teosinte *Silas Tittes*[1,2], Anne Lorant[3], Sean McGinty[3], John Doebley[4], James Holland[5,6], Jose Sánchez-González[7], Arun Seetharam[8,9], Maud Tenaillon[10], Jeffrey Ross-Ibarra[1,2,11] 1) University of California, Davis, CA; 2) Center for Population Biology, University of California, Davis, CA; 3) Department of Plant Sciences, University of California, Davis, CA; 4) Laboratory of Genetics, University of Wisconsin – Madison, Madison, WI; 5) United States Department of Agriculture – Agriculture Research Service, Raleigh, NC; 6) Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC; 7) Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco; 8) Department of Ecology, Evolution, and Organismal Biology, Ames, IA; 9) Genome Informatics Facility, Iowa State University, Ames, IA; 10) Génétique Quantitative et Evolution - Le Moulon, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Gif-sur-Yvette, France; 11) Genome Center, University of California, Davis, CA

Evidence for adaptation is ubiquitous, but the geographic scale that it operates over and the forces that moderate the scale are less understood and documented. To study this, we used whole genome sequencing data from sympatric population pairs of cultivated maize and its wild relative teosinte (*Zea mays* subsp. *parviglumis*) sampled at six locations across the species' native range. Despite population structure within and between subspecies that could limit the spread of beneficial alleles, our findings suggest that adaptations (as quantified by proportion of mutations fixed by natural selection and signatures of selective sweeps), are frequently shared between multiple populations of maize and teosinte. We found that adaptation is facilitated by the rapid human-assisted migration of cultivated maize, allowing for the exchange

of beneficial alleles across the range and between subspecies. Together, our results provide evidence that the geographic scale of adaptation in Zea mays is unexpectedly relatively large when compared to the size of populations.

24 Identifying the genetic basis of convergent local adaptation in conifers *Tom Booker*[1], Pooja Singh[2], Brandon Lind[1], Mengmeng Lu[2], Sally Aitken[1], Michael Whitlock[1], Sam Yeaman[2] 1) University of British Columbia; 2) University of Calgary

Studying local adaptation provides a window into the process of evolution, yielding insights about the nature of evolvability. Conifers are a great model for studying convergent evolution as they often inhabit large spatial ranges that encompass wide environmental heterogeneity to which they exhibit strong local adaptation. If the same or similar genomic regions were repeatedly involved in local adaptation in distinct conifer lineages, it would inform us about constraints to diversification. A potentially powerful method to identify locally adaptive genetic variation is the genotype environmental association (GEA) study. In GEA studies, correlations between allele frequency and environmental variables that are presumed to reflect selection gradients are examined for many markers across the genome. In this presentation, I will describe a comparative population genomics analysis of conifers from both North America and Europe. In our study, we apply novel GEA methods and tests for convergent evolution to data from 7 species, representing roughly 200 million years of evolution. Our novel methods allow us to quantify the extent of convergent evolution among pairs of species and to identify gene orthologs with evidence of convergent evolution over deep time. I will describe results from our analyses and discuss what we can infer from them.

25 Evolutionary consequences of domestication on the selective effect of new mutations in canids *Carlos Eduardo Guerra Amorim*[1], Clare Marsden[2], Jonathan Mah[2], Miguel Guardado[3], Bernard Kim[4], Jacqueline Robinson[3], Robert Wayne[2], Kirk Lohmueller[2] 1) California State University Northridge; 2) University of California Los Angeles; 3) University of California San Francisco; 4) Stanford University

Domestication created radical phenotypic changes in many species and understanding the genetic basis of these changes is a major research objective. Recent studies suggest that the domestication of canids has increased the load of deleterious mutations, simplified the genetic architecture of complex traits, and increased runs of homozygosity in the dog genome. What remains unclear is whether domestication has also altered the distribution of fitness effects (DFE) of new mutations due to shifted selective pressures. Specifically, strongly deleterious mutations may have become less deleterious in domesticated populations living alongside humans, while neutral mutations underlying traits of interest may have been selected by breeders. We address this question by leveraging whole-genome sequence data from 23 grey wolves and 61 dogs. We observe nonsynonymous to synonymous ratios ranging from 0.61-0.71 across the different populations of wolves and dogs. After accounting for differences in demography and background selection, we find that the DFE is similar across canids, with ~36% of new amino acid-changing mutations being neutral ($s<0.0001$), and ~43% under strong purifying selection ($s>0.01$). We evaluate the robustness of our results to different model assumptions and conclude that the DFE is stable across short evolutionary timescales, despite putative drastic changes in the selective pressure caused by artificial selection during domestication and breed formation. On par with previous works describing DFE evolution across the tree of life, we hypothesize that the DFE depends more strongly on organismal characteristics and less so on shifting selective pressures.

26 The genetics of epigenetics Magnus Nordborg Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria

Epigenetics continues to fascinate, especially the notion that it blurs the line between “nature and nurture” and could make Lamarckian adaptation via the inheritance of acquired characteristics possible. That this is in principle possible is clear: in the model plant *Arabidopsis thaliana*, experimentally induced DNA methylation variation can be inherited and affect important traits. The question is whether this is important in nature. Recent studies of *A. thaliana* have revealed a pattern of correlation between levels of methylation, climate variables, and recent transposon activity that is consistent with methylation playing a role in adaptation. Unexpectedly, it has also become clear that much of this variation has a genetic basis, suggesting the methylation is more of a phenotype than an epigenetic inheritance system. I will describe currently knowledge of this, based on research from my own group and others.

27 Mating-related barriers to admixture shape ancestry patterns across the baboon genome *Arielle Fogel*[1], Tauras Vilgalys[1,2], Elizabeth Archie[3], Susan Alberts[1], Jenny Tung[1] 1) Duke University, Durham, NC; 2) University of Chicago, Chicago, IL; 3) University of Notre Dame, Notre Dame, IN

Admixture is a common feature across the tree of life. Nonetheless, many species that hybridize remain phenotypically and genetically distinct, even when natural hybrids are both viable and fertile. To understand what maintains taxonomic

integrity in the face of gene flow, we combined genome-wide local ancestry analysis with five decades of field data from naturally hybridizing wild baboons to investigate barriers to gene flow at the genetic and behavioral levels. Study subjects (n=442) are members of a hybrid baboon (*Papio cynocephalus* x *P. anubis*) population in which signatures of selection against introgression resemble those observed for hominins. Thus, they also serve as "living models" for understanding admixture dynamics in human history, including the role of behavior, which does not leave a trace in the fossil record.

To identify barriers to admixture at the genetic level, we found pairs of physically unlinked, ancestry-informative markers in which two-locus genotypes of the same ancestry were enriched in our sample (1,185 locus pairs corresponding to 1,276 unique loci; the most extreme 99.99% tested). Such cases produce a positive correlation in ancestry states across loci, consistent with the removal of mismatched ancestry states (i.e., genotype combinations from different species) due to assortative mating by ancestry or intrinsic or extrinsic hybrid incompatibilities.

Within this set of unlinked, ancestry-correlated sites, we found 92 locus pairs (131 unique loci) at which baboons with matched homozygous ancestry were more likely to mate than those with mismatched ancestry, controlling for sociodemographic factors known to influence baboon mating behavior (10% FDR; 11,597 potential mating pairs; 140 males, 103 females). Baboons mated more assortatively based on overall ancestry at these 131 loci than they did based on ancestry-associated morphological traits (ΔAIC=14.5) or ancestry genome-wide (ΔAIC=13.3). Introgressed ancestry is depleted near assortative mating-related loci compared to the genomic background, consistent with a role in restricting gene flow (mean=0.296 vs. 0.353 ± 0.008 s.d. for random markers). Assortative mating-related loci also fall near genes that are weakly enriched for nervous system traits (FDR=0.13) and expression (FDR=0.12), suggesting a role in behavioral modulation. Together, our results point to novel candidate loci involved in complex mating behavior in the wild. They emphasize a key role for behavior in shaping ancestry patterns along primate genomes, suggesting that behavioral barriers may also have played a role in hominin admixture.

28 Hybrid fitness effects modify fixation probabilities of introgressed alleles *Aaron Pfennig*, Joseph Lachance Georgia Institute of Technology, Atlanta, GA

Hybridization is a common occurrence in natural populations, and introgression is a major source of genetic variation. Despite the evolutionary importance of adaptive introgression, classical population genetics theory does not take into account hybrid fitness effects (HFEs). Specifically, heterosis (i.e., hybrid vigor) and Dobzhansky-Muller incompatibilities (DMIs) influence the fates of introgressed alleles. Here, we explicitly account for polygenic, unlinked HFEs when tracking an introgressed marker allele. These HFEs quickly decay over time due to repeated backcrossing, enabling a separation-of-timescales approach. Using branching process and diffusion theory in combination with computer simulations, we formalize the intuition behind how HFEs affect introgressed alleles. We find that HFEs can either hinder or boost the fixation probability of introgressed alleles, depending on the relative strength of heterosis and DMIs effects. We show that the inclusion of a correction factor (α, representing the compounded effects of HFEs over time) into classic population genetics theory yields accurate fixation probabilities. Despite this, HFEs only subtly change the distribution of fitness effects of introgressed alleles that reach fixation. Although strong DMI effects may expedite the loss of introgressed alleles, fixation times are largely unchanged by HFEs.

29 Asymmetric introgression between selfer and outcrosser subspecies of *Clarkia xantiana* across a zone of secondary contact *Shelley Sianta*, David Moeller, Yaniv Brandvain University of Minnesota

Secondary contact of taxa that have diverged in allopatry has historically been touted as the final stage of speciation, deciding the fate of incipient species. Ongoing work in the field of speciation genomics demonstrates that introgression upon secondary contact is common. However, it remains unclear whether differences in ecological, reproductive and genomic features of hybridizing taxa have repeatable effects on introgression. For example, shifts in mating system, from predominantly outcrossing to predominantly self-fertilizing, are predicted to affect both the magnitude and asymmetry of introgression. The few studies that have documented introgression between a selfer and an outcrosser have found asymmetric patterns, with more introgression occurring from the selfer to the outcrosser. Many of these studies, however, involve either sampling few sympatric populations and/or few individuals, making it unclear whether this pattern is generalizable. Moreover, it is unclear what mechanisms may govern this pattern. Here, we study two subspecies of *Clarkia xantiana* that have diverged in mating system in allopatry and that have recently come into secondary contact. Along their narrow zone of contact, we employ a high-density sampling scheme across four contact zone sites (n = 442 individuals). We test for the presence of introgression, asymmetry in the magnitude of introgression, and the extent of independence of introgression across the contact zone sites. Despite finding no evidence for early generation hybrids, we find introgression in all contact zone sites and a signal for independence of introgression among contact zone sites. While

we detect introgression in both subspecies within each contact zone site, there is strong asymmetry in introgression, with more introgression from the selfer to the outcrosser. Differences in track lengths of introgressed DNA suggest that introgression from the outcrosser to the selfer is either more recent and/or selection against introgressed DNA is more efficient because of reduced effective recombination in the selfer. Lastly, we test the hypothesis that differences in the architecture of genetic load between the outcrosser and selfer contribute to patterns of introgression within each subspecies. This study highlights the predictability of mating system shifts on patterns of introgression upon secondary contact.

30 The spatiotemporal patterns of major human admixture events during the European Holocene *Manjusha Chintalapati*[1], Nick Patterson[2], Priya Moorjani[1] 1) Department of Genetics, University of California Berkeley; 2) Broad Institute of Harvard and MIT, Cambridge,

Recent studies have shown that gene flow or admixture has been pervasive throughout human history. While several methods exist for dating admixture in contemporary populations, they are not suitable for sparse, low coverage data available from ancient specimens. To overcome this limitation, we developed *DATES* that leverages ancestry covariance patterns across the genome of a single individual to infer the timing of admixture. By performing simulations, we show that *DATES* provides reliable results under a range of demographic scenarios and outperforms available methods for ancient DNA applications. We apply *DATES* to ~1,100 ancient genomes to reconstruct gene flow events during the European Holocene. Present-day Europeans derive ancestry from three distinct groups, local Mesolithic hunter-gatherers, Anatolian farmers, and Yamnaya Steppe pastoralists. These ancestral groups were themselves admixed. By studying the genetic formation of Anatolian farmers, we infer that the gene flow related to Iranian Neolithic farmers occurred before 9,600 BCE, predating the advent of agriculture in Anatolia. We estimate the early Steppe pastoralist groups—Yamnaya and Afanasievo—were genetically formed more than a millennium before the start of steppe pastoralism, contrary to the archaeological evidence. Using ancient genomes across sixteen regions in Europe, we provide a fine-scale chronology of the Neolithization of Europe and the rapid spread of Steppe Pastoralist ancestry across Europe. We confirm previously discovered signal of the resurgence of hunter-gatherer ancestry during the Neolithic period and provide new details about the spread of Corded Ware and Bell Beaker Complexes across Europe. Our analyses provide new insights about the origin and spread of farming and Indo-European languages, highlighting the power of genomic dating methods to elucidate the legacy of human migrations.

31 Population genomics of an entire community of Galápagos finches *Erik Enbody*[1,2], C. Grace Sprehn[1], Ashley T. Sendell-Pierce1[1], Carl-Johan Rubin[1], B. Rosemary Grant[3], Peter Grant[3], Leif Andersson[1,4,5] 1) Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; 2) Department of Biomolecular Engineering, University of California Santa Cruz, CA, USA; 3) Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA; 4) Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, USA; 5) Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

Molecular data provides a powerful opportunity to study the historical processes that have led to speciation and the evolution of phenotypic diversity, but rarely is it linked to contemporary observations of population change. Here, we link long-term trends in natural selection and hybridization in an iconic adaptive radiation, the Galápagos finches, to their molecular drivers by whole-genome sequencing of an entire island community. Specifically, we sequenced every finch alive on Daphne Island in the Galápagos Islands between the years 1983 and 2012 (*Geospiza,* n = 3,957). We show that hybridization has had dramatic effects on the genetic ancestry of three species, leading to large shifts in population-wide ancestry coefficients over the course of the study period. We identify six loci of large phenotypic effect that together explain as much as 70% of beak morphology variation; a key ecological trait in this group that has facilitated speciation. We next use individual-level haplotype data to track the frequency of causal alleles through time. In doing so, we show that allele frequency shifts in *G. fortis* at loci contributing to beak size track a decline in beak size over time in this species. We next show evidence for introgression of these loci leaving lasting footprints on morphology over the time period and discuss evidence for introgression between two species via a third conduit species. Our community-based investigation identifies several important genetic loci governing beak traits that have been subject to natural selection. They are transferred between species through introgressive hybridization and may reflect a general mechanism for how gene exchange is contributing to biodiversity.

32 Demographic History Inference and the Polyploid Continuum Paul Blischak, Mathews Sajan, Michael Barker, *Ryan Gutenkunst* University of Arizona, Tucson, AZ

Polyploidy is an important generator of evolutionary novelty, occurring within a single lineage (autopolyploidy) or by hybridization between different lineages (allopolyploidy). Historically, these scenarios have been treated as completely separate cases based on chromosome pairing, with autopolyploids pairing among all homologous chromosomes and

allopolyploids pairing only among chromosomes from the same parent lineage. But if they are similar enough, chromosomes from distinct lineages may pair and thus alleles may be exchanged between subgenomes. Quantitatively inferring demographic history and rates of subgenomic exchange within polyploids may shed light on the evolutionary history of key species, including many crops. We thus developed diffusion models for genetic variation in polyploids with variable inheritance patterns and implemented them in the dadi software. We used SLiM simulations to validate our models and test their inference properties, focusing on the case in which reads from two subgenomes cannot by distinguished bioinformatically. We then applied these models to infer demographic history and subgenome exchange for shepherd's purse (*Capsella bursa-pastoris*), a recently formed allotetraploid, to investigate the timing and impact of hybridization and genome duplication.

33 Replicate hybrid populations reveal shared genome evolution across multiple species *Quinn Langdon*[1,2], Daniel Powell[1,2], Molly Schumer[1,2,3] 1) Stanford University, Stanford, California; 2) Centro de Investigaciones Científicas de las Huastecas "Aguazarca", A.C., Calnali, Mexico; 3) Hanna H. Gray Fellow, Howard Hughes Medical Institutes, Chevy Chase, Maryland, United States of America

Hybridization is ubiquitous across the tree of life. With whole-genome sequencing we have revealed evidence for hybridization across scales; from ancient to recent and between closely and distantly related species. Our own species harbors genomic regions that are derived from hybridization. We are now entering an era where we can determine the mechanisms underlying hybrid genome stabilization and how these are shared across hybridization events. A powerful system to address these questions is in the swordtail fish genus *Xiphophorus*, where replicate hybrid populations of two species pairs (*X. birchmanni* × *X. cortezi* and *X. birchmanni* × *X. malinche*) have allowed us to identify predictable factors driving genome stabilization in hybrids. We find high levels of repeatability in ancestry across the genomes of these populations despite their independent origins. This can be attributed in part to shared broad scale genomic organization. This includes a shared recombination landscape and shared locations of coding and conserved basepairs. With the discovery of additional hybrid populations between *X. birchmanni* × *X. cortezi*, we are beginning to disentangle the relative roles of different sources of selection in shaping ancestry across the genome. Our results highlight an important role for adaptively introgressed elements and hybrid incompatibilities that are shared across multiple species. Together these findings of parallel evolution after hybridization have important implications for mechanisms shaping hybrid genomes across the tree of life.

34 Molecular and morphological evolution across the most species-rich radiation in mammals *Gregg Thomas*[1,2], Carl Hutter[3], Ana Paula Assis[1], Emily Kopania[1], Sebastian Mortimer[1], Colin Callahan[1], Kevin Rowe[4], Jacob Esselstyn[3], Jeffrey Good[1] 1) University of Montana, Missoula, MT, USA; 2) Harvard University, Cambridge, MA, USA; 3) Louisiana State University, Baton Rouge, LA, USA; 4) Museum Victoria, Melbourne, VIC, AU

Adaptation and speciation ultimately depend on evolution at the molecular level, yet relatively little is known about the dynamics of molecular evolution during rapid species radiations. The Old World rats and mice (Murinae) represent one of the most species-rich radiations in mammals with extensive phenotypic and ecological variation. This diverse group is comprised of more than 630 species accounting for over 10% of all mammalian species, arising in only ~15 million years. Laboratory mice and rats are embedded within this massive radiation, providing a compelling opportunity to leverage these powerful model systems to understand the evolution of the extensive phenotypic and ecological variation found across this group. Here, we integrate comparative genomics with quantitative ecomorphology, biogeography, and life history data to understand the diversification of lineages, phenotypes, and genes across this speciose radiation. Using whole exomes from 209 species, we resolve their evolutionary history in the face of substantial phylogenetic discordance. We developed methods to accurately estimate substitution rates in the species tree while accounting for gene tree discordance and to heuristically prune large phylogenies to maximize concordance between the species tree and underlying gene trees. Using these methods, we identified lineages and genes that have undergone episodic positive selection during the evolution of rats and mice. We examined correlations between molecular evolution and the repeated colonization events in the Indo-Australian archipelago to test if signatures of convergent or adaptive molecular evolution coincide with bursts in phenotypic evolution. Finally, we map rates of cranial shape and body size evolution onto the history of the rodent radiation. This study is the largest comparative molecular evolution study of rodents to date and provides an opportunity to understand molecular evolution during rapid radiations in the context of two of the most important model organisms for basic and biomedical genetic research.

35 Testing for a role of parent-offspring conflict in the emergence of postzygotic barriers in *Mimulus* using a combination of genetic mapping and RNA sequencing analysis *Elen Oneal*[1], Miguel Flores-Vergara[2], Robert Franks[2], John Willis[1] 1) Duke University, Durham, NC; 2) North Carolina State University, Raleigh, NC

A fundamental constraint on organismal development and evolution – that natural selection does not lead to diminished individual viability and fertility – is encapsulated in the Dobzhansky-Muller model for the evolution of hybrid incompatibilities: substitutions of alleles that are adaptive or nearly neutral in the native genomic background of one species can be dysfunctional in the genomic background of other species, resulting in inviable and/or sterile hybrids. Evidence increasingly suggests that the fixation of incompatible alleles in divergent populations may be driven by unexpected forms of selection such as intra-genomic conflict. Parent–offspring conflict, where mothers and offspring spar over evolutionary time over the extent of maternal investment, is a pervasive feature of parent-offspring interactions, and is especially prominent when mothers mate multiply and maternal investment is substantial and allocated dynamically via a placenta-like connection between mother and developing offspring. Such conflict is thought to explain, in part, the evolution of genomic imprinting, in which alleles are differentially expressed depending upon their parent-of-origin, and which has only been detected thus far in the placental tissues and developing offspring of viviparous mammals and the seed endosperm of angiosperms. Intriguingly, misregulation of endosperm development is a prominent cause of hybrid seed failure, the most common form of postzygotic isolation in plants. Is parent-offspring conflict a major factor driving plant speciation? Unfortunately, explicitly testing this hypothesis genetically is often difficult because F1 hybrids are dead. We are able to circumvent the endosperm barrier in *Mimulus* species by rescuing F1 embryos prior to seed death. We use these F1s in a reciprocal backcross design to map loci contributing to hybrid seed inviability and endosperm failure and test for paternal and maternal interactions between loci. We combine our mapping results with RNA sequencing of endosperm tissue to characterize species divergence in imprinting status and regulation. Our results shed light on the role of parent-offspring conflict in driving imprinting divergence and the evolution of this common postzygotic barrier in *Mimulus*.

36 When and how is introgression adaptive? A tale from two widely distributed sympatric oak species *Ruirui Fu*[1], Yuxiang Zhu[1], Ying Liu[1], Yu Feng[2], Rui-Sen Lu[3], Yao Li[4], Pan Li[1], Antoine Kremer[5], Martin Lascoux[6], Jun Chen[1] 1) College of Life Sciences, Zhejiang University, China; 2) Chengdu Institute of Biology, Chinese Academy of Sciences, China; 3) Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, China; 4) College of Biology and the Environment, Nanjing Forestry University, China; 5) UMR BIOGECO, INRAE, Université de Bordeaux,France; 6) Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Sweden

Introgression is a crucial source of new alleles, perhaps even more important than standing variation, for adaption under rapidly changing environment. Though introgression has been extensively studied in many plants and animals, key questions on the underlying mechanisms of introgression still remain unanswered. In particular, what is the genomic distribution of introgressed regions along the genome, are the level and pattern of introgression influenced by ecological factors, and if so when and how does introgression contribute to adaptation? By investigating introgression between two sympatric widespread Asian oak species, Q. acutissima and Q. variabilis sampled in multiple forests and for which we have generated high quality genomic resources, we show that introgressed regions are broadly distributed along the genome. Introgression depends on genetic divergence between pairs of populations and on the similarity of the environments in which they live. Oak populations occupying similar ecological sites tend to share the same introgressed regions and introgressed genomic footprints of adaptation are preferentially located in regions with suppressed recombination rate. Introgression confers adaptation by introducing allelic variation in cis-regulatory elements, in particular through TE insertions, thereby altering the regulation of genes related to stress. Our results provide new avenues of research for uncovering the mechanism of adaptation in sympatric species.

37 Predicting evolutionary divergence and parameters of relocated genes from their expression data *Antara Anika Piya*, Raquel Assis Florida Atlantic University, Boca Raton, FL

Relocation is a large-scale mutation event that places genes in new genomic locations and chromatin environments. Consequently, a relocated gene may incur novel expression patterns and such changes are widely hypothesized to drive evolutionary divergence and speciation. Here, we design the first methods for predicting evolutionary divergence and its underlying parameters from expression data of relocated genes. In particular, our methods utilize random forests, support vector machines, and multi-layer neural networks built on a model of gene expression evolution. Application of our methods to simulated data shows that, whereas the globally best performance is achieved with a neural network, all architectures have high power and accuracy in predicting evolutionary divergence and its underlying parameters across a diversity of evolutionary scenarios. Further, application of our methods to empirical data from *Drosophila* shows that evolutionary divergence occurs in 15-20% of relocated genes, is driven by both neutral and selective forces, and tends to affect genes involved in development. Thus, by providing a mechanism for assaying whether and how evolutionary divergence occurs after gene relocations, our suite of machine learning methods represents a major advancement in studying the roles of gene relocations in the origins of novel functions, phenotypes, and species.

38 Critical role of insertion preference for the invasion trajectory of transposons *Manisha Munasinghe*, Nathan Springer, Yaniv Brandvain University of Minnesota

Transposable elements (TEs) are mobile repetitive DNA sequences that have been highly successful at invading eukaryotic genomes. TEs can be divided into classes, subclasses, superfamilies, and families depending on their replicative mechanism, ancestral origin, and sequence similarity highlighting the genetic and mechanistic diversity of TEs. The proportion of the genome occupied by TEs varies greatly between species, as well as how that proportion is distributed between the present TE families. Some TE families contribute relatively little to this overall proportion, with only a handful of copies present in the genome, while other families contain tens of thousands of copies. It is unclear how TE families reach a high copy number given the expectation that novel insertions be deleterious. Previous population genetic models have largely ignored the tendency of TE families to preferentially insert into specific DNA sequences or features. This insertion preference may not only be the result of structural differences between TEs but also an evolved trait that dictates an underlying distribution of fitness effects for each new insertion. It has been hypothesized that TE families may evolve a preference for neutral insertion sites to minimize their cumulative deleterious load and avoid driving the host population extinct. To test this hypothesis, we use a non-Wright-Fisher framework in SLiM 3 to explore how transposition probability and insertion preference influence the evolution of mean TE copy number and host population size. We consider a diploid population that gains a single copy of a TE in the genome of a single individual (analogous to horizontal transfer). This TE belongs to a unique family with an assigned transposition probability and range of fitness effects for novel insertions that represent insertion preference. We can track the spread of the TE family through the population by measuring its mean copy number and population frequency over time, and we allow population size to fluctuate depending on the fitness of individuals, such that populations can go extinct. Using our model, we find that population extinction occurs most often under high transposition probabilities, but, as we reduce transposition, we find that extinction persists when the neutral insertion preference is high, rejecting our initial hypothesis. Somewhat counterintuitively, a preference for neutral insertion sites allows the mean copy number of both neutral and deleterious insertions to grow most rapidly, consequently, leading to extinction. Our results show insertion preference is not protective on its own and suggest mechanisms that regulate TE transposition are key to obtaining high TE copy numbers.

39 Chromosomal inversions shape the genomic landscape of the deer mouse *Olivia Harringmeyer*, Hopi Hoekstra Harvard University

Chromosomal inversions are an important type of structural variation, with increasing evidence for their role in local adaptation across diverse species. However, since inversions can be challenging to detect, the prevalence and hence significance of inversions segregating within species remains largely unknown, especially in natural populations of mammals. Here, by combining population-genomic and long-read sequencing analyses in a single, widespread species of deer mouse (*Peromyscus maniculatus*), we identified 25 novel polymorphic inversions and explored their role in adaptation. Overall, these inversions are large, ranging in size from 1.5-43 Mb, and in total, affect 17% of the *P. maniculatus* genome. By localizing the inversion breakpoints, we found that genomic regions harboring breakpoints are enriched for long inverted repeats (0.5-50 kb), suggesting that these inversions likely arose via ectopic recombination. We then genotyped the inversions in 13 populations across the species range and found that the inversions are highly polymorphic, not only within the species, but also within populations, such that the inversions are frequently heterozygous. Because we found that the inversions cause near complete suppression of recombination in heterozygotes (genome average: 0.81 cM/Mb; heterozygous inversion: 0.01 cM/Mb), the inversions have a strong effect on recombination rates and patterns of linkage disequilibrium in natural populations. In addition, the inversions have accumulated significant levels of nucleotide divergence (0.2-0.9%) from the ancestral haplotypes, as a consequence of suppressed recombination. Finally, we found that the inversions substantially contribute to differentiation between locally adapted populations: for example, a closely-related pair of populations adapted to forest versus prairie habitats differed by >50% in their frequencies of 13 inversions despite ongoing gene flow. One inversion in particular plays a critical role in maintaining multiple adaptive trait differences between forest-prairie ecotypes, both by carrying adaptive mutations (for pigmentation and tail length) as well as enabling their co-inheritance through suppressed recombination. Together, we find that inversion polymorphisms have a significant impact on recombination, genome structure and genetic diversity in deer mice, and are possibly a key source of genetic variation facilitating local adaptation across this species' widespread range.

40 Recombination patterns in corn snakes suggest a tug of war between PRDM9 and promoter-like features *Carla Hoge*[1], Marc De Manuel[1], Zachary Fuller[1], Zachary Baker[1, 2], Izabel Cavassim[1, 3], Shreya Banerjee[4], Molly Schumer[4], Athanasia Tzika[5], Molly Przeworski[1] 1) Columbia University, New York, NY; 2) University of Cambridge, Cambridge, United Kingdom; 3) University of California, Los Angeles, CA; 4) Stanford University, Stanford, CA; 5) University of Geneva, Geneva, Switzerland

Comparisons among model organisms make clear that, despite the fundamental importance of recombination in sexually-reproducing species, the mechanisms by which it is directed to the genome can vary markedly. Notably, in mammals such as humans and mice, recombination occurs at the binding sites of the protein PRDM9 and the binding affinity of PRDM9 is rapidly evolving. In other species such as birds or canids, PRDM9 has been lost and recombination occurs preferentially at promoter-like features, such as CpG islands, through unknown mechanisms. Our previous work showed that PRDM9 arose before the origin of vertebrates and is rapidly evolving where the PRDM9 ortholog is intact. This finding led us to postulate that some species outside of mammals also use PRDM9 to direct recombination. To test this prediction, we focused on the corn snake *Pantherophis guttatus*, a species with a complete PRDM9 ortholog that is rapidly evolving. We improved the assembly and annotation of the reference genome and resequenced 22 unrelated corn snake samples to high coverage in order to infer historical recombination rates across the genome from patterns of linkage disequilibrium. In contrast to what is seen in mammals with PRDM9, we found evidence for elevated recombination around computationally predicted PRDM9 binding sites but also near CpG islands. To verify these findings, we resequenced two families with five offspring, identified the PRDM9 alleles segregating in the families and called crossover events that occurred in the parents. This analysis confirmed that crossover events overlap both PRDM9 binding sites and CpG islands more than expected by chance. These findings indicate that corn snakes behave neither like mammalian species with PRDM9 nor like species that lack the gene. We are now testing the possibility that they use a mixture of features to direct recombination because of changes in interacting genes that recruit the recombination machinery to sites of PRDM9 binding.

41 Genomic and Epigenomic Insights into Formation and Evolution of Polyploid Plants and Crops Z. Jeffrey Chen
The University of Texas at Austin

Polyploidy or whole genome duplication (WGD) is a prominent evolutionary feature for all flowering plants and many animals. Most crop plants including wheat, cotton, and oilseed rape are polyploids, and many other crops such as maize and soybean are ancient polyploids. In allopolyploids, interspecific hybridization can induce “genome shock,” leading to genomic and epigenomic changes, the effects of which can be amplified by genome doubling (ploidy changes). Moreover, heterozygosity between hybridizing parents is permanently fixed in allopolyploids. However, many polyploid plants and crops were formed long time ago, and their exact progenitors may become extinct. To replay the evolutionary type, we resynthesized *Arabidopsis suecica*allotetraploids between *A. thaliana* and *A. arenosa* and compared them with natural *A. suecica* that was formed ~300,000 years ago. Using the trackable *Arabidopsis* allotetraploid system and allotetraploid cotton crop model, we employed integrated approaches of genomics, epigenomics, and computational biology to investigate genomic variation, gene expression changes, and epigenomic modifications that have shaped phenotypic diversification in plant polyploids. Here, I will update recent findings and perspectives of comparative genomics and epigenomics that have uncovered the molecular bases for nonadditive gene expression, altered genetic recombination, morphological diversity, and adaptive evolution in *Arabidopsis* and cotton allopolyploids. These conceptual advances and genomic and epigenomic resources will help us better understand polyploid genome evolution and improve the production of polyploid crops for food security and carbon capture and storage.

42 Synergistic epistasis of the deleterious effects of transposable elements Grace Lee Department of Ecology and Evolutionary Biology & Center for Complex Biological Systems, University of California, Irvine

Transposable elements (TEs) are widespread genetic parasites that copy and insert themselves across host genomes. The replicative nature and generally deleterious effects of TEs raise an outstanding question about how TE copy number is stably contained in host populations. Classic theoretical analyses predicted that, when the decline in fitness due to each additional TE insertion is greater than linear, or when there is synergistic epistasis, selection against TEs can result in a stable equilibrium of TE copy number. While several mechanisms have been predicted to yield synergistic deleterious effects of TEs, empirical studies of the presence of such epistatic interactions are still lacking. To investigate the presence and prevalence of synergistic epistasis among TEs, we leveraged the fact that purifying selection with synergistic epistasis generates repulsion linkage between deleterious alleles. We investigated this population genetic signal in the likely ancestral *Drosophila melanogaster* population and found evidence supporting the presence of synergistic epistasis among TE insertions, especially TEs expected to exert large fitness impacts. Even though synergistic epistasis of TEs has been predicted to arise through ectopic recombination and TE-mediated epigenetic silencing mechanisms, we only found mixed support for the associated predictions. We observed signals of synergistic epistasis for a large number of TE families, which is consistent with the expectation that such epistatic interaction mainly happens among copies of the same family. Curiously, significant repulsion linkage was also found among TE insertions from different families, suggesting the possibility that synergism of TEs’ deleterious fitness effects could arise above the family level and through mechanisms similar to those of simple mutations. Our findings set the stage for investigating the prevalence and importance of epistatic interactions in the evolutionary dynamics of TEs.

43 Tracing the evolutionary dynamics of gene retrocopies in house mouse natural populations *Wenyu Zhang*, Diethard Tautz Max Planck Institute for Evolutionary Biology

While the contribution of retrogenes (new genes originated from the reintegration of reverse-transcribed RNA into the genome) to genome evolution and adaptations has long been recognized, the evolutionary patterns of very recently derived gene retrocopies that are still polymorphic within natural populations have not been much studied so far. To trace the evolutionary dynamics of these new gene retrocopies, we use a unique genomic dataset from nine house mouse populations and show the biological effects of these new gene retrocopies from two distinct aspects. In the first study, we show for natural house mouse populations that the primary rate of gene retroposition is orders of magnitude higher than the long-term rate. Comparisons with single-nucleotide polymorphism distribution patterns in the same populations show that most new gene retrocopies are deleterious. Transcriptomic profiling analysis shows that new gene retrocopies become easily subject to transcription and have an influence on the expression levels of their parental genes, especially when transcribed in the antisense direction. While most gene retrocopies are detrimental and quickly purged, we find a subset of them that appears neutral or even adaptive in the second study. We show that gene retrocopies from X-chromosomal parental genes have a higher likelihood of reaching elevated frequencies in the populations, confirming the notion of adaptive effects for "out-of-X" retrogenes. Also, retrocopies in intergenic regions are more likely to achieve higher population frequencies than introns of genes, implying a more detrimental effect when they land within transcribed regions. For a small subset of retrocopies, we find signatures of positive selection, indicating they were involved in a recent adaptation process of these house mouse natural populations. Taken together, these two lines of work represent a comprehensive coverage of the evolutionary dynamics of gene retrocopies in a range of natural populations.

44 Epistatic drift causes gradual loss of predictability in molecular evolution *Yeonwoo Park*, Brian Metzger, Joseph Thornton University of Chicago

Epistatic interactions have the potential to make the course and outcomes of evolution unpredictable, but no comprehensive data are available on the extent, direction, rate, and consequences of changes in the effects of mutations in any protein across evolutionary time. Here we characterize the temporal dynamics of epistasis by measuring changes in the effect of every possible point mutation along a densely sampled, 700-million-year phylogenetic trajectory of reconstructed and extant proteins in the steroid receptor DNA-binding domain. We found that pervasive epistatic interactions caused the functional effects of the majority of mutations to steadily drift from their initial effects and become completely or substantially contingent on the particular set of intervening substitutions that occurred during history. By statistically characterizing each mutation's effect as an evolving trait on the phylogeny, we found that epistatic change occurred at a rate that was largely constant across time but varied dramatically among mutations; this pervasive drift in the effects of mutations occurred gradually and without bias via many steps of small magnitude, suggesting a neutral evolutionary process and an underlying architecture of dense but weak epistatic interactions. By transiently opening and closing windows of evolutionary accessibility, substitutions that occurred under purifying selection shaped the historical fate not only of those mutations that fixed during history but also the far greater number that never did. Our findings indicate that the process by which protein sequences drift inevitably into contingency and unpredictability is itself predictable, given sufficient phylogenetic and experimental data.

45 Environmental adaptation in house mice: genetic and non-genetic effects on gene expression *Mallory A. Ballinger*[1], Katya L. Mack[2], Sylvia M. Durkin[1], Michael W. Nachman[1] 1) University of California, Berkeley; 2) Stanford University

A major goal in evolution biology is to identify the gene regulatory mechanisms underlying adaptive evolution. However, it is well known that gene regulation is highly plastic and dependent on the environment. Nevertheless, the genetic and non-genetic effects of gene regulation are rarely considered together in the context of adaptive evolution. Specifically, relatively little is known about the contributions of *cis*- and *trans*-regulation to gene expression plasticity. House mice (*Mus musculus domesticus*) provide an excellent opportunity to study the gene regulatory mechanisms of plasticity as they have recently expanded into a wide range of novel environments. Here, using RNA-seq data collected from liver and brown adipose tissue, we studied gene expression in mice reared in warm and cold environments in parents adapted to warm and cold environments and in their F1 hybrids. We identified strong patterns of expression divergence across environments, largely attributable to *cis*-regulatory changes. Patterns of expression plasticity were largely attributable to *trans*-effects as *trans*-effects showed greater sensitivity to temperature change. Finally, we identified genes for which there were significant effects of temperature on regulatory divergence, with genes exhibiting *cis*-by-environment and *trans*-by-environment effects. Among these genes, we identified two candidates - *Scd1* and *Cdh13* - that show patterns of adaptive plasticity and are regulated by *cis*-by-environment effects. These two genes play important roles in

adiposity and thermoregulation. Overall, these findings demonstrate the utility of allele-specific expression to identify regulatory mechanisms associated with plasticity and highlight the role the environment plays in adaptive gene regulation in house mice.

46 Using an Evolve + Resequencing experiment to estimate the strength of selection on candidate genes underlying local serpentine adaptation in Mimulus guttatus *Amelia Lawrence*[1], Jessica Selby[2], John Willis[1] 1) Duke University, Durham, NC; 2) Bayer - Crop Science, St. Louis, MO

Local adaptation occurs when spatially divergent selection at a locus is stronger than the rate of migration between habitats. Population genomic scans for loci contributing to local adaptation often identify tens or hundreds of candidate genes, but usually cannot estimate the strength of divergent selection at each locus and answer the key question: which loci are actually under the strongest selection, and therefore contribute the most to local adaptation? Serpentine outcrops are ideally suited for the study of local adaptation in plants because their patchy distribution results in a landscape with abrupt changes from "normal soils" to serpentine soils. Serpentine soils are unconducive to most plant life because of their characteristically low Ca:Mg ratio, a limited availability of macronutrients (K, P, N), and high levels of heavy metals (Ni, Zn). We have been studying parallel local serpentine adaptation of the wildflower Mimulus guttatus. Reciprocal transplant experiments in the field and greenhouse both reveal extreme local adaptation, where plants from non-serpentine populations are unable to survive and reproduce when transplanted onto serpentine soils. Population genomic scans of many pairs of adjacent serpentine and non-serpentine populations have identified dozens of candidate genes on all chromosomes that are under divergent selection. What are the relative fitness contributions of these genes to local adaptation? We marry a classic F2 mapping approach with an Evolve + Resequence experiment to estimate fitness effects of candidate loci. We cross a pair of well-characterized inbred lines, from adjacent serpentine and non-serpentine habitats, to form F2 progeny. We planted thousands of these hybrids on serpentine and control soils in the greenhouse and allowed them to die or survive and reproduce, using captive bumblebees to ensure outcrossing. By planting the seeds of survivors on their parental soils, we allow each population to adapt to its soil for several generations. We then use pool-sequencing of each population in each generation and apply likelihood test statistics to detect loci which responded to divergent selection and estimate marginal fitnesses of the alleles in each environment. Here I report the initial results from the first few generations.

47 Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation *Grant Kinsler*, Olivia Ghosh, Dmitri Petrov Stanford University

Building a genotype-phenotype-fitness map of adaptation is a central goal in evolutionary biology. Despite the importance of building such a map, it is difficult to do so even when adaptive mutations are known because it is hard to enumerate which phenotypes make these mutations adaptive. We address this problem by first quantifying how the fitness of hundreds of adaptive yeast mutants responds to subtle environmental shifts. We then model the number of phenotypes these mutations collectively influence by decomposing these patterns of fitness variation. We find that a small number of inferred phenotypes can predict fitness of the adaptive mutations near their original glucose-limited evolution condition. Importantly, inferred phenotypes that matter little to fitness at or near the evolution condition can matter strongly in distant environments. This suggests that adaptive mutations are locally modular — affecting a small number of phenotypes that matter to fitness in the environment where they evolved — yet globally pleiotropic — affecting additional phenotypes that may reduce or improve fitness in new environments. By isolating additional mutations from further evolution experiments and studying their fitness effects in subtle environmental perturbations both near and far from their original evolution condition, we further demonstrate that adaptive mutations indeed affect only a small number of phenotypes that contribute to fitness only in the environment in which they evolved but have diverse effects in distant environments. Using these approaches, we also show that further adaptive steps depend on the specific phenotypic effects of initial mutations, showcasing that directional selection alone can generate a large amount of latent phenotypic diversity with substantial effects on further adaptive evolution.

48 Investigating the genetic determinants of *L. monocytogenes* stress tolerance to food-industry relevant stressors through adaptive laboratory evolution *Tyler Bechtel*[1], John Gibbons[1,2,3] 1) Department of Food Science, University of Massachusetts, Amherst, MA; 2) Molecular and Cellular Biology Graduate Program, University of Massachusetts, Amherst, MA; 3) Organismic and Evolutionary Biology Graduate Program, University of Massachusetts, Amherst, MA

Listeria monocytogenes is a foodborne bacterial pathogen that causes thousands of illness and billions in medical and food recall costs annually in the United States. *L. monocytogenes* can survive, replicate, and persist in a variety of harsh environments, including conditions encountered in food processing facilities and equipment. This presents a major problem for public health and the food industry as stress resistant strains of *L. monocytogenes* are more likely

to persist in the food environment, and subsequently contaminate food products. Here, we devised an "evolve and resequence" experiment to shed light on the frequency, temporal patterns, and genetic determinants of stress resistance evolution in *L. monocytogenes*. Specifically, we grew three replicates of three phylogenetically and phenotypically distinct strains of *L. monocytogenes* for 500 generations in sub-lethal concentrations of (i) salt-induced osmotic stress, (ii) benzalkonium chloride (BAC), a sanitizer commonly used in the food industry, and (iii) tryptic soy broth (control). We have sequenced and annotated the genomes of each of the ancestral strains. Every 50 generations, we have (i) frozen and stored cultures, (ii) measured the minimum inhibitory concentration (MIC) of each stressor to track temporal changes in fitness, (iii) conducted a quantitative virulence assay to investigate the potential relationship between stress response and pathogenicity, and (iv) resequenced the genomes of each lineage to track changes in allele frequency and to identify candidate mutations underlying adaptive phenotypes. Lastly, we will create gene deletion mutants using the suicide plasmid pHoss1 to validate the function of candidate genes. Our results indicate that BAC-sensitive isolates, ALE_10_0415 and ALE_20_0415, exhibited a two-fold increase in MIC of BAC after ~300 generations. In comparison, BAC-tolerant isolate ALE_16_0415 showed minimal improvement at ~400 generations, which suggests that there is a physiological limit to *L. monocytogenes* BAC resistance. Our results will provide insight into the evolution and genetic mechanisms underlying *L. monocytogenes* stress-tolerance. In addition, the results of this study will identify candidate genes that will enable bacterial surveillance systems to detect and characterize persistent *L. monocytogenes* strains isolated from food production facilities.

49 Pleiotropy is associated with the parallel gene expression evolution during adaptation to a novel environment *Wei-Yun Lai*[1,2], Sheng-Kai Hsu[1,2], Andreas Futschik[3], Christian Schlötterer [1] 1) Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria.; 2) Vienna Graduate School of Population Genetics, Vienna, Austria.; 3) Department of Applied Statistics, Johannes Kepler University Linz, Linz, Austria.

Repeatable genomic/phenotypic changes across replicated population pairs provide evolutionary biologists with strong evidence for adaptive evolution. Nevertheless, adaptation to a similar environment often yields inconsistent outcomes, particularly at the genic level. The evolutionary response can range from a subset of adaptive genes evolving in a highly repeatable manner to highly heterogeneous evolutionary responses in replicate experiments. Understanding the potential factors affecting the extent of parallelism in gene evolution is an important research question as this provides an important step towards the prediction of evolution. The degree of pleiotropy has been proposed as a potential factor affecting the degree of evolutionary parallelism/repeatability of a gene. Although several theoretical predictions have been made, their empirical validation is still limited. To fill this gap, we studied the evolution of gene expression in 10 replicated *Drosophila simulans* populations adapting to the same new temperature regime. Comparing the parallelism of expression evolution among putatively selected genes, we observed a positive relationship between the degree of evolutionary parallelism and the pleiotropic effects. In addition, the degree of evolutionary parallelism was negatively associated with the magnitude of natural variation in the expression of the genes. Because gene expression variation was also associated with the extent of the pleiotropic effect, we performed causal analysis to disentangle the causal influence of natural variation and pleiotropic effects on parallel evolution. Our analyses identified pleiotropy as the causal factor for the parallel evolution of gene expression and levels of natural variation. We conclude that pleiotropy is a key factor determining evolutionary predictability.

50 Structure for a Functional Genomics CURE: Comparison of 26 RNAseq pipelines from mapping to functional pathway enrichment using Daphnia's response to caloric restriction *Tonia Schwartz*, Amanda Clark Auburn Univeristy

Curriculum-based undergraduate research experiences (CUREs) provide a valid research experience addressing a scientific question. Bioinformatics and functional genomics, topics once reserved for graduate students, are now entering the undergraduate curriculum. RNAseq is now a standard molecular method to test for differential gene expression among treatment groups and identify functional pathways that are responding to stress and diverging across populations and species. The analysis of RNAseq data is still considerably variable with new programs continually being develop. RNAseq analysis typically consists of very distinct steps, each of which utilizes specific programs: mapping to a reference (whether genome or transcriptome), defining genes and transcripts to obtain the counts of reads that mapped to each, statistical differential gene expression analysis, enrichment for functional pathways. Mixing and matching the programs across these steps can produced a maze of potential pipelines. Choosing a path through these pipelines can be intimidating to students. We use RNAseq dataset developed from Daphnia that were exposed to full feed or caloric restriction to teach the basic analysis and use of bioinformatic pipelines for RNAseq analysis, with students choosing their paths, to test the question "to degree does the choice of a particular program matters in the biological interpretations of the experiment". We have now formally developed a user friendly "chose your own adventure" of 26 different pipelines using popular programs that can be easily utilized in a classroom setting an applied to most any dataset facilitating CURE based structure

for bioinformatics courses that can be easily transported across institutions.

51 The Pond Team: An example of connecting interdisciplinary research and place-based learning for undergraduate research *Ashley Elias*, Dawn Drake, Carissa Ganong, Michael Grantham, Karen Koy, Mark Mills Missouri Western State University

Undergraduate research experiences benefit student scientific skills, and recent trends in undergraduate research and education include a focus on interdisciplinary projects and on place-based learning. We formed an interdisciplinary undergraduate research program focused on a local aquatic ecosystem that included five faculty from multiple scientific disciplines (virology, genomics, ecology, zoology, and geosciences). Students participated in approximately half a day of research per week, weekly hour-long lab meetings of the entire group, and end-of-semester poster presentations to the university. This presentation will include a discussion of self-assessment data from this student research group, including the perceived benefits and challenges, and how this program's structure could be modified for use at other undergraduate institutions. This presentation will also include a highlight of one facet of the research group as an example of the type of data that is being collected. A survey of the fish in the nine ponds on the Missouri Western State University campus was completed by students. However, potential species hybridization was identified. Sunfish hybrids can be phenotypically difficult to distinguish, so identification can be done using molecular techniques. This method is also referred to as DNA barcoding which involves sequencing a short, standardized DNA region that is known to distinguish species. Differentiation of hybrids from green or bluegill sunfish is important for not only adaptive management and survey information, but also for ecological questions. The data from the DNA barcoding for fish species identification are being combined with the other data from the research group to address the question of what factors are affecting the biodiversity of these freshwater ponds. Overall, bringing together faculty from multiple disciplines with projects focused on the same local ecosystem is a valuable technique for providing undergraduate students with hands-on research experience, exposure to a diversity of research areas and methods, and a better understanding of the ecosystems in and around their campus.

52 Coding and data science programs for every biology student Pleuni Pennings San Francisco State University

Over the last seven years, I have been involved in the creation of several programs at SF State University to increase the number of biology and biochemistry students that learn coding and data science skills. With my colleagues, I work hard to make sure that all our computational classes and programs are welcoming to students from groups underrepresented in (computational) research. We run two part-time summer programs and the following academic programs: the PINC minor (CS for undergrads), the GOLD certificate (data science for MS students), and the gSTAR certificate (biotech/ML for Bio/Chem/CS students), with support from NSF and Genentech. Together, these programs now reach about 200 students each year. The success of these programs shows that biology students at SFSU, which are majority women of color and include many first-generation students, are very interested and highly successful in learning computational skills—we just need to create more inclusive classes that speak to more students.

53 Improving the estimation of DFE using paired allele frequency and allele age information *Vivaswat Shastry*, Jeremy Berg University of Chicago, Chicago, IL

Estimating the fitness effect of de-novo mutations is an important problem in population and evolutionary genetics. The accurate estimation of the distribution of fitness effects (DFE) is crucial in understanding a broad variety of processes, from selection shaping genetic diversity in natural populations to the evolution of complex traits in the human population. Broadly, the DFE can be inferred through two approaches: experimentally, through mutagenesis or mutation-accumulation and computationally, via analysis of existing variation in natural populations. We focus on the latter approach in which the site frequency spectrum (SFS) at potentially functional sites is used to infer the DFE for deleterious and nearly neutral mutations. **However, the allele frequency alone at an individual site contains relatively little information about its selection coefficient, limiting the utility of the SFS as a source of information about the DFE.** Here, we propose incorporating more information about the trajectory of these deleterious mutations into our estimation by jointly modeling the frequency and age of deleterious mutations. We compute the likelihood for a certain selection coefficient given the paired data of allele frequency and age by evolving the SFS forward-in-time (starting with a de-novo mutant) across a range of parameters. This iterative algorithm is based on a close approximation to the Wright-Fisher model with additive selection previously proposed by Jouganous et. al., 2017, and as a result, can incorporate the effect due to complex demographic histories. Using simulated data, we show that, for the same number of sites, using allele age jointly with allele frequency allows us to achieve higher levels of accuracy in estimating the selection coefficient compared to using the SFS alone. With the development of methods that construct genome-wide genealogies, we can calculate the allele age for mutations fairly quickly and accurately from large empirical data sets. Finally, we hope to use this light-weight

and flexible approach outlined above to improve our estimation of the DFE in specific functional contexts of the human genome (especially regions in which we have very few variants) to provide a finer view into the evolution of complex traits.

54 To Scale or Not to Scale: The Influence of Scaling on Forward-in-time Population Genetics Simulations *Amjad Dabi*, Daniel Schrider University of North Carolina at Chapel Hill

Forward-in-time simulations are a valuable tool in population genetics that offer a chance to capture complex evolutionary dynamics but are more computationally expensive than commonly used coalescent simulations. Scaling of such forward-in-time simulations is a popular method of alleviating this computational burden by scaling the number of simulated individuals and the duration of the simulated time period (in generations), while upscaling other parameters such as the mutation rate by the same value. This rescaling approach is based on assumptions that may be violated in commonly simulated scenarios (e.g. those with natural selection), potentially biasing results. However, there have been few studies on the effects of such scaling on the accuracy of results. In this study, we carry out both unscaled and scaled simulations under a variety of demographic and selective scenarios using various scaling factors, and then use statistical and machine learning tools to investigate how key simulation results and summary statistics diverge from their unscaled counterparts. Our results indicate that some properties of simulations, such as the site frequency spectrum, are relatively robust to modest rescaling, while others such as the distribution of mutation loss times can be biased substantially.

55 The fitness of an introgressing haplotype *Andrius Dagilis*, Daniel Matute University of North Carolina at Chapel Hill

The genomic era has made clear that introgression, or the movement of genetic material between species, is a common feature of evolution. Examples of introgression that is both adaptive and selected against exist in a variety of systems. What is unclear is how the fitness of an introgressing haplotype changes as species diverge, or as the size of the introgressing haplotype changes. In a simple model, we show that early in the process of divergence, introgression of large haplotypes can be favored more than introgression of individual alleles. This is because introgressing haplotypes bring in not only potential incompatibilities in the form of deleterious epistasis with the receiving genome, but also positive epistasis in the form of co-adapted alleles on the same haplotype. The build up of incompatibilities between diverging species begins to favor the introgression of small haplotypes in the long run, while in highly diverged species even single alleles with positive direct effects can be selected against. This model generates several novel predictions - while it is consistent with observations of a positive relationship between recombination rate and introgression frequency across the genome, the model suggests this relationship may not exist or be entirely inversed in recently diverged species pairs. The model also generates predictions about asymmetry in the direction of introgression.

56 Uncovering footprints of natural selection through time-frequency analysis of genomic summary statistics *Sandipan Paul Arnab*, Michael DeGiorgio Florida Atlantic University

Natural selection leaves a spatial pattern along the genome, with a distortion in the haplotype distribution near the selected locus that becomes less prominent with increasing distance from the locus. Evaluating the spatial signal of a population-genetic summary statistic across the genome allows for patterns of natural selection to be distinguished from neutrality. Different summary statistics highlight diverse components of genetic variation and, therefore, considering the genomic spatial distribution of multiple summary statistics is expected to aid in uncovering subtle signatures of selection. In recent years, numerous methods have been devised that jointly consider genomic spatial distributions across summary statistics, utilizing both classical machine learning and contemporary deep learning architectures. However, better predictions may be attainable by improving the way in which features used as input to machine learning algorithms are extracted from these summary statistics. To achieve this goal, we apply three time-frequency analysis approaches (wavelet, multitaper, and S-transform analyses) to summary statistic signals. Each analysis method converts a one-dimensional summary statistic signal to a two-dimensional image of spectral density or visual representation of time-frequency analysis, permitting the simultaneous representation of temporal and spectral information. We use these images as input to convolutional neural networks and consider combining models across different time-frequency representation approaches through the ensemble stacking technique. Application of our modeling framework to data simulated from neutral and selective sweep scenarios reveals that it achieves almost perfect accuracy and power across a diverse set of evolutionary settings, including population size changes and test sets for which sweep strength, softness, and timing parameters were drawn from a wide range. Given that this modeling framework is also robust to missing data, we believe that it will represent a welcome addition to the population-genomic toolkit for learning about adaptive processes from genomic data.

57 Heritable epigenetic variation facilitates maintenance of genetic variation *Amy Webster*, Patrick Phillips Univer-

sity of Oregon, Eugene, OR

Understanding how genetic variation is maintained is a major problem in population genetics. A variety of factors have been implicated to explain the maintenance of genetic variation in some contexts (*e.g.* balancing selection), but the potential role of epigenetic regulation to influence population dynamics has been understudied. It is well recognized that epigenetic regulation, including histone methylation, small RNA expression, and DNA methylation, helps to define differences between cell types and facilitate phenotypic plasticity. In more recent years, empirical studies have shown the potential for epigenetic regulation to also be heritable for at least a few generations, raising the possibility that differences in epigenetic regulation can act alongside genetic variation to shape evolutionary trajectories. Like genetic mutation, heritable differences in epigenetic regulation can arise spontaneously; these are termed 'epimutations'. Epimutations differ from genetic mutations in two key ways – they occur at a higher rate, and the loci at which they occur often revert back to their original state within a few generations. Here, I present an extension of the standard population-genetic model with selection to incorporate epigenetic variation arising via epimutation. My model assumes a diploid, sexually reproducing population with random mating. In addition to spontaneous genetic mutation, I included parameters for spontaneous epimutation and back-epimutation, allowing for four potential epialleles at a single locus (two genetic alleles, each with two epigenetic states), each of which affect fitness. I then analyzed the conditions under which stable epialleles were maintained. My results show that, for certain sets of parameters, epimutations affecting fitness can facilitate the maintenance of recessive deleterious alleles that would otherwise be purged from the population. This demonstrates that heritable epigenetic regulation that itself is maintained only a few generations may be an important factor in the long-term maintenance of genetic variation in populations.

58 The population genetics of collateral resistance and sensitivity *Sarah Ardell*, Sergey Kryazhimskiy University of California, San Diego

Antibiotic resistant infections are predicted to kill millions of people per year within the next 30 years. As the rate of new antibiotic development dwindles, the outlook of many single drug treatments is grim. This major challenge has sparked interest in developing multi-drug treatments which utilize current antibiotics together to effectively eliminate patients' infections before dual (or higher order) resistance emerges. Successful multi-drug treatments rely on the phenomenon known as collateral sensitivity – when evolution of resistance to one drug leads to the simultaneous decrease in resistance to another.
However, if mutations with diverse collateral effects are available, the treated population may evolve either collateral sensitivity or its opposite - collateral resistance. How to design treatments robust to such uncertainty is unclear. We show that many resistance mutations in Escherichia coli against various antibiotics indeed have diverse collateral effects. We propose to characterize such diversity with a joint distribution of fitness effects (JDFE), which describes the probability of a new mutation having any given set of selection coefficients across the antibiotic environments of interest. We then develop a theory for describing and predicting collateral evolution based on simple statistics of the JDFE. We show how to robustly rank candidate drug pairs to minimize the risk of collateral resistance and how to estimate JDFEs. In addition to practical applications, these results have implications for our understanding of evolution in variable environments.

59 Examining polygenic adaptation in time-stratified genome samples with diffusion-based hidden-Markov models *Xiaoheng Cheng*[1], Matthias Steinrücken[1,2] 1) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL

With the rapid accumulation of ancient DNA (aDNA) genomes and evolve-and-resequence (E&R) data, more time-stratified population genomic datasets are emerging. Such time-series data allow us to examine the temporal dynamics of natural selection and can lend power to detecting its footprints. Few selection-detecting methods, however, are tailored to jointly consider multiple samples of the same population at different time, especially when more than one locus is involved. Meanwhile, increasing evidence is supporting the polygenic nature of most traits under selection, underscoring the need for approaches that account for multiple loci. Here, we constructed a hidden-Markov model (HMM) framework based on the Wright-Fisher diffusion model explicitly for directional or stabilizing selection on polygenic traits. To reduce the computational load, we use a normal approximation for the step-wise transition between generations in the underlying diffusion model. We implemented this framework as a Python package with its command-line-interacting script, and benchmarked its performance with both forward and backward simulations. Further, for each major complex trait in the UKBioBank, we extract variation records of their significant loci in historical British populations during the past ~3500 years from the Allen aDNA Resources dataset, and assign these loci their respective effect size estimates in UKBioBank. With the composite likelihood incorporating all significant loci, we were able to obtain estimates of the predominant mode of selection, *i.e.*, directional or stabilizing, and its intensity, for each trait, gaining insights into recent human evolu-

tion.

60 Interpretable machine learning improves performance in association, discovery, and prediction *Mariano Alvarez*[1], Emily Abernathy[1], Cynthia Rudin[2] 1) Avalo; 2) Duke University, Durham, NC

The incorporation of machine learning (ML) techniques in genomics has enhanced a variety of routine tasks, from predicting gene function to genomic selection. However, tasks outside of prediction have received little attention, largely because of the black box nature of ML algorithms. The emerging literature on interpretable ML offers useful innovations that might allow the use of ML techniques for discovery and deeper understanding. We use the model-agnostic concept of conditional model reliance and show how simple, interpretable summary statistics can be generated from any black box prediction algorithm. We then test the performance of these statistics across several common tasks in quantitative genetics, including association mapping, genomic prediction, and tests for natural selection, and find that estimations of conditional model reliance provide substantially higher performance than existing methods. Specifically, using simulations and real data in Arabidopsis thaliana and Oryza sativa, we show that interpretable ML techniques can more accurately identify causal loci in association mapping and loci under selection when the selective environment is known. We also show that modeling these loci alone can meet or exceed prediction accuracy from models using whole-genome data. We suggest that interpretable ML techniques offer new opportunities to model and derive insight into difficult problems in evolutionary biology.

61 Response of Quantitative Traits to Directional Selection in Finite Populations *Hannah Götsch*[1,2,3], Reinhard Bürger[1] 1) Faculty of Mathematics, University of Vienna, Austria; 2) Vienna School of Mathematics, Austria; 3) Vienna Graduate School of Population Genetics, Austria

Phenotypic adaptation can occur rapidly as a response to a sudden - natural or artificial - change in the environment. In the early phase of adaptation, the evolution of gene frequencies will be mainly driven by directional selection. While molecular population genetics focuses on the dynamics of single loci, quantitative genetics has a more trait-centered view. We combine these two approaches to describe adaptation of complex traits by studying the evolutionary dynamics at individual loci. For a finite panmictic population, we derive accurate approximations for the distribution of newly arising beneficial mutations as a function of time, first for a single locus. Then, using an infinite sites model and assuming unlinked loci (sites) loci contributing additively to a quantitative trait under weak non-epistatic directional selection, we derive highly accurate approximations for the evolutionary dynamics of the phenotypic mean and variance. Thus, we examined the effect on polygenic adaptation of selection, random genetic drift, and mutation rate and mutational effects drawn from a distribution. We also present results for the number of segregating sites during adaptation. The mathematical model is based on a combination of branching process theory (for the initial stochastic phase) and deterministic theory. Although diffusion theory leads to simple expressions for many quantities such as fixation probabilities and times, time-dependent results seem to be out of reach. Our approach yields this time dependence and is especially accurate in the initial phase of adaptation. However, we also derive highly accurate results for the stationary phase under long-term selection, i.e., when the phenotypic variance has stabilized. These refine classical results. Analytic approximations are tested by comprehensive simulations based on a Wright-Fisher model. As an application, we explored when the response of the trait is mainly caused by selective sweeps at few loci and when it is due to subtle allele-frequency shifts at many loci. We found that the selection strength determines primarily the rate of adaptation. The central parameter determining the relative importance of sweeps vs. shifts is the population-wide mutation rate. However, unequal mutational effects of different loci blur this distinction, and intermediate patterns occur in a wider parameter range if mutational effects have large variance.

62 Phylogenomic comparative methods: accurate evolutionary inferences in the presence of gene tree discordance *Mark Hibbins*[1], Lara Breithaupt[1,2], Matthew Hahn[1,3] 1) Department of Biology, Indiana University, Bloomington, IN, USA; 2) Department of Statistical Science, Duke University, Durham, NC, USA; 3) Department of Computer Science, Indiana University, Bloomington, IN, USA

Phylogenetic comparative methods have long been a mainstay of evolutionary biology, allowing for evolutionary inferences across species while accounting for their common ancestry. These analyses typically assume a single, bifurcating species tree that describes this shared history. However, modern phylogenomic analyses have shown that genomes are often composed of a mosaic of different histories that can disagree both with the species tree and each other. These discordant gene trees describe shared histories among species that are not captured by their history of speciation, and therefore that are unaccounted for in standard comparative approaches. The application of standard methods to species histories containing discordance leads to incorrect inferences about the timing, direction, and rate of evolution. Here, we develop two approaches for incorporating gene tree histories into comparative methods: one involves constructing a full-

er phylogenetic variance-covariance matrix that includes branches not found in the species tree, while the other applies the pruning algorithm over a set of gene trees to calculate trait histories and likelihoods. Both approaches are agnostic to the biological cause of gene tree discordance, which may include incomplete lineage sorting and introgression. Using simulation, we demonstrate that our new approaches generate much more accurate estimates of evolutionary parameters than standard methods. Finally, by analyzing empirical data from a rapid radiation, we highlight how discordance can potentially lead to false inferences of clade-specific shifts in trait evolutionary rates.

63 Indirect genetic effects across life cycle stages in a cooperatively breeding bird Gladiana Spitz[2], Elissa Cosgrove[3], Reed Bowman[4], John Fitzpatrick[3], Andrew Clark[3], *Nancy Chen*[1] 1) University of Rochester, Rochester, NY; 2) University of Colorado, Boulder, Boulder, CO; 3) Cornell University, Ithaca, NY; 4) Archbold Biological Station, Venus, FL

Indirect genetic effects, which occur when an individual's phenotype is influenced by the genotype of another conspecific individual, are an often-overlooked yet potentially important factor impacting phenotypic variation in natural populations. In many organisms, interactions between parents and offspring can have widespread consequences on offspring traits and fitness. Most indirect genetic effects studies to date have focused on estimating maternal effects, even though males also care for offspring in many species. Furthermore, the strength of maternal and paternal effects is expected to vary for different offspring traits, and few studies have investigated the ontogeny of paternal effects. Here, we estimated the environmental and genetic effects of maternal, paternal, and helper care on offspring survival and body condition at different life stages in an intensively-studied population of cooperatively breeding Florida Scrub-Jays (*Aphelocoma coerulescens*). Using an animal model approach, we found that the strength of maternal and paternal effects decrease with offspring age, but the pattern differs for helper effects. Paternal effects are higher than maternal effects for early nestling survival, which is consistent with variation in provisioning behavior. We also estimate the effects of non-transmitted parental alleles and incorporate parental genotypes in genome-wide association studies of different fitness components to disentangle direct effects from indirect effects. Our study highlights the importance of broadening the consideration of indirect genetic effects beyond maternal effects when studying the evolution of offspring traits and performing genotype-phenotype associations.

64 Association and Fine-Mapping with Bayesian Machine Learning Methods Lorin Crawford Microsoft Research New England, Cambridge, MA

A common goal in genome-wide association (GWA) studies is to characterize the relationship between genotypic and phenotypic variation. Linear models are widely used tools in GWA analyses, in part, because they provide significance measures which detail how individual single nucleotide polymorphisms (SNPs) are statistically associated with a trait or disease of interest. However, traditional linear regression largely ignores non-additive genetic variation, and the univariate SNP-level mapping approach has been shown to be underpowered and challenging to interpret for certain trait architectures. While machine learning methods such as neural networks are well known to account for complex data structures, these same algorithms have also been criticized as "black box" since they do not naturally carry out statistical hypothesis testing like classic linear models. This limitation has prevented machine learning approaches from being used for association mapping tasks in GWA applications. In this talk, we present flexible and scalable classes of Bayesian multi-layer perceptron models which provide interpretable probabilistic summaries such as posterior inclusion probabilities and credible sets for association and fine-mapping tasks in high-dimensional GWA studies. We illustrate the benefits of our methods over state-of-the-art linear approaches using extensive simulations. We also demonstrate the ability of these methods to recover both novel and previously discovered genomic associations using traits from the Wellcome Trust Case Control Consortium (WTCCC), the Framingham Heart Study, and the UK Biobank.

65 How hybrid incompatibilities agglomerate on gene networks Rafael Guerrero North Carolina State University

Postzygotic reproductive isolation--manifesting as sterile or inviable hybrids--is often the result of Dobzhansky-Muller incompatibilities, epistatic interactions involving the substitutions accumulated between diverging populations. Previous theoretical work has found that the number of incompatibilities should grow faster than linearly with time, the so-called "snowball effect". In this talk, we develop a gene network framework for the evolution of genetic incompatibilities and explore how underlying properties of the network--topology, density, and size--affect the snowball effect. We find that, while a snowball is generally expected for pairwise genetic incompatibilities, the agglomeration of these incompatibilities into higher-order clusters is more common than previously thought. These findings imply the need to revise some simplifying assumptions about the genetic basis of incompatibilities, and suggest that the buildup of reproductive isolation will not typically track the accumulation of incompatibility loci.

66 Timing and causes of evolution of human germline mutation spectrum Ziyue Gao[1], Yulin Zhang[2], Molly Przewor-

ski[3], *Priya Moorjani*[2] 1) University of Pennsylvania, Philadelphia, PA; 2) University of California, Berkeley, CA; 3) Columbia University, New York, NY

Germline mutations are the source of all heritable variation. Understanding the rate and mechanisms by which mutations occur is of paramount importance for studies of human genetics (to interpret heritable disease prevalence) and evolutionary biology (to date evolutionary events). Over the past decade, there has been a flood of data in genomics—within pedigrees, among populations and across species—that is fundamentally revising our understanding of the process of mutagenesis. In my talk, I will first briefly summarize the key findings from these different datasets and then discuss recent findings investigating differences in mutation rate and spectrum (i.e., proportions of different mutation types) across human populations. To investigate inter-population differences, we developed a framework to compare polymorphisms that arose in different time windows in the past while controlling for the effects of selection and biased gene conversion. Applying this approach to Europeans, Africans and East Asians from 1000 Genomes Project, we uncovered multiple significant differences in the mutation spectrum within and across human populations, including at least two independent changes that occurred after the split of the continental groups. Interestingly, we also found that non-Africans and Africans differ significantly in their mutation spectra even for ancient polymorphisms that predate out-of-Africa migration likely due to mutational differences between the ancestors of modern humans and archaic hominins. By relating the observed variations in polymorphisms to the parental age effects on de novo mutations, we show that plausible estimates of reproduction ages cannot explain the joint patterns observed for different mutation types, implying that changes at the molecular level such as genetic modifiers and varying environments have had a non-negligible impact in shaping the human mutation landscape. The composite nature of mutation rate underscores the challenges of using it as the molecular clock for dating evolutionary events even for recent timescales.

67 Could medical privacy be compromised by associations between forensic loci and the expression levels of neighboring genes? *Rori Rohlfs*[1], Jhony Zavaleta[1], Mayra Banuelos[1, 2], Alennie Roldan[1], Rochelle-Jan Reyes[1], Miguel Guardado[1], Berenice Chavez Rojas[1], Thet Nyein[1], Ana Rodriguez Vega[1], Maribel Santos[1], Emilia Huerta Sanchez[2] 1) San Francisco State University; 2) Brown University

A set of 20 short tandem repeats (STRs) is used by the United States criminal justice system to identify suspects, and to maintain a database of genetic profiles for individuals who have been previously convicted or arrested. Some of these STRs were identified in the 1990s, with a preference for markers in putative gene deserts to avoid forensic profiles revealing protected medical information. We revisit that assumption, investigating whether forensic genetic profiles reveal information about gene expression variation, or potential medical information. We find six significant correlations (FDR = 0.23) between the forensic STRs and the expression levels of neighboring genes in lymphoblastoid cell lines. We explore possible mechanisms for these associations, showing evidence compatible with forensic STRs causing expression variation, or being in LD with a causal locus in three cases, and weaker or potentially spurious associations in the other three cases. Together, these results suggest that forensic genetic genotypes may reveal expression levels and, perhaps, medical information.

68 Uncovering the genetic basis of local adaptation in maize with large-scale multi-environment trials *Daniel Runcie*[1], Daniel Gates[1], Garrett Janzen[2], J. Alberto Romero Navarro[3], Martha Willcox[4], Kai Sonder[4], Samantha Snodgrass[2], Fausto Rodriguez-Zapata[5,6], Ruairidh Sawers[7], Ruben Rellan-Alvarez[6], Edward Buckler[3], Sarah Hearne[4], Matthew Hufford[2], Jeffrey Ross-Ibarra[1] 1) University of California Davis, Davis, CA; 2) Iowa St University, Ames, IA; 3) Cornell University, Ithica, NY; 4) International Maize and Wheat Improvement Center (CIMMYT), Mexico; 5) Laboratorio Nacional de Genomica para la Biodiversidad/Unidad de Genómica Avanzada, Cinvestav, Mexico; 6) North Carolina State University, Raleigh, NC; 7) The Pennsylvania St University, State College, PA

Threats to crop production due to climate change are one of the greatest challenges facing society today. Considerable adaptive variation exists in traditional landraces and germplasm collections, but effective use of this diverse germplasm requires separating adaptive alleles from linked deleterious variants. One strategy to identify adaptive loci is to grow panels of varieties in multiple environmental contexts and use genome-wide association analyses (GWAS) to identify loci exhibiting adaptive patterns such as fitness benefits in specific environments or geographic clustering of alleles by environmental characteristics. However, the statistical analysis of such gene-environment interactions and gene-environment associations is challenging because of issues of the considerable population and spatial structure in traditional landrace collections and the high-dimensional nature of performance traits measured across multiple locations. We will present a genome-wide association analysis of performance traits in maize using nearly 4000 traditional landraces broadly representing the breadth of genetic diversity of maize in Central and South America. Partially overlapping subsets of these varieties were grown in 23 field trials spanning multiple environmental gradients in Mexico. We will highlight several

methodological developments in the analysis of multi-environment trials allowing accurate and efficient genome-wide modeling of allele-environment interactions and the patterns of plasticity along environmental gradients. Our results produce a detailed and fine-scale map of the properties of potentially beneficial alleles available in the traditional landraces housed in maize germplasm collections.

72W Cross-species incompatibility between a DNA satellite and the *Drosophila* Spartan homolog poisons germline genome integrity *Cara Brand*, Mia Levine University of Pennsylvania

Satellite DNA spans megabases of eukaryotic sequence and evolves rapidly. Paradoxically, satellite-rich genomic regions mediate strictly conserved, essential processes like chromosome segregation and nuclear structure. A leading resolution to this paradox posits that satellite DNA and satellite-associated chromosomal proteins coevolve to preserve these essential functions. We experimentally test this model of intra-genomic coevolution by conducting the first evolution-guided manipulation of both chromosomal protein and DNA satellite. The 359bp satellite spans an 11Mb array in *D. melanogaster* that is absent from its sister species, *D. simulans*. This species-specific DNA satellite colocalizes with the adaptively evolving, ovary-enriched protein, Maternal Haploid (MH)–the Drosophila homolog of Spartan. To determine if MH and 359 coevolve, we swapped the *D. simulans* version of MH ("MH[sim]") into *D. melanogaster*. MH[sim] triggers ovarian cell death, reduced ovary size, and loss of mature eggs. Surprisingly, the *D. melanogaster mh* null mutant has no such ovary phenotypes, suggesting that MH[sim] is toxic in a *D. melanogaster*background. Using both cell biology and genetics, we discovered that MH[sim] poisons oogenesis through a DNA damage pathway. Remarkably, deleting the *D. melanogaster*-specific 359 satellite array completely restores *mh[sim]*germline genome integrity and fertility, consistent with a history of coevolution between these two fast-evolving loci. Germline genome integrity and fertility are also restored by overexpressing Topoisomerase II (Top2), suggesting that MH[sim] interferes with Top2-mediated processing of 359. The observed 359-MH[sim] cross-species incompatibility supports a model under which ostensibly inert repetitive DNA and essential chromosomal proteins must coevolve to preserve germline genome integrity.

73W Cracking Coevolution: Consequences of Space and Genetic Architecture Victoria Caudill University of Oregon

Coevolution between two species can lead to variation in phenotypes across geography, sometimes creating areas where the phenotypes of species closely match and other areas where they do not match at all. However, it is not well-understood how the genetic basis of the traits underlying coevolution affect phenotypic and genotypic variation, especially across a geographical area. I investigate the genetic dynamics of coevolution of two species undergoing reciprocal adaptation across space and time, using both theory and simulations inspired by the newt-snake system. In this system, Taricha granulosa (rough-skinned newt) has developed tetrodotoxin that poisons predators. One of their predators, Thamnophis sirtalis (garter snake), has developed resistance to this toxin. I explore the effects that genetic architecture of the toxin and resistance traits has on the coevolutionary dynamics by manipulating the mutation rate and the effect size of mutations, demonstrating how different genetic architectures lead to different coevolutionary outcomes in a geographically explicit setting. The outcomes of these simulations reveal that a species with genetic architecture of intermediate polygenicity has higher phenotype and population size, suggesting that there might be an optimal combination of mutation rate and effect size in the coevolutionary arms race.

74T Coevolution between two essential telomere binding proteins preserves chromosome end-protection *Sung-Ya Lin*[1], Hannah Futeran[1], Mia Levine[1,2] 1) Department of Biology, School of Arts & Sciences, University of Pennsylvania, Philadelphia, PA; 2) Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Chromosome ends pose two primary threats to genome integrity: the erosion of unique DNA sequence and inappropriate DNA repair into lethal end-to-end chromosome fusions. Specialized telomere binding proteins combat these threats by 1) adding telomere-specific repetitive DNA to the chromosome ends and 2) protecting chromosome ends from telomere fusions. Although these telomere functions are essential in all eukaryotes, many telomere proteins evolve rapidly. Across the *Drosophila* species group, nearly half of the proteins required for telomere functions evolve rapidly under positive selection. Two such proteins, HOAP and HipHop, are essential components of the *Drosophila* end-protection complex – homozygous null mutations in either one result in lethal telomere fusions. These two proteins physically interact and HipHop depends on HOAP for recruitment to the telomere. We hypothesize that HipHop coevolves with HOAP to maintain this physical interaction and ultimately, to ensure its own recruitment to, and protection of, chromosome ends. To test this hypothesis, we transgenically replaced the native *D. melanogaster HipHop* with the highly diverged version from its close relative, *D. yakuba* ("*HipHop[yak]*"). We discovered that *HipHop[yak]* flies are homozygous lethal at the pupal-to-adult transition. Immunofluorescence on mitotic chromosomes from larval brains revealed that HipHop[yak] is depleted at telomeres compared to the control, HipHop[mel]. These cells also exhibited high rates of telomere associations, which are typical of the fused telomeres found in the *HipHop* null mutant. These data

suggest that *HipHop[yak]* behaves as a loss-of-function allele. To determine if lethality results from disrupted coevolution between HOAP and HipHop, we swapped *HOAP* from *D. yakuba ("HOAP[yak]"*) into the *D. melanogaster* fly encoding only *HipHop[yak]*. Remarkably, HOAP[yak] restored HipHop[yak] protein localization to telomere ends, restored telomere end-protection from fusions, and restored adult viability. These results reveal that recruitment of HipHop to telomeres requires the conspecific HOAP and support our hypothesis that HipHop coevolves with HOAP to maintain telomere integrity. We suspect that HipHop evolves to "keep up" with an ever-evolving HOAP, which we previously showed evolves adaptively to restrict selfish telomeric retrotransposons.

75T Distinct genetic mechanism underlining the trait evolution of male-specific wing spot in *Drosophila elegans* species sub-group *Ateesha Negi*[1], Ben-Yang Liao[2], Shu-Dan Yeh[1] 1) National Central University; 2) National Health Research Institutes

Pigmentation is one of the most variable traits among animals and is commonly subjected to either or both, natural and sexual selection. Male-specific wing pigmentation is one of the most variable and novel traits in several *Drosophila* species groups that have recently emerged and subsequently diversified. This male-specific wing spot has been proposed to be associated with frontal wing display, which serves as an advantage to males in the mating process. Thus, the evolution of male-specific wing spot and frontal wing courtship display in the *Drosophila elegans* species subgroup serves an intricate case of genetic correlation of distinct phenotypic traits. Based on the previous studies, the formation of pigmentation results from the conversion of melanin precursors, wherever converting enzymes is present. With insights from recent molecular studies, the expression of genes that encodes pigmentation influencing enzyme, alone are not sufficient to generate wing spots neither the loss of these genes are sufficient to lose wing spots. Therefore, the molecular bases underlying the genetic mechanism for even only wing spot evolution remained elusive. Hence, we plan to perform whole genome sequencing and transcriptome analysis of *D. elegans* (which possesses both male wing spots and frontal wing courtship display) and its sibling species *D. gunungcola* (which lacks both of these characters). By adopting contemporary sequencing technology we can perform a comprehensive characterization of the genome assembly which will include genomic size, satellite DNA content, comparative gene catalog, genetic switch(s), characterization of structural variants, protein-coding gene catalog, detect mutation(s), and base modification such as methylation. All this will substantially expand our knowledge regarding genetic bases that cause wing spot and frontal wing courtship display.

76W The role of continuous geography in parasite local adaptation *Bob Week*, Gideon Bradburd Michigan State University, East Lansing Michigan

Previous work modelling the consequences of interactions between gene-flow, random genetic drift and coevolution on local adaptation in host-parasite systems have made use of metapopulation models where geography is discretized and dispersal is global. These studies generally conclude that increased rate of gene-flow for the parasite relative to the host yields parasite local adaptation. The colloquial explanation for this result is that increased dispersal can lead to increased genetic variance and thus promote adaptive capacity. However, this explanation ignores the effect of continuous versus discrete geography and how patterns of limited versus global dispersal interact with drift to determine which species is locally adapted. To close this gap we present a phenotypic model of host-parasite coevolution in continuous space with finite local population densities. We find the parasite is locally adapted to its host when its average dispersal distance is shorter than the average dispersal distance of the host. Performing a similar analysis on a metapopulation model of host-parasite trait coevolution where dispersal is global and genetic variance is held constant, we recover the classical result described above. This highlights the role of geography and in particular suggests a subtle distinction between the effects of dispersal rate and dispersal distance on parasite local adaptation.

77W Predicting Antibiotic Resistance Through the Utilization and Comparison of Machine Learning Algorithms *Jameel Ali*, Meris Johnson-Hagler, Faye Oracles, John Matt Suntay, Kristiene Recto, Lucy Mocteczuma, Fayeeza Shaikh, Pleuni Pennings San Francisco State University

Antibiotic resistance has become a global public health concern. Bacteria are evolving resistance to the current arsenal of prescribed antibiotics resulting in strains that are developing multi-drug resistance. Currently, clinics are often performing traditional culture-based assays to determine antibiotic resistance in bacterial strains. However, this method is time-consuming and may be phenotypically inaccurate. To determine antibiotic resistance with a greater degree of accuracy and efficiency than traditional methods, we will be utilizing machine learning algorithms. The machine learning algorithms will process publically available whole genome sequences of E. coli strains to produce Decision Trees, Random Forest, and Gradient Boosted Trees models. We want to compare the machine learning models to determine which one has the best accuracy when using population structure, isolation year, and gene content as features. Through comparative analysis, we want to identify which features can predict antibiotic resistance.We aim to use what we have learned from this

study to contribute to a future where machine learning can be used as a diagnostic tool to accurately predict antibiotic resistance from whole genome sequencing data.

78T Uncovering the mysteries of antibiotic resistance through phylogenetic analysis *MaryGracy Antony*, Florentine van Nouhuijs, Faye Orcales , Pleuni Pennings San Francisco State University

Highly Drug-Resistant Bacteria or Superbugs are a result of Antibiotic Resistance, in which individuals affected with a resistant infection cannot heal by taking antibiotics. The mechanism in which a patient acquires a resistant infection is complicated and not well understood. Antibiotic resistance infections can either be transpired through transmission or via within-host evolution.
Transmitted resistance occurs when a patient contracts an already resistant strain.
De novo evolution occurs when a patient contracts a susceptible strain which through mutation or horizontal gene transfer evolves to become drug resistant.
Knowing the role of transmission and within-host evolution is important to design effective prevention programs. We propose to use phylogenetic trees of bacterial genomes to understand the roles of transmission and within-host evolution of resistance.
We are using a publicly available dataset (Kallonen et al YEAR) of E coli genomes and resistance phenotypes, along with the phytools package in R (Revell, 2008). Phytools is commonly utilized to analyze phylogenetic trees, but hasn't been used much to understand drug resistance.
By using the phytools package, the constructed phylogenetic trees will be compared with one another and clade sizes for resistant phenotypes will be determined for different drugs.
Preliminary findings show that resistance to the antibiotic Gentamicin, was observed in isolated places on the phylogenetic tree, which suggests that it is not easily transmitted, possibly because it is costly. On the other hand, Amoxicillin resistance was observed in larger clades, which suggests transmission playing a bigger role.
We hope after accomplishing this preliminary step to eventually estimate rates of within-host evolution and transmission of resistance for different drugs and pathogens.

79T Somatic Mutations in Mitochondrial DNA of Drosophila *Mara Baylis*, Jakob McBroome, Russell Corbett-Detig University of California, Santa Cruz

In humans, increasing heteroplasmy -or genetic variation of mitochondrial DNA within the body- has been associated with aging, neurodegenerative disease, cancer, and muscle degeneration disorders. However, the occurrence and distribution of the mitochondrial somatic mutations that create heteroplasmy remain poorly understood. Evaluating how these mutations accumulate through the lifespan of an organism and what genetic mechanics create them is key to understanding the aging process at a fundamental level. Using an ultra-accurate sequencing technique capable of detecting low-frequency variation within a single tissue sample, I have identified very low-frequency mutations at high accuracy in *Drosophila* mitochondrial DNA. To address this question, we are using *Drosophila melanogaster* as a model organism, as they are well suited to this research. By analyzing multiple individuals across different life stages, I compared mitochondrial mutational loads and evaluate the accumulation of heteroplasmy across the *Drosophila* lifespan. This life stage data will also be used to determine the different types of mutations occurring, can be used to draw conclusions about the presence of a selection event that occurs within the life stages, which could indicate that somatic selection occurs after the bottleneck in oogenesis. Additionally, I have identified conserved regions in mtDNA, evaluated mitochondrial diversity within and between cells, and revealed patterns related to underlying mechanisms of mutation.

80W Sex chromosome evolution in beetles *Ryan Bracewell*[1], Anita Tran[2], Kamalakar Chatla[2], Doris Bachtrog[2] 1) Indiana University Bloomington; 2) University of California Berkeley

Beetles are the most species-rich group of animals and harbor diverse karyotypes. Most species have XY sex chromosomes, but X0 sex determination mechanisms are also common in some groups. We generate whole-chromosome assemblies of two beetle species and utilize eight additional beetle genomes to reconstruct karyotype evolution across Coleoptera. We identify ancestral linkage groups that share a conserved set of genes across beetles, thereby revealing chromosome conservation over hundreds of millions of years. While the ancestral X chromosome is maintained across beetles, we find distinct cases of additions of autosomes to the ancestral sex chromosomes. These neo-sex chromosomes evolve the stereotypical properties of sex chromosomes, including the evolution of dosage compensation, and genes with sex-biased expression. Beetles thus provide a novel model to gain a better understanding of the diverse forces driving sex chromosome evolution and we propose a standardized chromosome naming scheme aimed at helping future studies of genome evolution within this exceptionally diverse radiation.

81W Mapping the genomic basis of trophic level adaptation in Lake Malawi cichlid fishes *Aldo Carmona Baez*[1,2,3], Kaitlin Coyle[2,3,4], Melissa Lamm[1,3,5], Emily Moore[2,3,6], Natalie Roberts[1,3,7], Ethan Dickson[1,3,8], Katara Griffith[1,3,9], Gargi Damle[1,3,10], Erin Peterson[1,2,3,4], Patrick Ciccotto[3,11], Maddy Arena[1,3], Reaghan Julke[1,3], David Reif[1,2,3], Reade Roberts[1,2,3] 1) Genetics and Genomics Academy, North Carolina State University; 2) Graduate Program in Genetics, North Carolina State University; 3) Department of Biological Sciences, North Carolina State University; 4) Q2 Solutions. Durham, NC; 5) Novogene, Research Triangle Park; 6) Division of Biological Sciences, University of Montana; 7) Prestage Department of Poultry Science, North Carolina State University; 8) RNA Biology Laboratory, National Cancer Institute; 9) Genetics, Bioinformatics, and Computational Biology, Virginia Tech University; 10) BiNGS Core, Mount Sinai Icahn School of Medicine; 11) Warren Wilson College

Trophic specialization is key to the phenotypic and species diversity observed across life. Despite its importance, very few studies have explored the genetic basis of trophic level itself as a complex trait. In this study, we used recently diverged Malawi cichlid species as a model to identify candidate genes involved in trophic level differences with a forward genetics approach. We performed pairwise Fst scans among carnivore, omnivore, and herbivore species. These comparative genomics scans detected numerous regions of divergence by trophic level, including genes previously implicated in trophic adaptations in cichlids and other species. We then compared the results of the Fst scans with the recombination patterns identified in two hybrid crosses of four genera representing three different trophic levels.
Our results suggest that regions of the genome identified in the Fst scans are located in the vicinity of putative structural variants identified via recombination patterns in the hybrid crosses. These analyses point out the potential impact that genomic structural rearrangements have had in the evolution of trophic level evolution in the Malawi cichlid radiation.

82T Using the *Eucalyptus polybractea* genome improved genetic variant identification compared to using a pseudo-reference *Swapan Chakrabarty*[1], Teng Li[2], David Kainer[3], William J. Foley Foley[4], Allen Rodrigo[2], Carsten Külheim[1] 1) College of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan 49931, USA; 2) School of Biological Sciences, The University of Auckland, Auckland 1142, New Zealand; 3) Center for BioEnergy Innovation, Bioscience Division, Oak Ridge National Laboratories, Oak Ridge, TN 37831, USA; 4) Research School of Biology, The Australian National University, Canberra 2600, Australia

Eucalyptus polybractea, commonly known as blue mallee is cultivated in South-Eastern Australia to produce Eucalyptus oil. *Eucalyptus polybractea* oil is valued for its flavor and medicinal properties. These essential oils are dominated by monoterpenes such as 1,8-cineole, and also contain some sesquiterpenes, and other volatile compounds. Using genetic markers based on genome-wide association studies may greatly enhance breeding programs aimed at improving oil yield and quality. A high-quality reference genome sequence can enable us to study the genetic architecture and identify candidate genes related to high-quality foliar essential oils. In this study, we utilized the hybrid assembly of the *E. polybractea* genome from both short- and long-read technology. We generated 44 Gb of Illumina HiSeq short reads and 8 Gb of Nanopore long reads, representing approximately 83× and 15× genome coverage, respectively. After polishing, the hybrid-assembled genome contained 24,864 scaffolds with an accumulated length of 523 Mb (N50 = 40.3 kb; BUSCO-calculated genome completeness of 94.3%). The genome contained 35,385 predicted protein-coding genes detected by combining homology-based and de-novo approaches. We tested if the hybrid assembled genome improved on the detection of high-quality genetic variants, compared to a published study, where we used an in-vitro generated pseudo genome reference based on the *E. grandis* genome edited with fixed alleles from *E. polybractea*. For this comparison we mapped cleaned reads from 480 *E. polybractea* samples to three reference sequences: 1. *E. grandis*, 2. *E. polybractea* pseudo reference, and 3. the hybrid assembled *E. polybractea* genome reference. Variants were identified from all three approaches and GWAS performed and compared. Thus, the high-quality genome of *E. polybractea* facilitated better mapping and identification of genetic variants that allowed us to identify candidate genes related to terpene production in *E. polybractea*.

83T Leveraging a de-novo long read assembly for comparative and functional genomics of the *Octopus bimaculoides* *Gabrielle Coffing*[1], Jeremea Songco[2], Judit Pungor[2], Denise Piscopo[2], Adam Miller[2], Cris Niell[2], Andrew Kern[1] 1) Institute of Ecology and Evolution, University of Oregon; 2) Institute of Neuroscience, University of Oregon

The California two-spot octopus (*Octopus bimaculoides*) was the first cephalopod genome to be sequenced (Albertin et al., 2015). While this was a landmark accomplishment, the genome assembly is highly fragmented, and thus problematic for in-depth transcriptional and single-cell analysis. As the genomes of more cephalopod species are being sequenced, comparative genomics studies will have improved power, thus it is essential to have high-quality genome assemblies and annotations (i.e., Li et al., 2019, Zhang et al., 2021). Here, we describe a de-novo assembly of a high-quality, near chromosome-level *O. bimaculoides* genome. The genome was sequenced to 28x coverage using Pacific Bioscience's (PacBio)

HiFi long read technology and subsequently scaffolded with Hi-C sequencing. The re-assembled *O. bimaculoides* genome is 2.56 Gb with a scaffold N50 of 2,761,593 bp from a total of 386.65 Gb unique PacBio HiFi data. To generate gene annotations, we mapped existing bulk RNA-sequencing data and new PacBio Iso-Seq reads onto this genome assembly. Using our novel genome, we conducted a comparative phylogenomics study to understand lineage specific evolution of octopuses and more generally, cephalopods. We examined the repeat landscape, identified novel *O. bimaculoides* genes, and have begun to infer whether any specific gene families played a role in developing the octopus's unique phenotypes.

84W Exploiting Natural Variation to Understand the Role of Mkt1p in Post-Transcriptional Gene Regulation *Crystal Crook*[1,2], Emma Cox[1], Jeffrey Lewis[1] 1) University of Arkansas, Department of Biological Sciences; 2) University of Arkansas, Cell and Molecular Biology

All organisms experience stress and have thus evolved sophisticated regulatory programs to coordinate stress defense. While traditionally research has focused on transcriptional responses to stress, substantial evidence suggests that cells also possess well-coordinated post-transcriptional responses. We have been taking advantage of natural variation in the Saccharomyces cerevisiae ethanol response to understand mechanisms underlying post-transcriptional stress defense, using the protein Mkt1p as a case study. Mutations in MKT1 are pleiotropic, affecting a number of traits including various stress tolerances. Additionally, we previously found that a polymorphism in MKT1 was responsible for a large number (>500) of differences in ethanol-responsive gene expression between laboratory and wild yeast strains. Because Mkt1p shares homology with RNA-binding proteins instead of transcription factors, and co-localizes with cytoplasmic processing (P) bodies known to regulate translation during stress in a strain dependent manner, we hypothesize that Mkt1p is a global post-transcriptional regulator. We have been employing proteomics and RNA immunoprecipitation in laboratory and wild yeast to better understand how Mkt1p may regulate expression post-transcriptionally during stress. We have found that sequence variation in Mkt1p results in large differences in protein interaction partners in wild versus laboratory yeast strains, thus offering a potential explanation for the dramatic pleiotropic effects of the MKT1 polymorphism.

85W Gene regulation, environmental adaptation, and parallel expression divergence in *Mus musculus domesticus* *Sylvia Durkin*, Mallory Ballinger, Michael Nachman University of California, Berkeley, Berkeley, CA

Gene regulatory divergence has long been appreciated as an important driving force of adaptive evolution. This divergence can be in *cis*-regulatory elements, which act locally, in an allele-specific manner, or *trans*-regulatory elements, which act distally, potentially affecting many genes. While the relative contribution of *cis*- and *trans*-acting regulation to expression divergence is a well-studied topic in evolutionary biology, few studies have investigated the evolution of gene regulatory mechanisms within the context of parallel adaptation. Here we use RNAseq of two wild-derived house mouse lines that independently adapted to northern, temperate environments and compare these to one southern, tropical-adapted line to address several open questions relating to gene regulatory evolution. To what extent are gene expression changes shared between cases of parallel adaptation? What is the relative contribution of *cis*- and *trans*-changes to adaptive evolution? Are *cis*- or *trans*- changes more likely to contribute to parallel expression divergence? This study in house mice represents one of the shortest divergence times to be investigated for the relative amount of *cis*- and *trans*-acting regulation (~500 generations), allowing us to understand the gene regulatory mechanisms involved in adaptive evolution at the earliest stages of divergence.

86T The genetic control of rapid genome content divergence in *Arabidopsis thaliana* *Christopher Fiscus*, Daniel Koenig University of California, Riverside

Repetitive sequences compose the majority of eukaryotic genomes and vary widely in both copy number and sequence content. Attempts to understand the evolution of repetitive sequences and their role in shaping genome content variation have been hindered by the lack of high-quality genome assemblies with resolved repeats. Here, we employ a novel K-mer based approach to study variation in genome content in 1,142 resequenced *Arabidopsis thaliana* genomes. We use our approach to study repeat evolution in this species and identify hundreds of repetitive sequences with copy number variation. We then map the genetic basis of this variation to both cis and trans-acting loci using genome wide association (GWAS). Furthermore, we use a meta-GWAS approach to identify loci associated with copy number variation of different types of repeats, suggesting that these loci are involved in modulating the rate of genome content evolution. Finally, we show that loci associated with repeat copy number change are under strong purifying selection with minor alleles largely associated with copy number decrease. Overall, our work provides insight into the genetic basis of genome content variation in plant genomes.

87T Eukaryote-wide survey suggests unified proximate and ultimate models of *de novo* intron creation *Landen Gozashti*[1,6], Scott Roy[2], Bryan Thornlow[3,5], Alexander Kramer[3,5], Manuel Ares Jr.[4], Russell Corbett-Detig[3,5] 1) Department of

Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Department of Biology, San Francisco State University, San Francisco, CA; 3) Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA; 4) Department of Molecular, Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA; 5) Genomics Institute, University of California Santa Cruz, Santa Cruz, CA; 6) Museum of Comparative Zoology, Harvard University, Cambridge, MA

The causes for the massive variation in intron numbers across eukaryotes have been debated for decades but remain poorly understood. Rapid intron loss and/or gain in some lineages contrasts with stasis in other lineages. Episodic intron gain could be explained by specialized transposons called Introners, if Introners are present in only a subset of organisms. Systematic searches for Introners across all available eukaryotic genomes identified 27,563 Introner-derived introns from 548 distinct families, with Introners found in 175/3325 (5.2%) of studied genomes. Unexpectedly, marine organisms were 6.5 times more likely to contain Introners, and 74% of Introner-containing marine genomes harbored multiple distinct Introner families. Overrepresentation in marine organisms could reflect higher rates of lateral gene transfer. While we find that Introners are efficiently splice, preferential presence in lowly-expressed genes suggests that new insertions are costly. Most Introner families exhibit one or more signatures of DNA-based propagation. However, observations of families with partial or absent signatures suggest possible additional alternative mechanisms. These results suggest unified proximal and ultimate causes of intron gain, with susceptibility to acquire weakly deleterious Introners by lateral gene transfer playing a major role in a taxon's tendency to gain introns.

88W Island-specific evolution of a sex-primed autosome in the planarian *Schmidtea mediterranea* *Longhua Guo*[1,6], Joshua Bloom[1,6], Daniel Dols Serrate[2], James Boocock[1,6], Eyal David[3], Olga Schubert[1,6], Kaiya Kozuma[1], Katarina Ho[1], Emily Warda[1], Clarice Chui[1], Yubao Wei[4], Daniel Leighton[1,6], Tzitziki Lemus Vergara[1,6], Marta Riutort[2], Alejandro Sánchez Alvarado[5,6], Leonid Kruglyak[1,6] 1) Department of Human Genetics, Department of Biological Chemistry, and University of California Los Angeles, CA, USA; 2) Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, and Institut de Recerca de la Biodiversitat (IRBio), Barcelona, Catalonia, Spain; 3) Department of Biochemistry and Molecular Biology, The Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Ein Kerem, Jerusalem, Israel; 4) Institute of Reproductive Medicine, Henan Provincial People's Hospital, Zhengzhou University, Zhengzhou, China; 5) Stowers Institute for Medical Research, and, Kansas City, MO, USA; 6) Howard Hughes Institute for Medical Research, Chevy Chase, MD, USA

The sexual strain of the planarian *Schmidtea mediterranea* is a hermaphrodite indigenous to Tunisia and several Mediterranean islands. Here, we isolated individual chromosomes and used sequencing, Hi-C and linkage mapping to assemble a chromosome-scale genome reference. The linkage map revealed an extremely low rate of recombination on chromosome 1. We confirmed suppression of recombination on chromosome 1 by genotyping of individual sperm and oocytes. We showed that previously identified genomic regions that maintain heterozygosity even after prolonged inbreeding comprise essentially all of chromosome 1. Genome sequencing of individuals isolated in the wild indicated that this phenomenon has evolved specifically in populations from Sardinia and Corsica. We found that most known master regulators of the reproductive system are located on chromosome 1. We used RNA interference to knock down a gene with haplotype-biased expression and observed that this led to the formation of a more pronounced female mating organ. Based on these observations, we propose that chromosome 1 is a sex-primed autosome primed for evolution into a sex chromosome.

89W Evolution of the WRKY gene family in Metrosideros *Maryam Hadi*, Anne Villacastin, Elizabeth Stacy University of Nevada, Las Vegas

The WRKY gene family is primarily a green lineage-specific group of genes that encode regulators involved in growth, development, and response to various biotic and abiotic stress. Variation in the structure and arrangement of WRKY genes has important implications for their roles in stress response and modulation in other gene regulatory networks. The WRKY gene family is well characterized in model species such as *Arabidopsis thaliana* and *Oryza sativa*, however, it has not yet been examined in an island adaptive radiation.
Metrosideros (Myrtaceae) is the dominant woody genus in the Hawaiian Islands and a model island adaptive radiation. This monophyletic group comprises >20 taxa or morphotypes that are non-randomly distributed across the islands' heterogeneous landscape and show heritable phenotypes and evidence of differential local adaptation. Recent studies have produced a high-quality chromosome-level reference genome assembly for *Metrosideros polymorpha* var. *incana and shown* divergent selection as a primary mechanism of diversification within the group. We posit that selection also drives variation in the structure and arrangement of WRKY genes across taxa. Using published reference genomes for *M. polymorpha* var. *incana, M. polymorpha* var. *glaberrima*, the *Eucalyptus grandis (also Family Myrtaceae),* and the

model species, *A. thaliana, we are performing* a genome-wide search to identify, classify and determine the chromosomal localization of WRKY genes. Further, we will investigate the origin of gene duplication and loss, the mechanisms behind genomic rearrangements, type of selection, and homologous regions of the WRKY gene family across these genomes. This study aims to find a relationship between the evolution of the WRKY gene family and differential local adaptation of taxa within an island adaptive radiation. This information will further our understanding of the molecular mechanisms underlying stress tolerance and adaptive divergence in trees.

90T Evolutionary dynamics of stress-activated mobile elements in *Mimulus guttatus* Lauren Hamm University of California, Berkeley

To effectively cope with projected warmer and drier growing seasons, plant populations must rapidly adapt to novel conditions through phenotypic evolution or acclimate to environmental shifts through phenotypic plasticity. Adaptation across generations requires populations to acquire heritable genetic variation over short timescales. Although most studies have focused on base pair changes (i.e. single nucleotide polymorphisms), their mutation rates are relatively low compared to transposable elements (TEs), which may self-replicate and generate new insertion variants across the genome in relatively rapid bursts. These TE proliferation events may even be spurred by stress-activated epigenetic derepression, further increasing their potential value as sources of new genetic variation for phenotypic adaptation to worsening climate extremes and oscillations. One expression of these ideas, the "TE-Thrust hypothesis", predicts that eukaryotic lineages which tolerate short term decreases in fitness brought on deleterious TE insertions may also experience beneficial insertions that fuel adaptive evolution during periods of stress, but whether this mechanism bears out in nature has not been well studied. My research will focus on testing this hypothesis in the common monkeyflower, *Mimulus guttatus*, a widespread wildflower species that has adapted to thrive in diverse habitats across its range. Expanding empirical and computational methods we recently applied to discover environment-associated genetic variants in one part of the range (Colicchio & Hamm et al. 2021) to a broader survey and by examining TE expression in response to drought and heat, we will gain insight into TE family dynamics over space and time, elucidate patterns of stress activation, and identify evidence that natural TE variation mediates genomic plasticity. Determining the prevalence of positive selection on TEs in closely related lineages could shed light on the potentially conserved capacity of plants to domesticate TEs from selfish, nucleic acid parasites into mutualists that promote host fitness and facilitate adaptive evolution.

91T The genome of shepard's purse (Capsella bursa-pastoris) and the genetic basis of extreme cosmopolitanism *Daniel Koenig*[1], Chris Fiscus[1], Christina Wesse[3], Max Collenberg[2], Barbara Neuffer[3], Detlef Weigel[2], Christa Lanz[2], Rebecca Schwab[2] 1) University of California Riverside; 2) Max Plank Institute for Biology Tübingen; 3) University of Osnabrück

The allopolyploid Capsella bursa-pastoris is one of the world's most successful plant species being found in abundance on six of seven continents across a wide diversity of environments. Shepard's purse only recently spread to much of this distribution having been brought to the Americas and Australia by colonizers in the last few hundred years. In contrast, the diploid progenitors of C. bursa-pastoris reside in relatively constricted ranges in Eurasia. Here, we explore the birth and evolution of Capsella bursa-pastoris using several high quality genome assemblies from it and it's diploid relatives. We that varied patterns of genome evolution in the diploid progenitors have continued to shape genomic diversity in the polyploid species and may predict the outcome of ongoing diploidization. Furthermore we characterize genomic and phenotypic diversity in hundreds of shepherds purse samples drawn from across the globe. We find that the wide range of climate environments occupied by this species is partly explained by strongly segregated ecotypic diversity. The two dominant ecotypes remain segregated across climates even after recent colonization of the Americas. However, hybrid zones found in intermediate climates facilitate mapping of adaptive loci. Our work links standing genetic variation to extremely rapid colonization of new habitats and advances C. bursa-pastoris as a model system for studying the contribution of molecular variation to climate adaptation.

92W Rodents of Unusual Sperm: Molecular and Phenotypic Evolution of Male Reproduction in Murine Rodents *Emily Kopania*[1,2], Gregg Thomas[1,3], Carl Hutter[4], Sebastian Mortimer[1], Colin Callahan[1], William Breed[5], Nathan Clark[2], Jacob Esselstyn[4], Kevin Rowe[6], Jeffrey Good[1] 1) University of Montana; 2) University of Utah; 3) Harvard University; 4) Louisiana State University; 5) University of Adelaide; 6) Museums Victoria

Male reproductive traits often evolve extremely rapidly, likely due to post-mating sexual selection. In parallel, genes expressed in reproductive tissues tend to diverge rapidly in protein-coding sequence, which is often attributed to positive selection. However, few studies have connected rapid phenotypic and molecular evolution for reproductive traits or directly tested the role of positive selection underlying divergence. Doing so requires species that vary in reproductive phenotypes, a well-resolved phylogeny, and information on genes expressed in reproductive tissues. Murine rodents

provide an ideal system for studying reproductive evolution because they include the well-studied mouse and rat model systems and also represent a massive radiation (>10% of mammal species) with a remarkable diversity of reproductive phenotypes among species. In this study, we sequenced exomes from 209 murine species to infer a well-resolved phylogeny and modeled male reproductive trait evolution across this phylogeny for 96 species. We showed that relative testes mass, a proxy for sperm competition, was evolving independently of phylogeny. Most murine species have a hook on the sperm head, and our analyses showed that hook length and angle were correlated with relative testes mass, suggesting that these traits are likely co-evolving in response to sperm competition. We found that genes expressed in the seminal vesicles and during the later (post-meiotic) stages of spermatogenesis tended to be the most rapidly evolving among reproductive genes, and that this rapid evolution was due in part to more frequent positive directional selection. We also used these data to examine if some genes evolve rapidly in lineages with convergent reproductive traits such as large testes mass to directly connect the rapid molecular and phenotypic evolution of male reproduction. We tested if reproductive genes experienced episodic bursts of positive directional selection in lineages with elevated divergence in reproductive traits. Collectively, our results demonstrate that sperm competition and positive selection likely play a central role in the rapid phenotypic and molecular evolution of male reproductive traits in murine rodents, but the intensity of these forces and their importance in shaping evolutionary patterns is highly variable across traits, tissues, and species.

93W Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness *Amanda Kowalczyk*[1], Maria Chikina[2], Nathan Clark[3] 1) Carnegie Mellon University, Pittsburgh, PA; 2) University of Pittsburgh, Pittsburgh, PA; 3) University of Utah, Salt Lake City, UT

Although hair is a defining mammalian characteristic, several mammals, such as elephants, naked mole-rats, and humans, have substantially reduced hair coverage. In these so-called “hairless” mammals, their limited hair differs both in quantity and localization from other mammals, suggesting potential roles of gene regulation and gene sequence changes underlying phenotypic diversity. Thus, the hairless phenotype is a dynamic system in which to study interplays in evolution of coding and regulatory sequence.
Elucidating these genotype-phenotype relationships is a key question in biology. One way to answer this question is a computational method called RERconverge that performs a genome-wide scan to find associations between evolutionary rates of genetic elements and convergently-evolving traits. As similar traits evolve independently in different species, similar selective pressure shifts in genetic sequence can accompany and drive phenotypic changes. RERconverge detects these selective pressure shifts as quantified by shifts in evolutionary rate.
RERconverge was used to identify the genetic basis of reduced hair quantity in mammals using sequence for protein-coding genes and approximately 350,000 noncoding regions. RERconverge successfully identified genes and noncoding regions associated with keratinization, cornification, and hair-related mouse knockout phenotypes. A skin- and hair-associated microRNA, MIR205, was likewise found to have a significant enrichment of quickly evolving noncoding elements in its vicinity. RERconverge also identified genes and non-coding elements associated with changes in skin, many of which may be complementary to hair loss, such as changes in genomic regions related to UV light DNA damage repair. Other genes detected remain poorly functionally characterized, and nearly all noncoding regions remain untested. These regions represent good candidate genes and regulatory elements to test experimentally for association with hair growth. Despite the similar functional categories identified in both coding and noncoding sequence analyses, precise regions identified were strikingly different. Most genes that showed significant evolutionary rate shifts associated with hairlessness did not show the same in their nearby noncoding regions and vice-versa. Such findings suggest a complementary interplay between evolution of regulatory and protein-coding regions underlying mammalian hairlessness.

94T Rice chromosome architecture at nucleosome-level resolution *Amina Kurbidaeva*, Michael Purugganan New York University

The three-dimensional structure of genome plays an integral role in gene expression regulation. Chromosomes are partitioned into compartments, characterized by active or repressive chromatin states. Topologically associated domains (TADs) are defined as regions of high chromatin interconnectivity and restrict enhancer-promoter crosstalk maintaining proper gene expression and co-regulation with the domain. In addition, chromatin loops bring into proximity distant regulatory parts of the genome, such as enhancers, promoters, and tethering elements. These three levels of genome organization can be best characterized by Micro-C, a Hi-C-based method in which micrococcal nuclease is used instead of restriction enzymes to fragment chromatin, enabling nucleosome resolution chromosome folding maps. For the first time to our knowledge, we applied Micro-C to study the 3D genome organization of a plant genome, using Oryza sativa. This allowed us to look into the genome structure of rice with an unprecedented resolution, providing a basis for future comparative genomics studies. An understanding of 3D genome organization will allow us to explain regulatory evolution processes shaping speciation and domestication within the Oryza genus.

95T Development across evolutionary time at a single cell resolution in the *Caenorhabditis* nematodes *Christopher Large*[1], Rupa Khanal[2], Qin Zhu[3], Priya Sivaramakrishnan[1], Felicia Peng[1], Erik Nordgren[2], Jean Rosario[2], Junhyong Kim[2], John Murray[1] 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Department of Biology, University of Pennsylvania, Philadelphia, PA; 3) Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA

Complex gene regulatory networks specify the development of all multicellular organisms and determine the morphological complexity of life. What determines the rate of change and the evolutionary constraint of cellular gene expression patterns across development remains a fundamental question of biology. Previous measurements of morphology and gene expression across developmental stages in whole organisms have led to the establishment of the hourglass model which posits that the developmental plan is under the highest constraint during the midpoint of development. However, the relevance of these observations for individual cell and tissue types is unclear, harming the predictive power of the model and leaving open questions as to the model's biological underpinnings. Historically, the ability to systematically profile homologous cells across different organisms for their function and transcriptional profile has been limited by technology, however single cell sequencing has facilitated the ability to capture and molecularly label individual cells through microfluidics. Each cell's identity, developmental timing, and physical location in the organism can then be determined by comparing the cell's transcriptional profile to reference developmental markers with known expression patterns. Using single cell sequencing, we are currently measuring the spatiotemporal divergence of gene expression across evolutionary time within the *Caenorhabditis* nematodes by comparing the transcriptomes of homologous cells and tissues between species across embryonic development. Altogether, we will elucidate the constraints on gene regulatory network evolution by building and comparing molecular atlases of development for >10 *Caenorhabditis* species. We plan to use the multispecies single cell-atlases to determine the rate of divergence of cell and gene expression patterns across the *Caenorhabditis* phylogeny, providing insight into how new cell types are born, the fate of recent gene duplications, and how the life histories of the *Caenorhabditis* species influence their evolution.

96W Investigation of convergent evolution with the southern marsupial mole and other subterranean mammals *Sarah Lucas*[1], Stephen Frankenberg[2], Charles Feigin[3], Andrew Pask[2], Nathan Clark[1] 1) University of Utah, Salt Lake City, UT, USA; 2) University of Melbourne, Parkville, Victoria, Australia; 3) Princeton University, Princeton, NJ, USA

Despite it being millions of years since the rare southern marsupial mole (*Notoryctes typhlops*) shared a common ancestor with any of the fossorial placental mammals, these species share several physiological adaptions for tunneling: tubular bodies, enlarged forepaws, reduced vision, and likely a tolerance for hypoxic conditions. When species exhibit similar phenotypes, these are often accompanied with similar genomic changes. Our group is specifically interested in what changes resulted in the species' severely degraded eye phenotype. For unlike the other subterranean placental mammals, this species lacks both a pupil and an optic nerve. Eye development is a highly conserved and regulated developmental process. To determine what genes are undergoing convergent relaxed evolution within the marsupial mole and other subterranean species, we used RERconverge, an R statistical package. By adding two additional subterranean species to our analysis (marsupial mole and Damaraland mole-rat), this will increase the power for detecting genes associated with degraded eyesight while burrowing. We will also be able to characterize novel changes found only within the marsupial mole. As one of the oldest fossorial mammals, this will allow for more inactivating mutations to fixate within this lineage generating a stronger relative evolutionary rate (RER) signal within the species. With the increased statistical power, some genes associated with vision which were nearly significant in the previous analysis will likely become significant. Knowledge of potentially inactivating substitutions will be useful in trying to diagnose individuals with rare congenital eye malformations.

97W Ultra-accurate sequencing unveils early somatic lineage selection in *Drosophila melanogaster* *Jakob McBroome*, Evan Pepper, Cade Mirachandi, Russell Corbett-Detig University of California, Santa Cruz

Multicellular organisms are often thought to be genetically homogenous, but their constitutive cells acquire distinct mutations during growth and development. Variation among somatic lineages allows natural selection to act within the body, influencing the genetic composition of the maturing organism. This composition, known as somatic mosaicism, in turn has a significant potential impact on phenotype. Despite its importance, the nature and strength of selection on somatic lineages throughout development is largely unknown. To address this question, we present an ultra-accurate sequencing technique and accompanying analytical pipeline to identify low-frequency mutations within a single tissue sample. We applied this technique to whole *Drosophila melanogaster* individuals and developed a statistical framework to evaluate the distribution of low-frequency somatic mutations. From the resulting somatic allele frequency spectrum, we inferred a distribution of somatic fitness effects and discovered evidence that ribosomal genes and other metabolic

genes are exceptionally conserved, in line with previous studies on the mechanisms of cell competition. Our results suggest that purifying selection plays a significant role in the fate of early developmental mutations and somatic mosaicism overall in *Drosophila melanogaster.*

98T Hybridization underlies localized trait evolution in cavefish *Suzanne McGaugh*[1], Rachel Moran[1], James Jaggard[2,3], Emma Roback[1], Alexander Kenzior[4], Nicolas Rohner[4,5], Johanna Kowalko[6], Patricia Ornelas-Garcia[7], Alex Keene[8] 1) 1. Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN; 2) 2. Department of Biological Sciences, Florida Atlantic University, Jupiter FL; 3) 3. Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 4) 4. Stowers Institute for Medical Research, Kansas City, MO; 5) 5. Department of Molecular & Integrative Physiology, KU Medical Center, Kansas City, KS; 6) 6. Department of Biological Sciences, Lehigh University, Bethlehem, PA; 7) 7. Departamento de Zoología, Instituto de Biología, Universidad Autónoma de México, CP 04510, Mexico City, Mexico; 8) 8. Department of Biology, Texas A&M University, College Station, TX 77843

Introgressive hybridization often drives patterns of phenotypic evolution and may play an integral role in the evolutionary processes of local adaptation and speciation. In the Mexican tetra *Astyanax mexicanus*, cave populations have repeatedly evolved traits including eye loss, sleep loss, and albinism. Of the 30 caves inhabited by *A. mexicanus*, Chica cave is unique because it contains multiple individual pools inhabited by putative hybrids between surface and cave populations, providing an opportunity to investigate the impact of hybridization on complex trait evolution. We provide evidence that hybridization between cave and surface populations may contribute to localized variation in traits associated with cave evolution, including pigmentation, eye development, and sleep. We also uncover an example of convergent evolution in a circadian clock gene in multiple cavefish lineages and burrowing mammals, suggesting a shared genetic mechanism underlying circadian disruption in subterranean vertebrates. Our results provide insight into the role of hybridization in facilitating phenotypic evolution.

99T Identifying molecular evolutionary rate shifts accompanying dietary transitions in mammals *Wynn Meyer*[1], Kathleen Foley[1,2], Simon Gajewski[1], Oluwafunmbi Ojo[1], Irene Kaplow[3], Daniel Schäffer[3], Maria Chikina[4], Andreas Pfenning[3], Zoonomia Consortium 1) Lehigh University, Bethlehem, PA; 2) University of Iowa, Iowa City, IA; 3) Carnegie Mellon University, Pittsburgh, PA; 4) University of Pittsburgh, Pittsburgh, PA

Mammals display striking contrasts in diet, particularly between species that eat other animals almost exclusively and those that predominantly eat plants. These dietary specializations have evolved multiple times independently within the mammalian phylogeny, providing an opportunity to identify molecular evolutionary signatures associated with these convergent phenotypes. Such signatures can implicate genes involved in mammalian adaptation to distinct diets. We here implement a robust computational method to identify associations between relative rates of molecular evolution in coding sequences and dietary phenotype across branches of the mammalian phylogeny, representing 71 strictly carnivorous or herbivorous species. We test for these associations at 14,883 protein coding genes for which high quality alignments are available from the Zoonomia Project. Overall, we find that substantially more genes evolve rapidly in herbivorous lineages relative to carnivorous lineages than the opposite pattern (evolving more rapidly in carnivorous lineages). By implementing a false discovery rate correction for multiple testing, we find strong correlations between diet and the relative evolutionary rates of multiple large axonemal dynein heavy chain genes, driven by rapid evolution among many members of fully herbivorous clades. When we implement phylogenetic simulations and permutations to account for non-independence in phenotype throughout the tree, we find a different set of genes as most strongly associated with dietary phenotype, driven by large convergent shifts in molecular evolutionary rates on internal branches representing phenotype transitions. Among these results, we find an enrichment of more rapid evolution in herbivores among genes expressed in the human thyroid and brain, specifically the cerebellar hemisphere and frontal cortex. Two of the genes with the strongest phylogeny-corrected signatures of rapid evolution in response to herbivory have previously been implicated in adipogenesis. While relaxation of selective constraint may explain some convergent shifts in evolutionary rates, particularly for genes evolving rapidly in all lineages in herbivorous clades, the strong signatures of convergent rate shifts on internal branches more likely represent changing fitness optima in response to dietary change. The genes we have identified with these signatures thus represent strong candidates for targets of adaptation to herbivorous diets in mammals.

100W A *Drosophila* Reference Pangenome Graph *Cade Mirchandani*[1], Russell Corbett-Detig[1,2] 1) Department of Biomolecular Engineering, University of California Santa Cruz. Santa Cruz, CA 95064, USA; 2) Genomics Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA

Although linear reference genomes are a powerful resource for population and evolutionary genomics, each represents only a fraction of the diversity in most natural populations. The Drosophila Genome Nexus is a population genomics re-

source composed of high quality D. melanogaster genome assemblies from many different geographic regions. However, comparative analysis of Drosophila genomes are constrained by the use of a single reference genome. Here, we present a significant improvement to the resource by integrating all of the genomes from the DGN into a reference pangenome graph. We align all previous DGN genomes, along with newly published genomes to the reference genome graph using a singular workflow to allow comparison across datasets. Using our reference genome graph, we observe significantly increased alignment sensitivity when compared to mapping to a linear reference genome. Furthermore, we show that genotype calls for highly variable isolates are unbiased when aligned to the pangenome graph. The addition of the reference genome graph to this resource will allow users to leverage known wild variation in alignments and downstream analysis of their own data. These improvements to this resource will help address the broad range of population genetic questions in D. melanogaster and provide a roadmap for improving variant calls in other highly variable genomes.

101W Natural selection and correlated landscapes of diversity in the great apes *Murillo Rodrigues*[1], Peter Ralph[1,2,3], Andrew Kern[1,3] 1) Institute of Ecology and Evolution, Department of Biology, University of Oregon; 2) Department of Mathematics, University of Oregon; 3) contributed equally

Levels of genetic diversity along chromosomes – or landscapes of diversity – are correlated between related groups of species, such as flycatchers, monkeyflowers, and aspens. Even though these correlations should decay with divergence time under neutrality, empirical data demonstrate strong correlations persisting over long evolutionary timescales. What could be maintaining strong correlations in landscapes of diversity across pairs species? Natural selection is known to couple genomic features, such as recombination rate and density of functional sites, with levels of diversity. These features are expected to be largely shared between related taxa, so selection and its linked effects could in principle maintain correlations over longer periods of time. Using the great apes as a model system, we investigated whether landscapes of diversity are correlated in this group. Using forward-in-time simulations of the great apes' evolutionary history under different selective regimes (neutral, background selection, sweeps), we explore how natural selection shapes correlations over time. Further, we test whether variation in mutation rate along chromosomes can generate patterns like the data. We found high correlations among the great apes' landscapes of diversity, even between pairs of species that diverged over 10Mya. Forward-in-time simulations of the great apes' history show that the observed correlations far exceed what can be generated under a neutral null model. In simulations with selection and a realistic distribution of functional sites, we found that correlations between landscapes of diversity might be maintained over longer periods of time. Future work includes simulation-based inference of selection parameters in our model.

102T *Saccharomyces cerevisiae* on the rise: Characterizing domestication of *S. cerevisiae* for breadmaking *Margot Ruffieux*, Alexxis Gutierrez, Nathan Brandt, Caiti Heil North Carolina State University

The production of leavened bread is an important aspect of culture and nutrition worldwide and has been produced since at least the second millennium BCE. Despite bread's integral role in society, the evolutionary history of the organisms central to breadmaking remain largely unexplored. *Saccharomyces cerevisiae* is the most common yeast in sourdough starters and is used exclusively in commercial bread baking. Breadmaking presents a number of selection pressures for yeast, including osmotic stress, metabolism of complex sugars, selection for particular aromas and tastes, and competition with a mixed microbial community. To understand how *S. cerevisiae* has adapted to these conditions, we have isolated and sequenced 47 *S. cerevisiae* strains from sourdough starters collected as part of a citizen science project from home-bakers across North America. We combine our sourdough strain genome sequencing with sequences representing a wide variety of fermentation and natural isolation sources to understand how adaptation to human-associated environments has shaped the genomes of *S. cerevisiae*. We characterize the population structure of *S. cerevisiae* found in sourdough starters and identify genomic signatures indicative of domestication in breadmaking yeast. We pair these population genomic analyses with phenotypic data collected on our sourdough strains and a selection of commercial baking strains, beer, sake, wine, and natural isolates to quantify dough rise, growth in the complex sugar maltose (prevalent in dough), and other baking traits. This study elucidates genetic and phenotypic traits under selection in *S. cerevisiae* as it has adapted to the unique bread dough environment.

103T Evolved genetic variation due to epistatic mitochondrial-nuclear interactions *Samantha Sierra-Martinez*, Douglas Crawford, Marjorie Oleksiak University of Miami

One explanation for large standing genetic variation and the diversity of nucleotide polymorphisms responding to similar evolutionary pressures is epistasis. If the adaptive value of a nucleotide polymorphism is context dependent (dependent on genetic background), then more than one adaptive solution is likely, providing a mechanism to maintain standing genetic variation. That is, the interactions among nucleotide polymorphisms define their adaptive value, which translates into more than one adaptive solution. One of the more approachable genetic interactions occurs between the two

cellular genomes: nuclear and mitochondrial. These two genomes define the 96 proteins that interact within the oxidative phosphorylation pathway to produce most cellular ATP. These interactions could affect ATP production, altering an individual's fitness. To examine mito-nuclear interactions, I analyzed nucleotide polymorphisms among pairs of polluted and non-polluted populations, where there is an adaptive divergence between populations. There are three predictions: 1) population pairs will differ in the frequencies of mito-haplotypes, 2) population pairs will have significant allele frequency differences in nuclear OxPhos genes, and 3) within a population between mitochondrial haplotypes significant differences will occur in the nuclear polymorphism frequencies. These data suggest that adaptation involves mitochondrial genomic changes that alter the adaptive importance of nuclear polymorphisms.

104W Gene expression and functional co-evolution in the Drosophila female reproductive tract *Rachel Thayer*, David Begun University of California, Davis

One goal of evolutionary genetics is to understand the mechanisms and drivers of gene expression evolution and innovation. In Drosophila melanogaster, the male reproductive tissues that produce ejaculate exhibit rapid sequence and gene expression evolution, as well as high rates of genetic novelty. Sexual conflict over the optimal outcomes of female-ejaculate interactions is one favored explanation for this accelerated evolution. Importantly, this sexual conflict model predicts strong female co-evolution, which we sought to test. Using flies from high and low latitude populations in the classic North American cline, from which male gene expression has been previously characterized, we assayed gene expression divergence across five female tissues that are directly exposed to ejaculate products. To investigate female-ejaculate co-evolution, we measured post-mating gene expression in females with either a same-population or a diverged mate. If females and males are co-adapted, then mismatched matings should show aberrant gene expression. We are also working to characterize female reproductive gene expression at single-cell resolution for the first time, and to interpret functional gene expression divergence in this cellular context.

105W Genomic signatures of desert adaptation at gene-rich regions in zebu cattle from the African drylands *Abdulfatai Tijjani*[1,2,3,4], Bashir Salim[5], Marcos Vinicius Barbosa da Silva[6], Hamza A. Eltahir[7], Taha H. Musa [8], Karen Marshall[9,10], Olivier Hanotte[2,3,4], Hassan H. Musa[6,11] 1) The Jackson Laboratory, Bar harbor, ME; 2) International Livestock Research Institute (ILRI) PO 5689, Addis Ababa, Ethiopia; 3) Centre for Tropical Livestock Genetics and Health (CTLGH), ILRI Ethiopia, PO Box 5689, Addis Ababa, Ethiopia; 4) Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, United Kingdom; 5) Faculty of Veterinary Medicine, University of Khartoum, Sudan; 6) Embrapa Gado de Leite, Juiz de Fora, Brazil; 7) Institute of Molecular Biology, University of Nyala, Sudan; 8) Biomedical Research Institute, Darfur College, Sudan; 9) International Livestock Research Institute (ILRI) PO Box 30709, Nairobi 00100, Kenya; 10) Centre for Tropical Livestock Genetics and Health (CTLGH), ILRI Kenya, P.O. Box 30709, Nairobi 00100, Kenya; 11) Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan

Background
Sudan, the largest country in Africa, acts as a corridor between North and sub-Saharan Africa along the river Niles. It comprises warm arid and semi-arid grazing lands, and it is home to the second-largest African population of indigenous livestock. Indigenous Sudanese cattle are mainly indicine/zebu (humped) type. They thrive in the harshest dryland environments characterised by high temperatures, long seasonal dry periods, nutritional shortages, and vector diseases challenges. We investigated genome diversity in six indigenous African zebu breeds sampled in Sudan (Aryashai, Baggara, Butana, Fulani, Gash, and Kenana). We adopted three genomic scan approaches to identify candidate selective sweeps regions (*ZHp*, $F_{ST,}$ XP-EHH).
Results
We identified a set of gene-rich selective sweep regions shared across African and Asian zebu or unique to Sudanese zebu. In particular, African and Asian zebu candidate gene-rich regions are detected on chromosomes 2, 5, and 7. They include genes involved in immune response, body size and conformation, and stress response to heat. In addition, a 250 kb selective sweep on chromosome 16 was detected exclusively in five Sudanese zebu populations. This region spans seven genes, including *PLCH2*, *PEX10, PRKCZ,* and *SKI,* which are involved in alternative adaptive metabolic strategies of insulin signalling, glucose homeostasis, and fat metabolism.
Conclusions
Together, these genes may contribute to the zebu cattle resilience to heat, nutritional and water shortages. Our results highlight the putative importance of selection at gene-rich genome regions, which might be under a common regulatory genetic control, as an evolutionary mechanism for rapid adaptation to the complexity of environmental challenges.

106T Genetic structure and multiple paternity in invasive Red Swamp Crayfish in southeastern Michigan, USA *Nicole Adams*[1], Jared Homola[2], John Robinson[1], Kim Scribner[1] 1) Michigan State University, East Lansing, Michigan, USA; 2) U.S.

Geological Survey, Wisconsin Cooperative Fishery Research Unit, Stevens Point, Wisconsin, USA

The Red Swamp Crayfish, *Procambarus clarkii*, is a prolific invader that has successfully colonized every continent except for Australia and Antarctica. Native to the Mississippi River Basin, the southeastern United States and northern Mexico, *P. clarkii* was first detected in southeastern Michigan in 2017. Following detection, extensive state and local resources have been expended to eradicate established populations and minimize further spread of this species. Using genetic tools we sought to characterize founding history, levels and direction of gene flow, demographic history, and the reproductive biology of *P. clarkii* in Michigan waterbodies to inform management practices. RAD capture (RAPTURE) sequencing produced 580 million paired reads across 1100 individuals from 22 infested waterbodies. Using 842 individuals from 20 locations that successfully passed our filtering criteria, we found that samples clustered largely based on geography. Even within short distances we found evidence for genetic differentiation among sampled waterbodies. We assigned parentage to offspring sampled from egg or larva-bearing females. Employing likelihood-based pedigree reconstruction methods, we found that a single female could mate with 2-6 potential males. The estimated number of breeding adults varied between waterbodies. Additional collected samples are being characterized to look at source-sink founding population dynamics and landscape barriers to gene flow. Understanding how *P. clarkii* colonize vacant waterbodies across heterogeneous and heavily human-impacted environments and characterizing fundamental aspects of their reproductive biology are critical for eradication and prevention of further spread.

107T Putative drought-adapted SNPs increase in frequency during severe drought *Daniel Anstett*[1,2], Julia Anstett[2], Dylan Moxley[2], Mojtaba Jahani[2], Kaichi Huang[2], Marco Todesco[2], Rebecca Jordan[3], Loren Rieseberg[2], Amy Angert[2] 1) Michigan State University, Lansing, MI; 2) University of British Columbia, Vancouver, BC; 3) CSIRO, Land and Water, Sandy Bay, Tasmania, Australia

Genotype-environment associations (GEA) allow for the identification of variants and genomic regions that may provide possible adaptations to climate change. With the advent of large landscape genomics data sets the use of GEA is becoming increasingly common. However, studies in natural populations are usually not able to observe if the variants increase in frequency as populations undergo strong selection due to climate change. Here we track the impact of the recent California mega-drought on associated SNPs in *Mimulus cardinalis*, a riparian, montane plant found across California and Southern Oregon. We carried out 8-21X whole genome sequencing on 55 populations sampled range-wide before the 2012-2015 drought. Using BayPass we identified over 20,000 SNPs associated with climate, and track their change yearly from 2010 to 2016. SNPs conferring possible adaptation to greater heat, lower precipitation and increased moisture deficit increased in frequency through the drought when compared to randomly selected neutral SNPs. We will illustrate the importance of these regions through genetic offset analyses that outline what regions of *M. cardinalis* are particularly genetically vulnerable to future climate change conditions.

108W Rapid evolution of abdominal pigmentation in *Drosophila melanogaster* *Skyler Berardi*[1], Subhash Rajpurohit[1,2], Seth Rudman[1,3], Tess Grainger[1,4], Mary Catherine Berner[1], Nicolas Betancourt[1], Paul Schmidt[1] 1) Dept. of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA; 2) Biological and Life Sciences, Ahmedabad University, Ahmedabad, India; 3) School of Biological Sciences, Washington State University, Vancouver, WA, USA; 4) Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada

Climatic conditions vary dramatically across spatiotemporal scales, and understanding how populations adapt to this heterogeneity remains a fundamental goal of evolutionary biology. Elucidating the mechanisms of polygenic adaptation is critical to our understanding of how populations evolve in response to rapid environmental change. In this study, we leveraged *Drosophila melanogaster* as a model to tease apart the roles of demography, migration, and selection on the evolutionary dynamics of a complex, fitness-associated trait: pigmentation. The extent of abdominal melanization is highly variable both within and among species of *Drosophila*; in *D. melanogaster*, it exhibits latitudinal and altitudinal clines on multiple continents, suggesting that pigmentation is an important component of local adaptation to climate. We first examined patterns of pigmentation across a latitudinal cline on the east coast of North America, which was not previously characterized. We sampled flies from six natural populations ranging from 25.28°N to 42.45°N, and we measured patterns of pigmentation after generations of culture in a common garden, laboratory environment. We found that abdominal melanization increases predictably with latitude, consistent with previously documented patterns on other continents. This supports the hypothesis that phenotypic clines are driven by local adaptation to climate. We further explored the dynamics of these putatively adaptive patterns by probing whether seasonal trends are concordant with this latitudinal trend. We sampled flies from an orchard in southeastern Pennsylvania in the spring and fall across six years. We found that populations in the spring exhibit a higher degree of abdominal melanization, which then declines predictably and in parallel across seasonal time in the six replicate years. Finally, we seeded replicate outdoor mesocosms

in southeastern Pennsylvania with an outbred panel of flies collected in the spring from local orchards across multiple years. This enabled us to further test the hypothesis that selection drives the observed seasonal patterns by eliminating the confounding effects of gene flow and cryptic population structure. The trend we had observed in the wild Pennsylvanian population was recapitulated: abdominal melanization decreased across seasonal time, and this pattern was repeated over several years in our mesocosms. Together, these findings implicate adaptive tracking as a key driver of the phenotypic response of populations to rapid environmental change.

109W Stability of the genetic structure and association with microhabitat of a wild wheat population over 36 years *Tal Dahan-Meir*[1], Thomas James Ellis[2], Fabrizio Mafessoni[1], Hanan Sela[3,4], Jacob Manisterski[4], Naomi Avivi-Ragolsky[1], Amir Raz[1,5], Moshe Feldman[1], Yehoshua Anikster[4], Magnus Nordborg[2], Avraham A. Levy[1] 1) Department of Plant and Environmental Sciences, Weizmann Institute of Science; Rehovot, Israel; 2) Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter; Vienna, Austria; 3) Institute of Evolution, University of Haifa; Haifa, Israel; 4) The Institute for Cereal Crops Improvement, Tel-Aviv University; Tel Aviv, Israel; 5) Migal, Galilee Technology Center; Kiryat Shmona, Israel

Wild progenitors of major crops can provide the genetic resources needed for ensuring food security. Long-term studies of progenitor species in their natural ecological niches are especially important for understanding the likely impacts of climate change, and inform strategies for conservation of such resources *in-situ* as well as in gene banks. We examined the genetic structure of Ammiad wild emmer wheat population which was sampled over 36 years while both temperature and atmospheric CO_2 concentration increased significantly. At each sampling, seeds were collected from plants at 100 marked locations along four linear transects traversing seven ecologically distinct microhabitats. The genotypes of 832 individuals revealed high genetic diversity over scales of tens of meters and spatial clustering of the population. We found a striking concordance between genotype' groups and ecological microhabitats, which were previously defined based on topographic and floristic data. This pattern was remarkably stable over time. Analyses and simulations indicate that neutral processes alone are unlikely to fully explain the spatial and temporal stability of the population. This suggests that natural selection, in addition to limited dispersal, contributed to shaping population structure. Our work shows that conservation of wild populations should take into account ecological niches, even in a small area, in order to best sample their diversity.

110T Genetic diversity loss in the Anthropocene *Moises Exposito-Alonso*[1,2,3], Tom Booker[4,5], Lucas Czech[1], Tadashi Fukami[2], Lauren Gillespie[1,6], Shannon Hateley[1], Christopher Kyriazis[7], Patricia Lang[2], Laura Leventhal[1,2], David Nogues-Bravo[8], Veronica Pagowski[2], Megan Ruffley[1], Jeffrey Spence[9], Sebastian Toro Arana[1,2], Clemens Weiss[9], Erin Zess[1] 1) Department of Plant Biology, Carnegie Institution for Science, Stanford, USA; 2) Department of Biology, Stanford University, Stanford, USA; 3) Department of Global Ecology, Carnegie Institution for Science, Stanford, USA; 4) Department of Zoology, University of British Columbia, Vancouver, Canada; 5) Biodiversity Research Centre, University of British Columbia, Vancouver, Canada; 6) Department of Computer Science, Stanford University, Stanford, USA; 7) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA; 8) Center for Macroecology, Evolution and Climate, GLOBE Inst., Univ. of Copenhagen, Copenhagen, Denmark; 9) Department of Genetics, Stanford University, Stanford, USA

More species than ever before are at risk of extinction due to anthropogenic habitat loss and climate change. But even species that are not threatened have seen reductions in their populations and geographic ranges, likely impacting their genetic diversity. Although preserving genetic diversity is key to maintaining adaptability of species, we lack predictive tools and global estimates of genetic diversity loss across ecosystems. By bridging theories of biodiversity and population genetics, we introduce a mathematical framework to understand the loss of naturally occurring DNA mutations within a species. Analyzing genome-wide variation data of 10,126 geo-tagged individuals from 19 plant and animal species, we show that genome-wide diversity follows a power law with geographic area, which can predict genetic diversity loss in computer simulations of population losses. Given pre-21st century values of ecosystem transformations, we estimate that over 10% of genetic diversity may be lost, already surpassing the United Nations targets for genetic preservation. These estimated losses could rapidly accelerate with advancing climate change and habitat destruction, highlighting the need for new forecasting tools that facilitate implementation of policies to protect genetic resources globally.

111T Adaptive significance of flowering time plasticity: synthesising 10 years of *Arabidopsis* research in the field. *Alexandre Fournier-Level*[1], Andhika R Putra[1], Johanna Schmitt[2] 1) The University of Melbourne; 2) UC Davis

Genetic and environmental variation combine to promote trait plasticity. Despite well-laid theoretical expectations, empirical data are still indispensable to validate the importance of plasticity for fitness and evolution in natural conditions. Here we synthesise 10 years of field experiments on *Arabidopsis thaliana*'s flowering time where hundreds of natural accessions were planted across multiple European sites and seasons.

Extensive flowering time plasticity was observed, primarily across environment, but also between accessions. The pattern of plasticity was mostly driven by ecotypes originating from marginal regions (Nordic and high elevation) and expressed over the summer, not the most common growing season for this species. Selection analysis showed that highly plastic genotypes were negatively selected in most cases, suggesting that plasticity was primarily cryptic and expressed away from home conditions. However even if selection primarily favoured highly canalized, early flowering, we did observe the selection of genotypes with increased seasonal plasticity, able to delay flowering in favourable Mediterranean fall conditions while still flowering early in the spring. Genome-wide association study identified *GIGANTEA SUPPRESSOR 5 as main* candidate gene for this plastic response, consistent with previous lab observations. Nonetheless, the massive gap between the overall very high heritability for flowering time plasticity traits and the low amount of variance explained by individual loci underscores a very polygenic architecture.
The data collected were used to train a genomic prediction model of flowering time across Europe. We show that a LASSO-penalized linear-mixed model designed using daily minimum and maximum ground temperature and a genetic kinship matrix without individual locus effects performed well with a cross-validated accuracy of 0.93. We also showed that a genotype-by-environment effect substantially improved the predictions, emphasizing the importance of capturing plasticity. This predictive model was independently validated using data from Rhode Island showing a good transferable accuracy of 0.64. These predictions were used to identify genotypes that could be used to genetically offset the predicted consequence of climate change in an environment-specific manner. We highlight the importance of multi-site testing of large numbers of genotypes from multiple origins to best design climate change-ready restoration using genomic data.

112W Genetic variation, covariation, and constraints in the evolution of sexual and clonal reproduction in a plant species *Jannice Friedman*[1], Christina Steinecke[1], Matthew Rubin[2] 1) Queen›s University; 2) Donald Danforth Plant Science Center

Plants show incredible diversity in their reproductive strategies, and a central problem for evolutionary biologists is to understand the selective forces and genetic mechanisms that are responsible for the origins and maintenance of this diversity. Many species have the ability to asexually reproduce (for example, via stolons or corms) and sexually reproduce via flowers and seed. The question, then, is why some invest resources into both, and whether a trade-off exists when investing in traits associated with one mode or the other. Trade-offs, or constraints, can arise due to genetic correlations (through pleiotropy or linked genes); and if resources are finite through resource allocation trade-offs. Here, we use the wildflower, *Mimulus guttatus (*syn. *Erythranthe guttata)* to investigate phenotypic and genetic trade-offs between vegetative and reproductive allocation. We use a series of approaches to understand the maintenance of the variation in nature, the underlying genetic architecture, and whether trait variation and covariation constrains evolutionary trajectories. First we quantify variation in natural annual and perennial populations (n=80) from across the native range of *M. guttatus* (California to Alaska) and demonstrate strong correlations between vegetative and reproductive traits. To determine whether these multi-trait patterns arise from pleiotropic or independent loci, we mapped QTLs on a cross between divergent populations, and followed this with three new mapping populations using recombinant F4 individuals, to investigate fitness and identify QTL in a common garden field experiment in British Columbia. We find extensive pleiotropy for QTLs related to flowering time and stolon production, and reveal different genetic architecture among the crosses, and between field versus greenhouse experiments. Finally, we quantify standing genetic variation within a single perennial population, and conduct five generations of artificial selection on high and low stolon number in the greenhouse. We use these selection lines to examine the response to selection and identify traits that evolve or are constrained through correlated evolution. Overall, we find strong multivariate trait associations, pleiotropic QTL, and patterns of covariation among traits that may determine the trajectory of adaptive divergence.

113W Genomics Facilitates Evaluation and Monitoring of McCloud River Redband Trout (*Oncorhynchus mykiss stonei*) *Ensieh Habibi*[1], Michael Miller[1], Daphne Gill[2], Leigh Sanders[2], Jeff Rodzen[2], Molly Stephens[3], Amanda Finger [1] 1) University of California, Davis, CA; 2) California Department of Fish and Wildlife, Genetics Research Laboratory, Sacramento, CA ; 3) Natural Reserve System, University of California Merced, Merced, CA

The McCloud River Redband Trout (MRRT; *Oncorhynchus mykiss stonei*) is a unique subspecies of rainbow trout that inhabits the isolated Upper McCloud River of Northern California. A major threat to MRRT is introgressive hybridization with non-native rainbow trout from historical stocking and contemporary unauthorized introductions. To help address this concern, we collected RAD-sequencing data on 308 total individuals from MRRT and other California *O. mykiss* populations and examined population structure using Principal Component and admixture analyses. Our results are consistent with previous studies; we found that populations of MRRT in Sheepheaven, Swamp, Edson, and Moosehead creeks are nonintrogressed. Additionally, we saw no evidence of introgression in Dry Creek, and suggest further investigation to determine if it can be considered a core MRRT conservation population. Sheepheaven Creek

was previously thought to be the sole historical lineage of MRRT, but our analysis identified three: Sheepheaven, Edson, and Dry creeks, all of which should be preserved. Finally, we discovered diagnostic and polymorphic SNP markers for monitoring introgression and genetic diversity in MRRT. Collectively, our results provide a valuable resource for the conservation and management of MRRT.

114T Tracking adaptation to seasonal insecticide pressure in Drosophila *Marianthi Karageorgi*[1], Mark C. Bitter[1], Caitlynn To-Duyen Tran[1], Caileb Travier[1], Hayes Oken[2], Skyler Berardi[2], Paul Schmidt [2], Dmitri A. Petrov[1] 1) Stanford University; 2) University of Pennsylvania

Understanding how organisms adapt to chemical pressures (e.g., plant allelochemicals and insecticides) is a central theme in evolutionary ecology. Adaptation to chemical pressures over ecological timescales provides an excellent study system to address the question because adaptation can be directly observed. In this project, we track the pace and the underlying phenotypic and genomic architecture of adaptation to seasonal insecticide exposure in replicate populations of *Drosophila melanogaster* in a field mesocosm experiment. The experiment lasted from early summer to late fall, and we performed high resolution temporal sampling to track evolution of insecticide resistance, other fitness-associated phenotypes, and allele frequencies genome-wide. We found that high insecticide resistance rapidly evolved in parallel in replicate populations of *D. melanogaster* following insecticide exposure in summer; when pesticide application ceased, resistance steadily declined until winter. The observed decline in resistance is consistent with known fitness costs of resistance in the absence of insecticide exposure. We also determined whether the evolution of insecticide resistance interacted with the evolution of other seasonal fitness-associated phenotypes, allowing us characterize trade-offs between adaptation to insecticide exposure and adaptation to seasonality. Last, the genomic data allowed test for the contribution of known insecticide resistance genes to the observed patterns in an unbiased, genome-wide manner. Overall, we expect that our study will provide insights into the phenotypic and genotypic architecture underlying rapid adaptation to chemical pressures.

115T Runs of homozygosity reveal extensive inbreeding among K'gari Island dingoes *Ana V. Leon-Apodaca*[1], Manoharan Kumar[2], Gabriel Conroy[2], Steven Ogbourne[2], Kylie M. Cairns[3], Sankar Subramanian[2], Zachary A. Szpiech[1] 1) Pennsylvania State University; 2) University of the Sunshine Coast; 3) University of New South Wales

Dingoes (*Canis dingo*) are wild canids from an ancient canid lineage, now naturalized in Australia. Their evolutionary history remains contested, and it is thought that they arrived in Australia via South-East Asia at least 5,000 years BP, via one or more founder events. As Australia's largest native terrestrial predator, they play an important ecological role. They are found across many different bioregions of Australia, including on multiple offshore islands. A protected population exists on K'gari (Fraser Island) that is relatively free from the risk of hybridisation with domestic or wild dogs. K'gari is the world's largest sand island and a World Heritage listed national park. Over 600,000 people visit K'gari each year resulting in occasional negative human-dingo interactions. While many management strategies are in place to minimize this occurrence, lethal control has been utilized significantly in the past, and still occasionally occurs. Previous research on K'gari dingoes using microsatellites, genome-wide SNPs, and mtDNA sequencing demonstrated divergence from mainland dingoes and low genetic diversity. However, whole-genome data is lacking from this important population. In this study, we analyze 18 whole genome sequences of dingoes sampled from mainland Australia (n=12) and K'gari Island (n=6) to assess the influence of their demographic histories on patterns of genetic diversity. Preliminary results showed that mainland dingoes and K'gari Island dingoes have distinct patterns of genetic diversity. We identify runs of homozygosity (ROH), indicators of small population size and inbreeding, in each population finding elevated levels of long ROH (>1 Mb) in both. However, K'gari dingoes showed significantly higher levels of very long ROH (>5 Mb; mean nROH = 71.2 for K'gari and 27.0 for the mainland, p = 5.35E-6; and mean sROH = 607.2 Mb for K'gari and 279.2 Mb for the mainland, p = 4.39E-4), providing clear evidence for inbreeding, isolation, small population size, and a strong founder effect. In the case of the K'gari dingo population, it appears that bottlenecks and isolation have maintained low levels of genetic diversity, while mainland dingoes show slightly higher diversity. We hypothesize that these ROH patterns may affect the distribution of deleterious homozygotes between mainland and K'gari dingoes. This work helps to elucidate the genetic structure and evolution of K'gari dingoes to inform conservation efforts.

116W Effects of inbreeding, drift, and selection on mutation load in the Florida scrub-jay *Mitchell G Lokey*[1], Tram N Nguyen[2,3], Elissa J Cosgrove[1], Felix EG Beaudry[4], Nancy Chen[4], Reed Bowman[5], John W Fitzpatrick[2,3], Philipp W Messer[1,2,6], Andrew G Clark[1,2,6] 1) Department of Molecular Biology & Genetics, Cornell University, Ithaca, New York, USA; 2) Department of Ecology & Evolutionary Biology, Cornell University, Ithaca, New York, USA; 3) Lab of Ornithology, Cornell University, Ithaca, New York, USA; 4) Department of Biology, University of Rochester, Rochester, New York, USA; 5) Archbold Biological Station, Venus, FL 33960, USA; 6) Department of Computational Biology, Cornell University, Ithaca,

New York, USA

Recent human activity has driven population decline for thousands of species around the world. Population genetic theory predicts that as populations decline, genetic drift can result in deleterious variants attaining higher allele frequencies. The aggregate increase of deleterious variants in a population leads to a larger genetic load in small and declining populations relative to large equilibrium populations. Furthermore, if inbreeding is increased due to population decline, the homozygosity of deleterious variants will also increase, thereby raising the homozygous load. Recent studies have claimed, however, that these genetic consequences of a population crash can be impacted by past demographic processes. In particular, ancestral bottlenecks may result in the loss of substantial amounts of recessive deleterious variation, as a consequence of both genetic drift and the increased efficacy of selection on homozygous recessive deleterious variants due to inbreeding, i.e. purging. The intertwined effects of demography, inbreeding, drift, and selection on mutation load has thus become an area of great interest for conservation genetics. In particular, the role of purging relative to random genetic drift in the loss of recessive deleterious variants during a bottleneck has been little studied. Here we explore the interplay of these processes using simulations, and relate these results to whole-genome sequences of 295 individuals from five meta-populations of the charismatic Florida scrub-jay (*Aphelocoma coerulescens*; FSJ). The FSJ is a well-studied bird with decades of demographic and environmental data collected from multiple conservation sites. Importantly, the FSJ has undergone several recent population bottlenecks, and, due to its non-migratory nature and human development, is currently living in fragmented meta-populations across the Florida peninsula. Using both gene prediction and synteny-based approaches, we identify putatively deleterious variants in whole-genome sequencing data for both historic and contemporary birds from five meta-populations. Comparing load across meta-populations and time in the population sequencing data and in forward genetic simulations, we explore the link between genetic load and the population genetic processes of inbreeding, genetic drift, and purifying selection.

117W The impact of climate change on parasite infection of bumblebees depends on mtDNA haplotypes of the host *Oliver Manlik*[1, 2], Sunil Mundra[1], Regula Schmid-Hempel[3], Paul Schmid-Hempel[3] 1) United Arab Emirates University; 2) University of New South Wales; 3) ETH Zurich

Climate change is predicted to affect host-parasite interactions. For instance, parasite infection prevalence is expected to increase with rising temperatures for some species. Global population declines of important pollinators have been attributed to fungal parasite (*Nosema*) infections, which, in turn, are influenced by environmental and climatic factors. However, the role of climate in driving parasite infection and the genetic basis for hosts to respond under variable climatic conditions remain obscure. In this study we investigated the association between climate and *N. bombi* infection of buffed-tailed bumblebees (*Bombus terrestris*), and whether this association is dependent on the host genotypes. For this we genotyped 876 wild bumblebee queens and screened *N. bombi* infection of those queens between 2000 and 2010. We also recorded seven climate parameters during those eleven years, and tested for correlations between climate and infection prevalence of specific host genotypes. Here we show that climatic factors drive *N. bombi* infection and that the impact of climate is dependent on mitochondrial DNA cytochrome oxidase I (COI) haplotypes of the host. Infection prevalence was correlated with climatic variables during the time period when queens emerge from hibernation. Remarkably, two host mtDNA COI haplotypes ('A' and 'B') best predict this association between climatic factors and infection. Both haplotypes displayed phenotypic plasticity, but in opposite direction: Haplotype A conferred greater resistance to parasite infection during hotter, wetter years, while haplotype B was more susceptible under those conditions. Our multivariate analysis further showed that the impact of interacting climatic variables on parasite infection is best explained by the two host haplotypes. To the best of our knowledge, this is the first study that identifies specific host genotypes that confer differential parasite resistance under variable climatic conditions. Our results demonstrate the importance of mitochondrial COI haplotypes to ward off parasites in a changing climate. More broadly, this also suggests that COI may play a pertinent role in climate change adaptations of insect pollinators.

118T Genomic and population viability analyses predict extinction risk in the most endangered marine mammal, the vaquita (*Phocoena sinus*) *Jacqueline Robinson*[1], Christopher Kyriazis[2], Sergio Nigenda-Morales[3], Annabel Beichman[4], Lorenzo Rojas-Bracho[5], Kelly Robertson[6], Michael Fontaine[7,8,9], Robert Wayne[2], Kirk Lohmueller[2,10], Barbara Taylor[6], Phillip Morin[6] 1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA; 2) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA; 3) Advanced Genomics Unit, National Laboratory of Genomics for Biodiversity (Langebio), Center for Research and Advanced Studies (Cinvestav), Irapuato, Guanajuato, Mexico; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) PNUD-Sinergia en la Comisión Nacional de Áreas Naturales Protegidas, Ensenada, BC, Mexico; 6) Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA, USA; 7) MIVEGEC, Université de Montpellier, CNRS, IRD, Montpellier, France; 8) Centre de Recherche en Écologie et Évolution de la Santé (CREES), Montpellier, France; 9) Gronin-

gen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, The Netherlands; 10) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

Anthropogenic pressures are driving species declines across the globe. Given limited conservation resources, a key question in cases of severe decline is whether recovery is possible, or if it is likely to be impeded by genetic factors such as inbreeding depression. Predicting the threat of inbreeding depression remains challenging, particularly for non-model species for which genetic and pedigree information is often limited or unavailable. A potential solution is to use genomic data to infer the genetic and demographic parameters needed to predict extinction risk through population viability analysis. We applied this framework to model extinction risk in the critically endangered vaquita porpoise (*Phocoena sinus*), which has declined to ~10 remaining individuals following decades of excess mortality from gillnet fishing in the Gulf of California, Mexico. We analyzed whole genome sequences from 20 vaquitas to investigate the impacts of the recent population decline, infer the species' demographic history, and examine deleterious variation in comparison with other cetaceans. We then integrated this genomic and demographic information into stochastic, individual-based simulations to quantify the vaquita's recovery potential. We find that the recent catastrophic decline has not yet impacted vaquita genomes, but that vaquitas nonetheless contain extremely low diversity and an excess of homozygous deleterious mutations due to long-term small population size. However, as a result of this naturally low diversity, vaquita genomes also contain very few segregating deleterious mutations, implying fitness may not decline under future inbreeding that is inevitable in any recovery scenario. We confirm this prediction through simulations of population dynamics over the next 50 years, which suggest that inbreeding depression in vaquitas is likely to be weak, and recovery is possible if bycatch mortality from gillnets is immediately halted. However, even modest rates of continuing bycatch result in an appreciable extinction risk, underscoring the importance of enforcing gillnet limits to avert the vaquita's extinction. Our results provide hope for vaquita recovery and highlight the applicability of genomic data in conservation.

119T Severe Inbreeding and Mutation Load in the Critically Endangered Devils Hole Pupfish *David Tian*[1], Austin Patton [1], Bruce Turner [2], Chris Martin [1] 1) UC Berkeley, Berkeley, CA; 2) Virginia Tech, Blacksburg, VA

Small populations with limited geographic distributions are predicted to be threatened by inbreeding and lack of genetic diversity, both of which may negatively impact fitness and exacerbate population decline. One of the most extreme natural examples is the Devils Hole pupfish (*Cyprinodon diabolis*), an iconic and critically endangered species with the smallest known habitat range of any vertebrate. This imperiled species has experienced severe declines in population size over the last thirty years and suffered major, repeated bottlenecks in 2007 and 2013, when the population sunk to 38 and 35 individuals, respectively. In our previous work, we found higher levels of inbreeding (F_{ROH} = 0.34 - 0.81) and increased deleterious genetic variation in the form of loss-of-function homozygous derived genotypes in the Devils Hole pupfish, relative to nearby populations of *Cyprinodon nevadensis* and *Cyprinodon salinus.* Moreover, we discovered predicted loss-of-function mutations and deletions associated with reproduction and hypoxia tolerance. This includes a fixed early stop codon in *cfap43* (*n* = 8/8 samples), which is associated with sperm flagellum defects and causes infertility in humans and mice and a deletion in the promoter of *redd1* (*n* = 7/7 samples), an inhibitor of mTOR signalling. Here, we present two *de novo* reference genomes of *C. diabolis* and *C. nevadensis mionectes* and an additional 150 resequenced genomes of Death Valley and Ash Meadows desert pupfish to improve our understanding of inbreeding and mutation load in desert pupfish and assess whether the captive bred population of *C. diabolis* has successfully maintained wild genetic variation. We thus document inbreeding and mutation load in the Devils Hole pupfish and in pupfish across Death Valley to inform management of this conservation icon and hopefully reduce extinction risk.

120W Evolution of immunity to cestode parasites is a pyrrhic victory *Jesse Weber*[1], Natalie Steinel[2], Foen Peng[3], Kum Chuan Shim[4], Brian Lohman[5], Lauren Fuess[6], Stephen De Lisle[7], Daniel Bolnick[8] 1) University of Wisconsin-Madison, Madison, WI; 2) University of Massachusetts Lowell, Lowell, MA; 3) Haverford College, Haverford, PA; 4) University of Texas at Austin, Austin, TX; 5) Eccles Institute of Human Genetics, Salt Lake City, UT; 6) Texas State University, San Marcos, TX; 7) Lund University, Lund, Sweeden; 8) University of Connecticut, Storrs, CT

Parasites impose fitness costs on their hosts. Biologists therefore tend to assume that natural selection favors infection-resistant hosts. Yet, when the immune response itself is costly, theory suggests selection may instead favor loss of resistance. Immune costs are rarely documented in nature, and there are few examples of adaptive loss of resistance. Here, we show that threespine stickleback fish (*Gasterosteus aculeatus*) have repeatedly gained immunity to a tapeworm parasite, during their replicated colonization of freshwater in the past 12,000 years. Yet, some freshwater fish populations have more complex and effective immunity than others. We identify the phenotypic and genetic basis of this recently-evolved variation in immunity. In particular, tapeworm infection stimulates inflammation and extensive fibrosis throughout the body cavity, which in turn contributes to suppression of parasite growth (in both lab and wild fish),

and can kill the tapeworm. However, this fibrosis response is costly, drastically reducing female and male reproductive success (in both lab and wild fish). Consistent with these costs, our quantitative genetic, population genomic, transcriptomic, and phylogenetic analyses all suggest that, in multiple freshwater populations that currently lack fibrosis and tolerate tapeworm growth, selection acted to favor the loss of this costly immune response. We also highlight several striking genetic changes that likely underlie these gains and losses of immunity. These results are unique in that we show the repeated gain and loss of immune adaptations across closely related conspecific populations. Moreover, our findings showcase the biomedical relevance of exploring the genetics of infection variation in wild vertebrates. Fibrosis is a major pathology in humans, and we show that stickleback contain naturally evolved genetic variation in pro- and anti-fibrotic pathways, providing a new model system to better understand fibrosis and inflammation in our own species.

121W The Evolutionary Consequences of Host-Microbe Interactions: *Rapid seasonal evolution of multiple host phenotypes mediated by associated microbes* *Jack Beltz*, Paul Schmidt University of Pennsylvania

Adaptive evolution is a complex process that is shaped by countless biotic and abiotic interactions between a population and its environment. The relative contributions of these forces and their ultimate impact on evolutionary trajectories are only beginning to be understood. Tracking the influence of host-associated microbes on their host population can offer valuable insight into the role interacting organisms can play in dictating the adaptive landscape of a population. To determine the influence of microbial interactions on host trait evolution, we constructed replicate populations from a panel of D. melanogaster lines collected from local Pennsylvania orchards, which evolved in parallel to seasonal change under different microbial treatments in a natural setting. The microbial treatments included Drosophila-derived Acetobacter and Lactobacillus taxa, which are known to colonize the host and influence phenotypes. The bacterial treatments were inoculated directly onto the population's fruit-based (nutritionally low) food supply, in an effort to maintain an environmentally relevant context. Various host phenotypes, as well as host and environmental microbiome composition, were tracked as these populations evolved in 18 replicate outdoor mesocosms for 115 days (June - November). During this period we observed a significant effect of the microbial treatments on the evolutionary trajectory of multiple life-history phenotypes, compared to control populations that did not receive microbial additions. By early fall the microbial treatment populations had evolved greater starvation resistance, egg to adult viability, adult body size, and lipid concentration as well as shorter larval development time. Additionally, we identified consistent variation across season and treatment, in the composition of the host microbiome. These results demonstrate that interactions between the host and both microbial isolates, effectively "rescue" the phenotypic effect of flies reared under the realistic nutrient stress of a fruit-based diet, potentially altering their fitness in a natural context. The variation in microbiome composition observed across seasonal time and microbial treatment suggests an influence of the host's internal and external environment on its microbial composition. Taken together, we demonstrate that shifts in the abundance of critical microbial taxa can have wide-reaching effects on host trait evolution, microbial composition, and adaptive landscape.

122T Finding patterns of antibiotic-resistant infections through the diversity of pathogenic sequence types. *Lorena Benitez-Rivera*[1], Anjani Pradhananga[1], Candace Clark[2], Pleuni Pennings[1] 1) San Francisco State University; 2) University of California San Diego

The increasing number of antibiotic resistant infections is a global threat to human health. Antibiotic resistance is the inability of antibiotics to treat the infections that were once treatable. Antibiotic resistance is a defense mechanism for the survival of pathogens, and occurs due to genetic modifications. In order to fight these resistant infections, it is important to know their origin. The problem lies in the fact that in many cases, the origin of these resistant infections is unknown. Antibiotic resistance might have evolved within patients, or may have been transmitted between patients. We expect that in situations where transmission is important, the diversity of the resistant strains is lower than the diversity of the susceptible strains. Diversity is therefore key in finding patterns to know the origin of resistance. In this study we determined diversity for resistant and susceptible strains from published datasets. We performed a retrospective study of six datasets with sequence type data; we analyzed the data using the programming language R, and calculated Simpsons's Index to measure diversity of susceptible and resistant strains. We used a bootstrapping approach to estimate significance. Our result suggests that diversity of the resistant strains is less than that of the susceptible population in some but not all of the cases. Thus, these studies of antibiotic resistance provide insight into how antibiotic resistance evolves and spreads, so that antibiotic resistance could be minimized.

123T Barcoding the Lenski Long-Term Evolution Experiment for Massively Parallel Bulk Fitness Assays Tanush Jagdish, Eliot Fenton, *Jack Edwards*, Michael Desai Organismic and Evolutionary Biology, Harvard University, Cambridge, MA

How evolution in a constant environment affects a population's adaptation to other novel conditions is an open ques-

tion. While studies have used massively parallel sequencing and lineage tracking experiments to address this problem, they are generally limited by short evolutionary timescales. We intend to study pleiotropy by taking advantage of the longest-running evolution experiment, the Long-Term Evolution Experiment with *Escherichia coli* (LTEE), which consists of 12 parallel *E. coli* populations started from a single ancestral clone that have been propagated for roughly 75,000 generations. A fundamental limitation of the LTEE has been that fitness assays are limited to colony-counting-based methods, which do now allow for high-throughput multiplexed assays. Given the sheer size of the LTEE library, which spans 150 timepoints for each of the12 populations, and many more coexisting subpopulations, a more high-throughput approach to measure fitness is essential. Here we propose to uniquely barcode multiple clones at all timepoints in every population in the LTEE, which would roughly amount to ~10,000 barcoded clones. We will then use the barcoded LTEE library to measure fitness in a series of environmental conditions using barcoded bulk fitness assays. We show results from our barcoding effort and preliminary data from bulk fitness assays in select environmental conditions from our pilot experiments. Understanding the dynamics of pleiotropy, especially over long evolutionary timescales relevant to natural populations, is essential to building a complete framework of evolutionary theory. We take a small step in this direction by taking advantage of an existing model system.

124W Effect of inoculation dose on colonization success in gut-derived microbial communities *Doran Goldman*[1], Katherine Xue[1], Rashi Jeeda[2], Benjamin Good[1], Dmitri Petrov[1], Kerwyn Huang[1], David Relman[1] 1) Stanford University; 2) California Institute of Technology

The human gut microbiome remains remarkably diverse and stable over the course of adult life, despite constant exposure to new microbes from the environment. Various ecological factors and evolutionary processes may affect the likelihood that new strains successfully colonize an established community—for instance, an introduced strain might directly compete with a resident strain in a resource-limited environment, or it might adapt to utilize alternative resources in a resource-rich environment. However, it remains unclear to what extent each of these strategies plays a role in microbiome stability. The neutral theory of community ecology provides a highly simplified model for colonization in which the relative abundance of a newly introduced species is proportional to its inoculation dose. To test the predictions of neutral theory, we derived *in vitro* microbial communities from human stool samples and performed community coalescence experiments in which we systematically mixed eight pairs of communities at ratios ranging from 1:1000 to 1000:1. Neutral theory predicts that the relative abundance of a taxon will increase linearly with its inoculation dose, and that species diversity will be higher when communities are mixed at more even ratios, as more species persist at relative abundances above the sequencing limit of detection. In preliminary analyses, we observed that species diversity in the community mixtures often remained at levels similar to the parent communities regardless of mixture ratio, suggesting that diversity may be limited by resource competition even when many species are initially present in the community mixtures. While the relative abundances of some taxa increased with inoculation dose, as predicted by neutral theory, several species were also able to persist stably at relative abundances several times higher than predicted by their inoculation dose, indicating that these taxa may colonize by adapting to consume underutilized resources or engaging in mutualistic interactions with other species. Our results highlight the importance of both inoculation dose and other ecological and evolutionary processes in determining colonization success. These analyses will provide important insight into the ecological and evolutionary factors that promote colonization in gut microbial communities.

125W Experimental evolutionary genomics of herbivorous insects on multiple host plant species *Diler Haji*[1], Andrew Gloss[2], Noah Whiteman[1] 1) University of California, Berkeley; 2) University of Chicago

Despite decades of progress, a major question in evolutionary biology is how so much functional genetic variation within species is segregating when genetic drift and both positive and negative selection should drive its loss. Balancing selection may explain how much of this variation is maintained. In the broadest sense, balancing selection can act through a variety of mechanisms including spatially- and temporally varying selection (STVS). In 1953, Howard Levene proposed a compelling model of STVS that has largely been untested. In this model, viability selection across niches in a randomly mating diploid population ensures the maintenance of alternative alleles without heterozygous advantage. This mode of balancing selection is particularly relevant to insect herbivores – plant parasites that feed on an assortment of different host plants containing varying chemical defenses (i.e., host “chemo-niches”). We tested the Levene model by experimentally evolving the herbivorous leaf-mining drosophilid fly Scaptomyza flava for 11 generations on monocultures or mixtures of two natural and chemically divergent mustard host plants, Barbarea vulgaris and Turritis glabra, and then compared our experimental findings with wild-caught populations from both host plants. We found that S. flava evolved to become strongly local adapted to each monoculture. Using a pooled genome sequencing approach, we found preliminary evidence that evolution on monocultures led to a stronger reduction in genetic diversity genome-wide than evolution on mixtures in some monoculture populations. We compare these results to wild samples of S. flava reared from

their two natural host plant species. These results suggest that STVS may contribute to the maintenance of functional genetic diversity directly as a result of biotic interactions at the plant-herbivore interface.

126T The role of sex in evolution: Sexual conflict and Sexual selection *Sheng-Kai Hsu*[1,2], Wei-Yun Lai[1,2], Johannes Novak[3], Felix Lehner[2], Ana Marija Jakšić[1,2,4], Elisabetta Versace[5], Christian Schlötterer[2] 1) Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria; 2) Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria; 3) Institute of Animal Nutrition and Functional Plant Compounds, Vetmeduni Vienna, Vienna, Austria; 4) École polytechnique fédérale de Lausanne, Lausanne, Switzerland; 5) Department of Biological and Experimental Psychology, Queen Mary University of London, London, UK

Since the evolution of sex, sexually reproducing organisms have managed to occupy a wider range of habitats and thrived in greater environmental fluctuation owing that recombination increases genetic variation for adaptation. In addition, the distinct roles of males and females in sexual reproductive system brings novel forms of evolutionary dynamics, namely sex-specific fitness requirements and the sexual selection imposed by one sex to the other. On one hand, despite decades of investigation on the resolution of sexual conflict, it is not yet clear whether and how the two sexes are able to respond differently to a sudden environmental shift. On the other hand, how sexual selection interplays with the ecological and mutation-order factors in affecting the evolution of reproductive isolation has long been a popular research question. Nevertheless, empirical studies, particularly on the mutation-order process, are limited. In this work, we decide to take the approach of experimental evolution with the aim to identify demonstrative insights into these open questions. Within 100 generations of experimental thermal adaptation, we demonstrate rapid sex-specific changes in transcriptional, metabolic and behavioral phenotypes. We propose that the standing genetic variation on the sex-specific genetic architecture due to the historical resolution of sexual conflict is the key factor that allows the uncoupled evolution in males and females of the same population. This model is further illustrated with our forward computer simulations. Further, utilizing the high replication level of the same experimental evolution framework, we empirically investigate both the deterministic and stochastic mechanisms behind the evolution of reproductive isolation during the experimental thermal adaptation. Particularly, we reveal the importance of the intersexual epistatic dependency due to sexual selection on the operation of mutation-order process. This not only resounds the theoretical prediction but also highlights the potential incompatibilities among standing genetic variation. This study advances the knowledge on the sex-related aspects of local adaptation and associated ecological/mutation-order speciation with unprecedented experimental evidences.

127T Killer Yeast: Uncovering the evolutionary history and environmental/genetic underliers to the antimicrobial activity of three core metabolic enzymes in *Saccharomyces cerevisiae* *Hannah Kania*[1,2], Mohammad Siddiq[1,2], Nick Brown[3], Patricia Wittkopp[1,2] 1) Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI; 2) Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI; 3) College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI

Some proteins contribute to multiple functions within a cell, and these functions can vary drastically. The *GAPDH* homologs in *Saccharomyces cerevisiae*, *TDH1*, *TDH2*, and *TDH3*, encoding core enzymes used in glycolysis, have been identified as precursors to antimicrobial peptides with which *S. cerevisiae* uses to kill off other yeast species in mixed culture, synthetic grape juice (SGJ) fermentations. This discovery identifies a novel, secondary function of these metabolic genes that is an intriguing example of co-option.

We aim to better understand the environmental and genetic causes that underlie the propensity of *S. cerevisiae* to kill other yeast in mixed cultures by characterizing how and when the *TDH* genes were recruited into fungal warfare. Our experiments seek to uncover environmental factors affecting the antimicrobial phenotype, identify the genetic contributions of each *TDH* gene to *S. cerevisiae* fungicidal capabilities, and characterize when the *TDH*-derived antimicrobial property evolved on the yeast phylogeny. To characterize how environment affects *S. cerevisiae* fungicidal behavior, we grew wild-type *S. cerevisiae* and *Hanseniaspora guilliermondii*, a victim species of *S. cerevisiae*, in single and mixed cultures in different types of liquid media and quantified their abundance. We find a striking effect of growth environment on antimicrobial activity: in SGJ *S. cerevisiae* completely kills off its victim species, whereas in YPD *S. cerevisiae* has virtually no effect and is even outcompeted by *H. guilliermondii*. To quantify how each *TDH* gene affects antimicrobial activity, we are performing high resolution flow-cytometry based competition assays with *H. guillermondii* and engineered strains of *S. cerevisiae* that lack the *TDH* genes individually or in combination. We are also competing other species of *Saccharomyces* yeast against *H. guillermondii* to determine whether production of the *TDH*-based antimicrobial peptides is a unique feature of *S. cerevisiae* or an ancient tool of fungal warfare shared by the *Saccharomyces* genus.

128W Three range limit hypotheses tested in climate-manipulated common gardens *Laura Leventhal*[1,2], Megan Ruf-

fley[1,2], Shannon Hateley[1,2], Moises Exposito-Alonso[1,2] 1) Carnegie Institution for Science, Department of Plant Biology, Stanford, CA; 2) Stanford University, Stanford, CA

A species' range is a physical representation of the limits of evolution as it epitomizes a species' evolutionary niche in a specific period of time. Range limits are dynamic barriers affected by the abiotic and biotic environment, population dynamics, and, least well understood, genetic mechanisms. There are currently three hypotheses that best account for the genetic mechanisms behind the formation of range limits. 1) An increase in deleterious mutation accumulation as a result of low density at the edges of ranges relative to the center, making edge populations more susceptible to genetic drift and the fixation of deleterious mutations. 2) Higher levels of genetic swamping where because edge populations are also frequently locally adapted, gene flow from the center to the edge of a population can create maladaptive hybrids at the edge. And 3) a high variance in fitness in extreme environments which present the opportunity for natural selection to act but also may allow fitness to aggressively decline to zero. We are testing each of these hypotheses using an outdoor common garden experiment with Arabidopsis thaliana on Stanford University's campus in Central California. To disentangle the genetic and ecological components of range limit formation, we used a sophisticated irrigation system that simulated 14 regimes of drought stress, varying in frequency and abundance, simultaneously. We selected ecotypes from the 1001 Genomes Project (1001genomes.org/) that fit the criteria of the three range limits by screening for Ka/Ks ratio, polygenic score for survival in drought conditions, distance to edge of A. thaliana's range, and fitness in previous drought experiments. In total, we planted 25,920 plants on the same day in November 2021 and will monitor them throughout their entire life cycle in Spring 2022. Fitness will be measured via survival and seeds produced, along with phenological variables including first day of flowering, day of death, and lifetime duration, as well as ecologically-relevant phenotypes such as stomata, trichome indices, growth rate, and ∂C13. This experiment will allow us to elucidate which hypothesis or combination of hypotheses contributes the most to the formation of range limits. It is essential that we understand the mechanism behind range limit formation as climate change is altering ecosystems at an alarming rate, and the capabilities of a population to respond may be the only way to avoid extinction.

129W Diverse mating phenotypes impact the spread of *wtf* meiotic drivers in *Schizosaccharomyces pombe* *Jose Fabricio López Hernández*[1], Rachel M. Helston[1], Jeffrey J. Lange[1], R. Blake Billmyre[1], Samantha H. Schaffner[2], Michael T. Eickbush[1], Scott McCroskey[1], Sarah E. Zanders[1,3] 1) Stowers Institute for Medical Research, Kansas City, MO 64110, USA.; 2) Kenyon College, Gambier, OH 43022, USA.; 3) Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS 66160, USA.

Meiotic drivers are genetic elements that break Mendel's law of segregation to be transmitted into more than half of the offspring produced by a heterozygote. The success of a driver relies on outcrossing because drivers gain their advantage in heterozygotes. It is therefore curious that *Schizosaccharomyces pombe*, a species reported to rarely outcross, harbors many meiotic drivers. To address this paradox, we measured mating phenotypes in *S. pombe* natural isolates.

We found that the propensity to inbreed varies between natural isolates and can be affected both by cell density and by the available sexual partners. Additionally, using experimental evolution approaches and theoretical modeling, we found that the observed level of inbreeding slows, but does not prevent, the spread of a *wtf* meiotic driver in the absence of additional fitness costs. These analyses reveal parameters critical to understanding the evolution of *S. pombe* and help explain the success of meiotic drivers in this species.

130T The efficacy of QTL analysis to predict adaptive variation: a test using experimentally evolved populations of yeast *Helen Murphy*, Benjamin Epley, Brianna Meeks, Juliana Salcedo William and Mary

A major goal of evolutionary biology is to identify adaptive genetic variation and understand its fate in a population. Quantitative trait locus (QTL) analysis is a common approach used to uncover variants associated with adaptive traits. It is unclear whether QTL hypothesized to underlie traits of interest are in fact those that will be favored by natural selection, as complex life histories, epistasis, and pleiotropy may affect the strength and direction of selection on these loci. We tested whether variants implicated in a QTL analysis increased in frequency when under selection using the yeast *Saccharomyces cerevisiae* as a model. The ability to adhere to plastic, which is the first step in biofilm formation and a clinically-relevant trait, was investigated in a highly-heterozygous strain isolated from a medical setting (YJM311). First, a bulk segregant analysis (BSA) was used to identify plastic adherence QTL; this assay implicated variants in a number of genes known to be associated with other biofilm-related traits, including *FLO8* and *FLO11*. Next, experimental evolution was performed on replicate sexual and asexual populations for 500 generations while selecting for the ability to adhere to plastic. In order to identify loci under selection, populations were subject to whole-population, whole-genome sequencing at 8 time points throughout the experiment. Thousands of variants were found to be under selection, many

likely related to adapting to the experimental conditions and not just increasing the ability to adhere to plastic. When the two datasets were compared, nearly all the loci with the strongest signal in the BSA increased in frequency throughout the selection experiment, suggesting the QTL analysis successfully uncovered evolutionarily-relevant loci. However, most of these variants did not fix in the experimental populations. These results suggest that the loci uncovered in the QTL analysis were subject to a more complex process than a simple selective sweep of adaptive loci and that populations with abundant segregating variation may not conform to simple evolutionary models.

131T Coordinating nutrition and energy allocation in *Drosophila melanogaster*: Genetic mechanisms and evolution *Enoch Ng'oma*, Joseph C. Gunn, Elizabeth Jones, Elizabeth G. King University of Missouri, Columbia, MO

Organisms have a basic necessity to acquire nutritional resources from the environment with which to build structures and maintain biological function. The quantity and quality of acquired nutrients, the 'criteria' for allocation to reproduction, maintenance, and storage to optimize the fitness-longevity trade-off is dictated by the quality of the environment and the organism's condition. However, we do not fully understand the genetic mechanisms which coordinate allocation patterns in natural populations, and what the sources of genetic variation in allocation 'decisions' between populations in variable nutritional conditions are. We investigate these questions in an outbred population generated from intercrossing over 800 recombinant inbred lines from a multiparent population, itself derived from a global set of eight inbred founder lines. We imposed three diet selection regimes on this base population for 30 generations. In each generation, 12 large replicates (N > 2000 flies each) were treated with a fluctuating, a deteriorating, and a constant high nutrient regime. We sampled at generation 0, 5, 10, 20 and 30 and report single nucleotide polymorphism (SNP) frequencies associated with adaptation to selection patterns. We relate these results to changes in life history traits during adaptation including lifespan. We identify the loci associated with adaptation to dietary selection and shed light on the underlying genetic mechanisms. Our results have important implications to nutritional, gerontological, and climate change research.

132W Upper bound on the mutational burden imposed by a CRISPR-Cas9 gene drive element *Michael Overton*, Sean Guy, Sergey Kryazhimskiy University of California San Diego

CRISPR-Cas9 gene drives (CCGDs) are poised to expand our ability to control wild populations, such as the eradication of disease vectors. However, Cas9-gRNA complexes can produce off-target dsDNA breaks. This raises the possibility that a CCGD element introduced into a genome that has not previously evolved in its presence could increase rates of mutations, genome instability, and loss of heterozygosity (LOH), with unpredictable long-term evolutionary consequences for the engineered species and the surrounding ecosystem.
To assess the potential evolutionary effects of a CCGD in a naive genome, we performed a large-scale mutation accumulation (MA) experiment in the yeast *Saccharomyces cerevisiae*. We introduced a CCGD construct into the genome of a diploid hybrid yeast strain, heterozygous at ~50,000 loci to detect LOH events at high resolution (Drive strain). We also created two control strains in the same genetic background, one carrying a neutral marker (WT control) and the other carrying the Cas9 gene alone without a gRNA (Cas9 control). We propagated at least 65 MA lines of each strain for 800 generations, and sequenced whole genomes of the founders and end-point clones.
Using these data, we estimated the rates of spontaneous mutations in the Drive, Cas9 control, and WT control strains to be 4.09×10^{-11}, 5.05×10^{-11}, and 4.68×10^{-11} per base-pair per generation, respectively, and the rates of LOH events to be 7.65×10^{-3}, 8.27×10^{-3}, and 7.93×10^{-3} per genome per generation, respectively, all of which were consistent with previous estimates. Neither de novo mutation nor LOH event rates were statistically distinguishable between the strains. Given the statistical power of our study, we estimate with 70% confidence that the presence of a CCGD element increases the genome-wide rates of LOH events and de novo mutations by less than 25% and 70%, respectively. Despite the lack of significant differences for the genome-wide rates, we found a significant difference in the distribution of LOH events across Chromosome V between the WT and Drive strains.
While it is important to examine the evolutionary consequences of CCGD elements in other species, our results suggest that these elements impose at most a weak and likely localized additional mutational burden to a naive genome, and that CCGD-based population control efforts are likely to be evolutionarily safe.

133W Experimental evolution reveals the synergistic genomic mechanisms of adaptation to ocean warming and acidification in a marine copepod Reid Brennan[1,2], James deMayo[3], Hans Dam[3], Hannes Baumann[3], Vince Buffalo[4], *Melissa Pespeni*[1] 1) Department of Biology, University of Vermont, Burlington, VT, USA; 2) GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany; 3) University of Connecticut, Groton, CT, USA; 4) Institute for Ecology and Evolution, University of Oregon, USA

Metazoan adaptation to global change will rely on selection of standing genetic variation. Determining the extent to which this variation exists in natural populations, particularly for responses to simultaneous stressors, is essential

to make accurate predictions for persistence in future conditions. Here, we identified the genetic variation enabling the copepod *Acartia tonsa* to adapt to experimental ocean warming, acidification, and combined ocean warming and acidification (OWA) over 25 generations. Replicate populations showed a consistent polygenic response to each condition, targeting an array of adaptive mechanisms including cellular homeostasis, development, and stress response. We used a genome-wide covariance approach to partition the allelic changes into three categories: selection, drift and replicate-specific selection, and lab adaptation responses. The majority of allele frequency change in warming (57%) and OWA (63%) was driven by shared selection pressures across replicates, but this effect was weaker under acidification alone (20%). OWA and warming shared 37% of their response to selection but OWA and acidification shared just 1%, indicating that warming is the dominant driver of selection in OWA. Despite the dominance of warming, interaction with acidification was still critical as the OWA selection response was highly synergistic with 47% of the allelic selection response unique from either individual treatment. These results are among the first to disentangle how genomic targets of selection differ between single and multiple stressors and to demonstrate the complexity that non-additive multiple stressors will contribute to predictions of adaptation to complex environmental shifts caused by global change.

134T A mitonuclear reality check on the evolutionary significance of Mother's Curse in Drosophila. *David Rand*, Faye Lemieux, Kenneth Bradley, Lindsay Marmor Brown University

Maternal inheritance allows selection to act on mtDNA-encoded effects in females, but prevents direct selection on mtDNA in males. Mutations that are deleterious in males but neutral or beneficial in females can persist in populations. This predicts that mtDNA-based disease or phenotypic variation should be more common in males, while haploid selection in females will purge mtDNA-based variation (Frank and Hurst (1996); repackaged as the 'Mother's Curse' by Gemmell et al. (2004)). There is conflicting evidence for this pattern in the literature. A key assumption in Mother's Curse is that mtDNA phenotypes must be sex limited with different effects, even different signs, in males and females. Extreme Mother's Curse scenarios invoke mtDNA mutations that are beneficial in females and deleterious in males and sweep through populations leading to extinction from male unfitness. Comparisons of sex-specific mtDNA phenotypic effects from different populations and species are needed to evaluate the evolutionary significance of Mother's Curse.
Most Mother's Curse analyses use alternative mtDNAs placed on one or more homozygous nuclear chromosomal backgrounds. Since most organisms are heterozygous at many loci, we sought to perform experiments in several different heterozygous backgrounds. MtDNAs from Drosophila melanogaster (OreR and Zimbabwe), D. simulans (siI and siII) and D. yakuba were each placed on a common D. melanogaster w1118 nuclear background. Virgin females from these strains were crossed to males from each of several deficiency stocks carrying a hemizygous segment of chromosome 2L. F1 female and male flies carrying the deficiency chromosome and the w1118 chromosomes were tested for starvation, climbing and flight performance. For all three traits in the majority of chromosomal backgrounds, the variance among mtDNA genotypes was greater in females than in males. This result is the opposite of the Mother's Curse prediction. Moreover, the impact of the foreign D. yakuba mtDNA was equally neutral or beneficial in both males and females, suggesting some form of phylogenetic heterosis. The mitonuclear epistatic interactions across the different heterozygous backgrounds and the five mtDNA haplotypes are more pronounced in females than males. This suggests that mtDNA interactions with regional hemizygosity or dominance effects are more pronounced in females than males, overshadowing any effect of Mother's Curse or even hidden Y chromosome variation.

135T Interplay of structural and regulatory evolution in functional evolution of glycolytic enzymes *Mohammad Siddiq*[1,2], Hannah Kania[1,2], Nick Brown[3], Patricia Wittkopp[1,2] 1) Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI; 2) Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI; 3) College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI

Genes often have multiple functions, and some of these functions change over time while others remain conserved. How do different factors constrain or enable the evolution of a gene's functions at the molecular level? One way of addressing this question is by concurrently studying the structural and regulatory evolution of a gene because what a gene product (e.g, protein) does and can do is determined by when, how, and in what form that product exists. Here, we investigate how gene structure and regulation have coevolved and shaped functional conservation and diversification of the three GAPDH genes in *Saccharomyces* yeast. The *GAPDH*s—named *TDH1*, *TDH2*, and *TDH3* in *S. cerevisiae*--catalyze the sixth step of glycolysis, are among the most highly expressed proteins in yeast, and last shared a common ancestor >100 million years ago despite their high sequence similarity to each other (~90% sequence identity). Yet, the presence of three GAPDH enzymes is not necessary for glycolysis; numerous fungi outside of the *Saccharomyces* clade have only GAPDH. Here, we explore the extent to which the genes have functionally diverged from each other. We first use controlled perturbations of the genes individually and combinatorically to assess their effects on enzymatic activity and growth rate and find that paralogous pairs vary in their level of functional redundancy. We find that changes in

gene expression mediated through *cis*- regulatory sequences are sufficient to explain most of the functional divergence between paralogs with respect to their effects on growth rate. However, we also find that at least one of the paralogs is post-transcriptionally regulated such that its localization extends beyond the cytoplasm—the canonical domain of glycolytic enzymes—to the nucleus and also outside of the cell. Our ongoing work aims to resolve whether the paralogs are differentially distributed in the cells and characterize how protein coding changes affect functions in the different locales, where the roles of these canonical housekeeping genes remain largely unknown. Collectively, the results from this work will both highlight how coupling experimental and evolutionary frameworks can reveal new molecular functions of a gene and provide mechanistic insight into the ways those diverse functions have coevolved through changes in gene structure and regulation.

136W Colonization and evolution after antibiotic perturbation in the human gut microbiome *Katherine Xue*, Doran Goldman, Kerwyn Huang, Dmitri Petrov, Benjamin Good, David Relman Stanford University

Gut commensal bacteria frequently form long-term associations with human hosts but can also transmit between hosts to colonize new microbial communities. The ecological context and evolutionary pressures associated with strain colonization and transmission remain largely unknown, despite their importance for designing successful targeted microbiome therapeutics. Here, we investigate the dynamics of strain transmission and evolution in a longitudinal household cohort of 48 healthy adults from 22 households over two months, during which one subject in each household took a 5-day course of the antibiotic, ciprofloxacin. Prior work has shown that cohabiting individuals frequently carry closely related microbial strains, suggesting that households are sites of natural microbial transmission. We developed methods to identify strain transmission events in our study using metagenomic sequencing data, and we tracked the ecological context and evolutionary dynamics of these transmission events in our household cohort. Individuals had heterogeneous ecological responses to antibiotic perturbation: in some subjects, the gut microbiome rapidly returned to its initial composition after antibiotic exposure, but in others, the microbiome transitioned after antibiotic exposure to an alternative stable state that persisted through months of follow-up sampling. In one intriguing case of a subject who experienced a major antibiotic perturbation, a strain of *Bacteroides stercoris* transmitted from a cohabiting partner after antibiotics and maintained a relative abundance of more than 50 percent through a year of subsequent follow-up sampling, suggesting that, in some cases, strain transmission can play a major role in reshaping the post-antibiotic community. We are now working to identify mutations that arise in strains after transmission as they adapt to a new host context. Our work shows that colonization following antibiotic perturbations can substantially remodel gut microbial communities and helps shed light on the evolutionary pressures that gut microbes experience as they circulate and transmit between hosts.

137W Detection of structural variants among inland annual and coastal perennial ecotypes of the yellow monkey flower, *Mimulus guttatus* *Leslie Kollar*, David Lowry, Chad Niederhuth Michigan State University, East Lansing, MI

The natural world consists of a mosaic of environments that drives adaptive divergence among natural populations. A critical gap in our knowledge of understanding how heterogenous environments drive evolutionary adaptation is establishing the role of structural variants (SV) in those adaptations. Our ability to identify SVs, especially chromosomal inversions, is limited by genome assemblies (i.e. our power to resolve repeat regions). Here, we use long-read Oxford Nanopore sequencing to identify large structural variants that differ between inland annual and coastal perennial ecotypes of the yellow monkeyflower, *Mimulus guttatus*. The inland populations of *M. guttatus* are characterized by an adaptive early flowering annual life history driven by summer drought, whereas the coastal populations of *M. guttatus* have a late-flowering perennial life history, as these populations are protected from drought stress by a pervasive marine fog. Additional abiotic factors such as oceanic salt spray in coastal habitats may play a key role in adaptation within this species. Research has identified QTLs linked to adaptive divergence between the ecotypes, some of which are localized to a large chromosomal inversion. Using these genomes, we identified SVs and began to resolve a decade long obstacle of identifying the breakpoints of a large chromosomal inversion thought to be involved in the adaptive divergence of *M. guttatus.* Additionally, we have clarified a list of genes located within the chromosomal inversion with some previously identified as responsible for adaptive phenotypes. Collectively, these genomes will lay the foundation for future research in understanding the role of structural variants, specifically chromosomal inversions, in adaptation.

138T Evolution of gene expression patterns of paralogous hormones (IGF1 and IGF2) and paralogous receptors (IGF1R and INSR) across amniotes. *Tonia Schwartz*, Abby Beatty, Morgan Muell, Kelly Blackshear Auburn Univeristy

The insulin and insulin-like signaling (IIS) network plays an important role in mediating several life-history traits, including growth, reproduction, and senescence. In vertebrates there are two paralogous receptors (IGF1R and INSR) and three paralogous hormones (IGF1, IGF2, and Insulin) that can bind the receptors with varying affinities and these relationships are likely co-evolving across the vertebrate phylogeny. Along with evaluating the sequence variation, evolution of expression

patterns are also important for understanding these co-evolutionary relationships. Although IGF1 and IGF2 are both key hormones in the vertebrate IIS network, research on IGF2 in juveniles and adults has been largely neglected because early biomedical research on rodents found negligible IGF2 postnatal expression. Our previous research on gene expression in this system found that lizards express IGF2 throughout life, and IGF1R was expressed at a much lower level than INSR – in stark contrast to biomedical rodents. Here, we challenged these assumptions of IGF2 and IGF1R expression being consistent across the phylogeny. Primarily, we ask to what degree IGF2 is expressed during postnatal life across amniotes (reptiles and mammals) by quantifying the relative gene expression of *IGF1* and *IGF2* using publicly available RNAseq data for 82 amniote species and quantitative polymerase chain reaction on liver cDNA at embryonic, juvenile and adult stages for two lizard, bird and mouse species. The results we will present indicate that (i) *IGF2* is expressed postnatally across amniote species and life stages—often at higher relative expression level than *IGF1*, contradicting rodent models; (ii) the lack of rodent postnatal *IGF2* expression is due to phylogenetic placement, not inbreeding or artificial selection; and (iii) adult *IGF2* expression is sex-biased in some species. These results demonstrate that *IGF2* expression is typical for amniotes throughout life, suggesting that a comprehensive understanding of the mechanisms mediating variation in life-history traits will require studies that measure both IGFs, since each can bind both receptors (IGF1R and INSR). We are furthering this research in understanding the relative expression patterns of the receptors (IGF1R and INSR) across the amniote phylogeny that will be incorporated into this discussion of the evolution of these hormone receptor relationships across amniotes.

139T Single-cell RNA sequencing identifies a unique set of cells that give rise to a diverse bone found in the mammalian penis *Caleb Ghione*, Matthew Dean University of Southern California

The baculum, a bone found in the penis of many mammals, has been gained and lost multiple times and is morphologically diverse across species. Do independent derivations occur via switching on of conserved bone developmental pathways, or through novel pathways? To begin answering this question, we employed single-cell RNA sequencing to identify the precursors of the baculum and characterize their patterns of gene expression. By integrating our data with existing literature, we show that these precursors are very similar to forelimb precursors. We conclude that independently derived bacula are probably deploying conserved bone developmental programs. Our work sheds light on the evolution of morphological novelty.

140W Genetics of behavioral evolution in giant mice from a predator-free island. *Jered Stratton*, Mark Nolte, Bret Payseur University of Wisconsin - Madison

Organisms on islands often evolve extreme phenotypes. Novel environmental conditions such as a lack of predators can shift long-standing adaptive peaks to new optima. House mice from Gough Island are a prime example of rapid and extreme phenotypic evolution following island colonization. These mice have nearly doubled in body size and frequently predate on nesting seabirds. We hypothesized that an absence of both natural predators and man-made shelters stimulated the evolution of increased boldness and exploration in mice from Gough Island. To test this hypothesis, we conducted a series of behavioral tests in a controlled laboratory setting using wild-derived inbred strains of mice from Gough Island and the Eastern United States. Open field and light/dark box tests show Gough Island mice are more active and spend more time in open, brightly lit areas than mainland mice. Exploration-related behaviors are influenced by age, sex, and parental identity whereas boldness-related phenotypes are not, suggesting that these two behavioral classes are genetically separable. To identify loci associated with the evolution of exploration and boldness, we used open field and light/dark box tests to quantify the behavior of 638 F2 mice generated by intercrossing Gough Island mice with mice from the mainland strain. Measures of exploration and boldness are uncorrelated in F2s further suggesting they are genetically separable while body size is slightly negatively correlated with exploration. F2 mice were genotyped at 31,683 single-nucleotide polymorphisms informative in the cross. We report quantitative trait loci (QTL) for exploration-related and boldness-related behaviors, providing one of the first portraits of the genetic architecture of the island syndrome. We also describe genetic effects of mice on the behaviors of their cage-mates (indirect genetic effects).

141W Rapid evolution of microbial adherence by host protein domain shuffling. *EmilyClare Baker*[1], Ryan Sayegh[1,2], Kristin Kohler[1], Wyatt Borman[1], Matthew Barber[1] 1) Institute of Ecology and Evolution, University of Oregon, Eugene, OR; 2) Department of Molecular, Cellular & Developmental Biology, University of Colorado Boulder, Boulder, CO

Epithelial cells are often the first point of contact between microbes and their animal hosts. As such, epithelial interactions with pathogenic microbes play a key role in determining successful microbial colonization or clearance. Pathogens may therefore be an important source of selective pressure for proteins on epithelial surfaces. While the pathogen binding surfaces of dedicated immune proteins may have significant evolutionary flexibility, pressure to maintain critical cellular functions could constrain the evolution of other epithelial proteins in response to pathogen antagonism. Despite

the importance of epithelial surface proteins in establishing host-microbe associations, how these interactions evolve over time remains largely undefined.

Bacterial adhesins are a diverse class of surface proteins that mediate binding and attachment to host epithelia. A common target of bacterial adhesins are CEACAM proteins, a multifunctional vertebrate family of cell adhesion molecules that play important roles in development and tissue homeostasis. CEACAM proteins are widely expressed across epithelia, with some also expressed on neutrophils. Here we show that a subset of primate CEACAM proteins are evolving rapidly with high levels of divergence in the extracellular N domain. The CEACAM N domain is critical for CEACAM protein function, but it is also the binding site for bacterial adhesins. We observe that diversification of the N domain has been accelerated by recurrent gene conversion among bacterially antagonized CEACAM orthologs. Furthermore, gene conversion events in this domain appear to make up the majority of human polymorphisms for the bacterially targeted CEACAM proteins, CEACAM1, CEACAM3, and CEACAM5 (CEA). Using biochemical protein binding experiments, we show that both between and within species diversity determines recognition by a panel of bacterial adhesins from the human pathogens *Helicobacter pylori* and *Neisseria gonorrhoeae*. Collectively our study suggests that gene conversion is an important mechanism by which multifunctional proteins can evolve in response to pathogen antagonism.

142T The Genetics and Physiology of Switchgrass Local Adaptation Across North America David Lowry Michigan State University

Local adaptation is a fundamental driver of biodiversity on planet Earth. While recent experiments have begun to dissect the genetic basis of local adaptation, we still have a poor understanding of how individual genetic loci contribute to local adaptation over large-scale environmental gradients. To understand local adaptation at a continental scale, we conducted a long-term 13 field site study, spanning 24 degrees of latitude from central Mexico to the northern United States, in the major bioenergy/bioproducts crop switchgrass (*Panicum virgatum*). Much of the functional genetic variation in switchgrass is distributed clinally with latitude as well as among upland, lowland, and coastal ecotypes. Southern lowland populations are generally high yielding, tolerant to heat, drought, and pathogens, while northern upland populations are superior in their acclimation to cold and tolerance to freezing tolerant. To understand the genetic basis of local adaptation across central North America, we implemented multi-site quantitative trait locus (QTL) analyses and genome wide association studies (GWAS) across our network of field sites. This work has resulted in the identification of key loci contributing to variation in biomass, flowering time, overwinter survival, microbiome assembly, and resistance to pathogens. The vast majority of these loci have strong genotype x environment interactions, with additive effects varying greatly among field sites. Overall, many of these loci have major positive benefits with minimal fitness trade-offs across field sites. To understand how individual environmental factors contribute to local adaptation, we conducted in-depth laboratory studies of cold acclimation and freezing tolerance. Going forward, we are integrating high-throughput phenotyping, using drones, into our field studies to develop predictive models that can be utilized to increase the productivity of switchgrass.

143T Sexually concoradant and antagonistic genetic variation predicts the evolution of sexual dimorophism over millions of years Jacqueline Sztepanacz Department of Ecology and Evolutionary Biology, University of Toronto

Sexual dimorphism is widely viewed as adaptive, reflecting the evolution of males and females towards divergent fitness optima. Its evolution, however, may often be constrained by the shared genetic architecture of the sexes. Cross-sex genetic covariances determine the extent and timescales on which sexual dimorphism can evolve. Here we start to disentangle the role of multivariate cross-sex covariances in the evolution of sexual dimorphism of wing shape in *Drosophila*. *Drosophila* wing-shape has emerged as a model high-dimensional complex trait; it is highly evolvable in contemporary populations, and yet perplexingly stable across phylogenetic timescales. We show that when we transform the space of genetic variance and cross-sex covariance in wing shape into a space of sexually concordant and antagonistic genetic variation, we tend to find more genetic variation that would allow a response to sexually concordant selection. Using a phylogenetic analysis of sexual shape dimorphism for 82 taxa that have been diverging for at least 33 million years we show that shape dimorphism is qualitatively conserved among species, but that with males characterized by longer thinner wings than females. However, we did find quantitative variation among, with evidence that shape dimorphism has adapted to different evolutionary optima in different clades on timescales of about 10 million years. We also found that allometry constrained the evolution of shape dimorphism for the two most variable multivariate trait combinations we studied, but that shape dimorphism was evolutionary labile in other multivariate trait combinations. Our results highlight that the keys for disentangling alternative explanations for dimorphism evolution are a deeper understanding of how microevolutionary parameters of genetic variation relate to macroevolutionary patterns of divergence, together with studies of natural and sexual selection.

144W Viral load minimally affects the intra-host recombination rate of HIV *Elena Romero*, Alison Feder University of Washington, Seattle, WA

Intra-host recombination generates new mutational combinations which enable immune escape and multidrug resistance in HIV. Because HIV recombines via coinfection, denser HIV populations may have higher rates of coinfection and therefore recombination, but the potential impact of viral density on recombination has not been previously quantified. Using viral load as a proxy for viral density, we analyze longitudinal, high-throughput viral sequencing data from individuals with HIV viral loads varying by orders of magnitude to quantify different recombination rates as a function of viral load. We use two methods to quantify recombination in time series: one based on the autocorrelation of linkage over time and one based on the differential discovery rate of haplotypes likely to be created via recombination as a function of distance. We validate these methods on extensive simulated data and find that the recombination rate estimation is distorted significantly by natural selection and overbinning genetic variation. We then identify a set of best practices for estimating recombination in genetic time series similar to those sampled in intra-host viral evolution. Using these best practices, we estimate effective recombination rates in viral populations with low (<10^4 viral copies/mL) and high (>10^4 viral copies/mL) and find, surprisingly, that viral load does not significantly affect recombination rate. The independence of viral load and recombination rate suggests that recombination is an important driver of viral evolution even in viremic controllers (individuals with setpoint viral load <=2*10^3 copies/mL), and that important viral population parameters can not be well-approximated by convenience sampling in the blood. These results inform our understanding of recombination in viruses more broadly, and may help us understand the structure of where viruses replicate in the body.

145W Understanding the spread of SARS-CoV-2 clusters through an integrated pipeline using UShER, Cluster Tracker and StrainHub *Adriano de Bernardi Schneider*[1], Colby T Ford[2,3], Jakob McBroome[1], Jennifer Martin[1], Daniel Janies[2], Yatish Turakhia[4], Russel Corbett-Detig[1] 1) Biomolecular Engineering and Genomics Institute, University of California, Santa Cruz, USA; 2) Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA; 3) School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, USA; 4) Electrical and Computer Engineering, University of California, San Diego, San Diego, USA

Response by the scientific community to the SARS-CoV-2 pandemic has created an unprecedented amount of genomic data that has to be processed and analyzed in a timely manner to have public health impact. These circumstances push for a change in the current way genomic data has been evaluated in order to assist epidemiologists and public health officials make effective policy changes. To these ends, we integrate a pipeline to bring together three tools currently available and supported by our research groups and evaluate select SARS-CoV-2 clusters identified to understand the spread of these clusters in the United States. Our pipeline consists of three applications: UShER, a program for rapid, accurate placement of viral genomic samples to existing phylogenies, which allows the evaluation of very large datasets in a timely manner; Cluster Tracker, a program which automatically identifies and highlights groups of closely related SARS-CoV-2 infections resulting from inter-regional transmission across the United States through a phylogenetically-informed summary heuristic; and StrainHub, a web-based application that generates transmission networks based on character state changes in metadata mapped to a phylogeny. Strainhub allows the user to visualize networks in a network format, on a map as well as calculate centrality metrics to provide insights on the behavior of network nodes (i.e., source, sink or hub behavior of traits). Using these three tools, we created a workflow that allows the user to identify the genomic sequences of interest with Cluster Tracker, select and extract the sequences and metadata from the underlying dataset using UShER's MatUtils tools through a snakemake workflow, and evaluate the cluster structure and behavior using StrainHub. We evaluated four SARS-CoV-2 large and diverse nodes identified in ClusterTracker using StrainHub in order to present our tools capabilities. This pipeline offers genetics researchers, epidemiologists, and public health officials the tools needed to rapidly reduce the impact of SARS-CoV-2 in our communities. Additionally, this pipeline can be expanded to other pathogens, increasing the reach to more research groups and the scientific community.

146T State-Dependent Evolutionary Phylodynamic Model (SDevo) Infers Boundary-Driven Growth in Hepatocellular Carcinomas Maya Lewinsohn[1,2], Trevor Bedford[2,3], Nicola F. Müller[2,4], *Alison Feder*[1,4] 1) University of Washington, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Howard Hughes Medical Institute; 4) Equal contribution

Spatial properties of tumor growth have profound implications for cancer progression, therapeutic resistance and metastasis, yet how space governs tumor cell division remains an open question. Xenograft and organoid studies suggest that tumors grow preferentially on the periphery (i.e.,"boundary-driven growth"), while sequencing efforts have suggested faster progression in the tumor interior. Boundary-driven growth affects the shape of tumor phylogenies and is therefore theoretically observable from multi-region sequencing data. However, phylodynamic methods have been largely under-utilized to infer growth dynamics in clinical tumors. Here, we show that boundary-driven growth can be well-ap-

proximated by a two-state model permitting different growth rates in the tumor edge and center. To quantify these differential growth rates from sequencing data, we develop a State-Dependent Evolutionary phylodynamic model (SDevo), which links growth rate to tree branching and clock rates as a function of state (here, the edge or center position of the sample). We validate this approach on simulated tumors sampled across multiple spatial regions and demonstrate its ability to quantify spatially-varying growth rates under a range of growth conditions and sampling strategies. We then apply SDevo to multi-region sequencing data from hepatocellular carcinomas and find evidence that these tumors divide more rapidly near the tumor edge than in the center. As multi-region and single-cell sequencing increases in resolution and availability, this approach could interrogate spatial growth dynamics in diverse clinically-resected specimens and be extended to test other two-state growth models, e.g. metastasis or driver gene effects. More generally, this approach demonstrates the potential power of phylodynamic models to quantify tumor evolutionary dynamics.

148W Extensive Trans-Species Polymorphism at the Major Histocompatibility Complex in Primates *Alyssa Lyn Fortier*, Jonathan Pritchard Stanford University

Genes within the Major Histocompatibility Complex (MHC) exhibit exceptional diversity, with thousands of alleles per gene in humans and other primates. Even distantly-related species appear to share similar alleles, pointing to possible long-term balancing selection at this locus. The MHC may even exhibit trans-species polymorphism (TSP), an extremely rare phenomenon in which alleles between species are more similar than alleles within species. Although previous work has suggested TSP at this locus, most have examined sequences from only a single exon, are limited to a handful of species, or fail to quantify support for these trans-species clades. Additionally, most studies on TSPs lack the functional context needed to interpret the results. We aimed to comprehensively characterize the nature and extent of TSP at the classical MHC genes. Using modern data from entire genes and across the primates, we find strong support for TSP in all six classical genes, including between humans and old-world monkeys (OWM) in HLA-DRB1, and even between humans and new-world monkeys (NWM) in HLA-DQB1. Additionally, rapidly-evolving amino acids within each gene are concordant with disease- and immune-phenotype-associated amino acids from the literature. Taken together, definitive TSP and functionally-relevant rapidly-evolving sites suggest complex and competing selective forces and a possible trade-off between disease susceptibility and infection resistance.

147T Predicting Antibiotic Resistance Through the Utilization and Comparison of Machine Learning Algorithms *Meris Johnson-Hagler*, Jameel Ali, Faye Orcales, John Matt Suntay, Kristiene Recto , Lucy Mocteczuma, Feyeeza Shaikh, Pleuni Pennings San Francisco State University

Antibiotic resistance has become a global public health concern. Bacteria are evolving resistance to the current arsenal of prescribed antibiotics resulting in strains that are developing multi-drug resistance. Currently, clinics are often performing traditional culture-based assays to determine antibiotic resistance in bacterial strains. However, this method is time-consuming and may be phenotypically inaccurate. To determine antibiotic resistance with a greater degree of accuracy and efficiency than traditional methods, we will be utilizing machine learning algorithms. The machine learning algorithms will process publically available whole genome sequences of E. coli strains to produce Decision Trees, Random Forest, and Gradient Boosted Trees models. We want to compare the machine learning models to determine which one has the best accuracy when using population structure, isolation year, and gene content as features. Through comparative analysis, we want to identify which features can predict antibiotic resistance.We aim to use what we have learned from this study to contribute to a future where machine learning can be used as a diagnostic tool to accurately predict antibiotic resistance from whole genome sequencing data.

149W Treenome Browser: concurrent phylogeny-aware visualization of millions of genomes *Alexander Kramer*, Russell Corbett-Detig University of California, Santa Cruz

The ongoing pandemic led to an unprecedented global sequencing effort that has yielded over seven million complete SARS-CoV-2 genomes. This massive dataset presented important challenges for data exploration and visualization. By taking advantage of the evolutionary redundancy among sequences, the mutation-annotated tree (MAT) data format[1] uses phylogenetic compression to enable efficient storage and traversal of such large datasets. A MAT encodes the complete genetic variation of its component sequences, storing the same information as a multiple sequence alignment or VCF file in a fraction of the space. This reduction in size along with the phylogenetic structure makes visualization of very large trees and the underlying genome sequences feasible. We developed Treenome Browser, a genome browser for viewing millions of genomes in their phylogenetic context. Treenome Browser uses the MAT format to display the amino acid mutations present in each genome alongside a tree that displays mutations in the full phylogeny, remaining performant on trees of over 7M genomes. The MAT is traversed at run-time to quickly reconstruct and display mutations accumulated in samples relative to the root of the tree. Interactive exploration of the tree (provided by Taxonium[2]) reveals muta-

tion patterns that arise in Treenome Browser at the levels of clades and individual genomes, allowing visual identification of the defining signatures of SARS-CoV-2 strains, for example. Treenome Browser has potential applications in phylogenetically-informed primer design and variant monitoring of SARS-CoV-2. This application empowers data visualization at unprecedented scale for SARS-CoV-2 and future similarly enormous datasets.

1. Turakhia, Y. *et al.* Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat. Genet.* **53**, 809–816 (2021).
2. https://github.com/theosanderson/taxonium

150T The story behind the strains: Examining the phylogeography of wild yeast from woodlands *Jacqueline Pena*[1], Eduardo Scopel[2], Douda Bensasson[1,2] 1) Department of Plant Biology, University of Georgia, Athens, GA; 2) Institute of Bioinformatics, University of Georgia, Athens, GA

The model species, *Saccharomyces cerevisiae*, is the workhorse for scientific research and biotechnology because of its tractability and short generation time. With advances in sequencing technology, genomes are paving the way to resolve the natural history of *S. cerevisiae* and other *Saccharomyces* species. The phylogeography of wild *S. cerevisiae* is challenging because this species has been heavily influenced by domestication, which increases dispersal and reduces geographic structure. To better understand the migrations of wild *S. cerevisiae*, we focused our study on *S. cerevisiae* isolates from woodlands. We asked: Do woodland populations show genetic substructure that is consistent with isolation by distance? We used over 300 publicly available whole-genome sequences to determine the population structure and phylogenetic relationships among woodland yeast. We were able to recapitulate known phylogenetic relationships of domesticated *S. cerevisiae* lineages and found that woodland populations from around the world are genetically distinct. This suggests that population structure of natural *S. cerevisiae* has not been drowned out by the effects of domestication. Strains from woodlands can therefore serve as good models for evolution and ecology, for example to assess local adaptation.

151T The repeated evolution of multiple traits between forest and prairie ecotypes of the deer mouse *Brock Wooldridge*, Sade McFadden, Chris Kirby, Kemi Ashing-Giwa, Hopi Hoekstra Department of Organismic and Evolutionary Biology, Harvard, Cambridge, MA; Department of Molecular and Cellular Biology, Harvard, Cambridge, MA; Museum of Comparative Zoology, Harvard, Cambridge, MA; Howard Hughes Medical Institute, Harvard, Cambridge, MA

Local adaptation often involves changes in more than one trait. One classic example is the evolution of forest and prairie ecotypes within a single species of deer mouse (*Peromyscus maniculatus*), which have evolved striking morphological differences. First reported over a century ago, some populations of forest-dwelling mice were described as having longer tails, larger feet, bigger ears, and elongated whiskers relative to prairie populations. However, the extent to which these morphological differences are correlated with habitat across their range and if these multi-trait differences have evolved multiple times independently remains largely unknown. To answer these questions, we measured both phenotypic and genetic variation in deer mice from across North America. Using satellite-based land cover estimates and museum specimens, we found widespread and significant correlation between habitat type and morphological variation, suggesting these morphological traits broadly represent adaptation to local environments. Next, whole-genome sequencing of 177 mice from 44 of those populations reveals two contrasting patterns. First, we find a significant effect of genetic isolation-by-environment (IBE): more closely related populations are more likely to be phenotypically similar given the same geographic distance. Nonetheless, we were still able to identify several independent gains and losses of ecotype-specific morphological adaptations. Together, these data point towards a set of independently evolved forest-prairie population pairs, which consistently differ in morphology, yet have high levels of gene flow, thus representing an exciting system to study the repeated co-evolution of multiple adaptive traits in a single widespread species.

152W A high-resolution map of Drosophila hybrid pairing connects BLACK heterochromatin to pairing loss, reproductive incompatibility, and DNA underreplication *James Baldwin-Brown*, Nitin Phadnis University of Utah

Homologous chromosome pairing is essential to all eukaryotes, but we do not fully understand the molecular mechanisms underlying pairing. Although pairing is often associated with meiosis, it also occurs in somatic cells. Complete somatic pairing is the wild type state in Drosophila and other dipterans. More that 80 years ago, researchers studying Drosophila between-species hybrids discovered that somatic chromosome pairing broke down in hybrid individuals, with no explanation as to which regions were breaking down or why. Because pairing machinery exists in all eukaryotes, finding the genomic elements that drive this non-pairing will help us understand the drivers of pairing generally.
Technical hurdles to measuring pairing rates genome-wide have prevented most investigation into hybrid pairing breakdown. In the past few years, however, new technologies, principally Hi-C, have allowed for high-resolution measurement

of pairing across the genome by directly measuring physical contacts between chromosomes. To find the mechanistic basis of hybrid pairing breakdown, we used Hi-C in hybrids of D. melanogaster and D. simulans to measure pairing rates with unprecedentedly high resolution across the genome. Compared to within-species crosses, this hybrid Hi-C shows dramatic regions of high and low pairing. Contrary with expectations, pairing rates did not correlate with sequence similarity. Instead, they correlated with the presence of BLACK chromatin. This has important implications for understanding speciation: chromatin state, rather than sequence, drives loss of pairing between species.
To better understand the effect of tissue on pairing, we repeated our pairing measurements in multiple tissues. Contrary to expectations, we found that hybrid pairing loss uniquely affects polytene nuclei but not diploid nuclei. This raises the possibility that DNA overreplication influences pairing breakdown, a hypothesis bolstered by the fact that BLACK chromatin is the last chromatin to replicate during mitosis. Our results suggest rapid evolution of genomic regions that alter chromatin state and replication timing may contribute to hybrid dysfunction. We also hypothesize a connection between polytene pairing breakdown and hybrid incompatibility genes in D. melanogaster and D. simulans, as pairing breakdown is substantially reduced by knockout of incompatibility genes. Our current work focuses on the consequences of pairing breakdown on gene expression through the phenomenon of transvection (a process where cis-regulators can act in trans when paired). Together, these studies tease apart whether hybrid pairing breakdown is a cause or an effect of reproductive incompatibility. Here, I describe our insights into fundamental evolutionary phenomena at the intersection of chromosome pairing, reproductive isolation, gene expression, and chromatin state evolution.

153W Piecing Together the Periodical Cicada Puzzle *Robert Bush*, Paul Frandsen Brigham Young University

Periodical cicadas of the genus *Magicicada* are commonly found throughout the Eastern United States, and their loud mating calls can be heard every 13 or 17 years as various broods of nymphs emerge in massive numbers. These cicadas have a tremendous impact on their ecosystems, including serving as a critical food source for many species during years in which they emerge. *Magicicada* emergences have even been known to have measurable impacts on avian abundance in the years following large broods. Despite the important roles these insects play, researchers have struggled to make sense of the many unique characteristics of these peculiar insects, including their abnormally lengthy life cycles, population structure, and evolutionary history. Studies have suggested that low effective population sizes (N_e) might help explain observed life cycle patterns and subsequent speciation. To investigate these questions, we analyzed the recently published whole genome sequences of two 17-year periodical cicada species (*M. septendecim & M. septendecula*) from Brood X. We estimated their effective population sizes over time using the Pairwise Sequential Markovian Coalescent (PSMC) analysis. Additionally, we conducted quality assessments on the current genomic resources to investigate the strength of the inferences made. Finally, we share insights learned from our current work on sequencing, assembling and annotating the genome of the third Brood X species (*M. cassini*). We show here that whole genome analyses can provide unique and valuable insights into answering quesions surrounding the evolutionary history and population growth of periodical cicadas.

154T Identifying patterns of introgression in two species pairs of Texas *Phlox* *Samridhi Chaturvedi*[1,2,3], Danielle Khost[3], Austin Garner[2,3], Ben Goulet-Scott[2,3], Antonio Serrato-Capuchina[2,3], Tim Sackton[3], Robin Hopkins[2,3] 1) University of California, Berkeley; 2) The Arnold Arboretum at Harvard University, Roslindale, MA; 3) Harvard University, Cambridge, MA

Many diverging lineages have a history of hybridization and gene flow. Understanding these patterns of gene exchange across space and across their genomes can provide invaluable insights into how and when species form and diverge. Patterns of genomic diversity and divergence within and between species across space can be used to infer evolutionary histories of gene flow and admixture between species. Here, we study patterns of introgression between three Texas *Phlox* wildflowers – *P. drummondii, P. cuspidata,* and *P. roemeriana*. These three species live in adjacent and overlapping geographic ranges with sympatric zones shared between *P. drummondii* × *P. cuspidata* and *P. drummondii* × *P. roemeriana*. Notably, in the sympatric zone between *P. drummondii* and *P. cuspidata*, reinforcement has driven the evolution of flower color as a mechanism to decrease costly hybridization between the two species. No such pattern of reinforcement is evident in the other sympatric zone. In this study, we use genome-wide sequence data to estimate introgression between these two pairs of species – one with reinforcement (*P. drummondii x P. cuspidata*) and one pair without (*P. drummondii x P. roemeriana*). We combined extensive geographic sampling, greenhouse experiments, partial and whole-genome sequencing of multiple individuals, to study the history of admixture for each species pair and identify the direction and extent of introgression. Our results provide strong evidence of introgression and admixture between sympatric species with striking variation across populations depending on geographic location. This study characterizes patterns of introgression and local ancestry in these species at a genomic level.

155T Patterns of population structure and polymorphic reproductive isolation in Drosophila melanogaster *Jenn Coughlan*[1,2], Andrius Dagilis[1], Antonio Serrato-Capuchina[3], David Peede[4,5], Dean Castillo[6], Brandon Cooper[7], Daniel Matute[1] 1) University of North Carolina @ Chapel Hill; 2) Department of Ecology and Evolutionary Biology, Yale University; 3) Arnold Arboretum, Harvard University; 4) Department of Ecology, Evolution, and Organismal Biology, Brown University; 5) Center for Computational Molecular Biology, Brown University; 6) Biology Department, University of Nebraska, Omaha; 7) Division of Biological Sciences, University of Montana

Despite a century of genetic analysis, the evolutionary history underlying patterns of exceptional genetic and phenotypic variation in the model organism *Drosophila melanogaster* remains poorly understood. How genetic and phenotypic variation is partitioned across the range of *D. melanogaster*, particularly in its putative ancestral range in Subtropical Africa, remains unresolved. Here, we assess patterns of population genetic structure, admixture, mate preference, and genetic incompatibility across a global sample of *D. melanogaster*, including 174 new accessions from remote regions within Subtropical Africa. While almost all Out of Africa genomes correspond to few genetic ancestries, different geographic regions within Africa contain multiple ancestries, with substantial cryptic diversity in Subtropical Africa. Admixture between distinct lineages is prevalent across the range, but admixture rates vary between lineages. We next perform large scale behavioral surveys of flies from Subtropical Africa and find that female mate choice within Subtropical Africa is highly polymorphic and behavioral types are not monophyletic. This suggests either the repeated evolution of strong female mate choice or substantial admixture between mating types. By quantifying branch specific evolution for flies with strong female choice, we find that the genetic architecture of mate choice is highly polygenic, including loci associated with neurological development, behavior, olfactory perception, and learning. By pairing this population genetic survey with transgenic assays, we explicitly test the role of these candidate loci on behavioral phenotypes, adding to a limited list of genes known to contribute to female mate choice. Additionally, we find that many segregating putative incompatibilities likely evolved during or after expansion out of Africa. However, we find that neither female mate choice nor putative incompatibilities can fully explain overall patterns of population genetic structure, and many of these loci introgress freely between geographic ancestry types. This work contributes to our understanding of the evolution of polymorphic reproductive isolation and how it may contribute to patterns of population genetic structure in a key model organism.

156W Hybrid seed inviability maintains species barriers in *Diplacus (Mimulus)* sect. *Eunanus* *Matthew Farnitano*, Andrea Sweigart University of Georgia

Postzygotic reproductive barriers are important in the maintenance of species barriers over long time scales, though their role as a driver of speciation is controversial. Comparing across a larger number of taxa helps us to better understand which postzygotic barriers act at what stages of divergence, and how their distribution might influence patterns of speciation. *Mimulus* monkeyflowers are a model genus for studies of speciation, but most work focuses on a few well-known clades. I have conducted crosses between populations of four sympatric species in *Diplacus* sect. *Eunanus*, an understudied clade of *Mimulus*, to determine the distribution of postzygotic crossing barriers in this group. In addition, I generated the first genome-scale SNP dataset for these species. I find considerable genetic divergence between morphologically similar taxa and limited signatures of historical introgression between groups. Seed inviability is a near-complete hybrid barrier in all but one comparison (*D. fremontii* x *D. brevipes*, which in turn has reduced pollen fertility in viable hybrids). Seed phenotypes show parent-of-origin effects, suggesting a possible role for parental conflict in seed inviability. Three of the studied species have remarkably similar floral morphology, suggesting seed inviability may be more important than pollinator-mediated barriers in reducing gene flow. In total, these data suggest an important role for post-zygotic barriers in explaining the diversity of the *Mimulus* radiation.

157W Identification of reinforcement mutations with targeted long-read sequencing in Phlox *Austin Garner*[1,2], Angie Diana[1,2], Danielle Khost[3], Timothy Sackton[3], Robin Hopkins[1,2] 1) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 021382; 2) The Arnold Arboretum, Harvard University, Boston, MA 02131; 3) Informatics Group, Harvard University, Cambridge, MA 021382

Natural selection directly favors the evolution of reproductive isolation between species through the process of reinforcement. During reinforcement, costly hybridization between co-occurring lineages can generate selection favoring traits that increase reproductive isolation. This important process has generated species across the tree of life, yet the mutations underlying this process and how they evolved by selection have remained unknown. Flower color variation in Phlox drummondii is the best documented case of reinforcement to date. P. drummondii and P. cuspidata display similar light-blue colored flowers in allopatry. However, in sympatry, P. drummondii has evolved dark-red flowers via selection to prevent maladaptive hybridization with P. cuspidata. Previous work inferred the divergence from light to dark colored

flowers is controlled by cis-regulatory mutations at an R2R3-Myb transcription factor. We have applied population-level targeted long read genome sequencing in P. drummondii to identify and characterize mutations in the regulatory sequence of the R2R3-Myb associated with the transition from light to dark colored flowers. Through our sequencing efforts, we have begun to infer the evolutionary history of reinforcement mutations through space and time. This is the first characterization of the mutations underlying reinforcement, and a powerful demonstration of how targeted sequencing can elucidate the molecular basis of trait variation in non-model systems with large genomes.

158T The temporal and genomic scale of selection against hybrids *Jeffrey Groh*, Graham Coop University of California, Davis

Genomic data from hybrid populations contain valuable information about how selection, demography, and gene flow shape divergence and introgression between species. Genome-wide correlations between ancestry state and recombination rate in a growing number of hybrid systems provide evidence of selection broadly acting to remove ancestry of the species contributing the minority of alleles. However, it is unknown to what extent local ancestry patterns reflect selection vs. neutral processes, nor the time-scales of selection on hybrids. To address these questions, we demonstrate the use of the wavelet transform to partition the ancestry variance across loci, as well as the correlation between ancestry and recombination, into contributions from different spatial genomic scales. We use both theory and simulations to illustrate how temporally localized effects of drift and selection are embedded in the spatial decomposition of ancestry variance across loci. We detect signals of strong selection on early generation hybrids in several population genomics datasets. Extending this approach, we estimate the amount of ancestry variance explained by selection across genomic spatial scales, with estimates above 20% at the largest genomic scales. These methods provide a more detailed look into the timescale of selection on hybrids and should be widely applicable as more genomic data sets from hybrid populations are generated.

159T Hotspots of disruption in placental regulatory gene networks reflect a common genetic architecture underlying hybrid placental dysplasia in rodents *Emily Moore*[1], Fernando Rodriguez-Caro[1], Kathryn Wilsterman[2], Quynh McKelvey-Pham[3], Jeffrey Good[1] 1) University of Montana, Missoula, MT; 2) Colorado State University, Fort Collins, CO; 3) Hellgate High School, Missoula, MT

Mammalian hybrids often show abnormal, parent-of-origin specific placental growth, suggesting that disruption of this critical developmental tissue plays an important role in the early stages of speciation. The placenta is among the most rapidly evolving mammalian organ and is enriched for genes that show parent-of-origin bias in expression due to epigenetic imprints, making divergence in imprinted placental gene pathways a prime candidate for the rapid evolution of hybrid inviability. To gain insight into the genomic bases of hybrid placental disruption, we compared placenta size, tissue composition, pregnancy failure, and gene co-expression network structure for house mouse (*Mus*) hybrids at increasing levels of evolutionary divergence. Using reciprocal crosses between wild-derived inbred strains of *M. musculus musculus*, *M. musculus domesticus*, and *M. spretus,* we tracked parent-of-origin gene expression within each lineage as well as in F1 hybrids between subspecies and species. To evaluate co-expressed and imprinted genes, we used transcriptomes and methylomes from whole placenta, augmented by transcriptomes from functionally distinct placental layers. We identified previously cryptic placental and pregnancy phenotypes between house mouse subspecies, revealing parent-of-origin incompatibilities between these closely related lineages. Placental growth abnormalities and patterns of regulatory disruption scaled with lineage divergence, with co-expressed sets of genes showing transgressive expression in hybrid placentas. Comparison between F1 experiments revealed a core set of disrupted genes shared between subspecific and specific *Mus* crosses and enriched for fetally expressed genes at junction of the fetal-maternal interface. Using a backcross between *M. musculus musculus* and *M. spretus*, we then showed that extreme placental overgrowth and disrupted placental pathways were primarily caused by genetic incompatibilities on the maternal X chromosome. Comparison of these data with results from dwarf hamster (*Phodopus*) hybrids revealed a parallel X-linked genetic architecture of placental incompatibilities mapping to a homologous region of the X chromosome and resulting in disrupted expression of a common imprinted autosomal gene network. These findings demonstrate that the placenta is a hotspot for evolutionary divergence contributing to rodent speciation, and that recurrent patterns of hybrid inviability may reflect common genetic and regulatory bases.

160W Hybridization alters the shape of the genotypic fitness landscape, increasing access to novel fitness peaks during adaptive radiation *Austin Patton*[1], Emilie Richards[1], Katelyn Gould[2], Logan Buie[2], Christopher Martin[1] 1) University of California, Berkeley, CA; 2) University of North Carolina, Chapel Hill, NC

Estimating the complex relationship between fitness and genotype or phenotype (i.e. the adaptive landscape) is one of the central goals of evolutionary biology. Empirical fitness landscapes have now been estimated for numerous systems,

from phage to proteins to finches. However, adaptive walks connecting genotypes to organismal fitness, speciation, and novel ecological niches are still poorly understood and processes for surmounting fitness valleys remain controversial. One outstanding system for addressing these connections is a recent adaptive radiation of ecologically and morphologically novel pupfishes (a generalist, molluscivore, and scale-eater) endemic to San Salvador Island, Bahamas. Here, we leveraged whole-genome sequencing of 139 hybrids from two independent field fitness experiments to identify the genomic basis of fitness, estimate genotypic fitness landscapes, and measure the accessibility of adaptive walks. We identified 132 SNPs that were significantly associated with fitness in field enclosures. Six fitness-associated regions contained differentially expressed genes and fixed SNPs between trophic specialists; one gene (*mettl21e*) was also misexpressed in lab-reared hybrids, suggesting a potential intrinsic genetic incompatibility. We then constructed genotypic fitness networks from adaptive alleles and show that scale-eating specialists are the most isolated of the three species on these networks. Intriguingly, introgressed and *de novo* variants altered the topography of the fitness landscape, increasing the accessibility of genotypic fitness paths from generalist to specialists as compared to standing variation. Our results suggest that adaptive introgression and *de novo* mutations alter the shape of the fitness landscape, mitigating the need to cross adaptive valleys in adaptive walks, thus triggering the evolution of novelty during adaptive radiation.

161W Color adaptation during the repeated domestication of grain amaranth Tom Winkler[1], José Gonçalves-Dias[1], *Markus Stetter*[1,2] 1) Institute for Plant Sciences, University of Cologne; 2) Cluster of Excellence on Plant Sciences, University of Cologne

Colors fulfill important functions that allow populations to adapt to their environment. In addition to their obvious functions, e.g., camouflage, pollinator attraction and UV protection, colorants can alter physiological traits in plants. Hence, when observing color changes as result of adaptation, the actual target trait of selection remains often unclear. The domestication of plants is a well-suited model to study plant adaptation and to identify the genetic basis and regulatory networks of adaptation. While most wild plants were domesticated only once, others were recurrently selected in different geographic locations. These species represent ideal models to study how repeatable evolution is and what determines its outcome. We study the ancient grain crop amaranth which has been domesticated three times in Central and South America from one wild ancestor. All three grain species display a distinct seed color compared to wild *Amarnathus* species. While all wild amaranths have dark seeds, cultivated amaranths have pale seeds. Despite strong gene flow between crop species and wild relatives, the pale seed color has likely been selected independently, on different genetic backgrounds but altering the same genomic regions. We were able to map the genetic control of the seed color adaptation to two genomic regions and identify a MYB transcription factor gene as potential regulator for the seed color change. A long-read transcriptome assembly and differential gene expression analysis identified variation in flavonoid-pathway genes between individuals with pale and dark seeds. Our results link the genomic change in a transcription factor gene to altered expression in color pathway genes and metabolites with multiple physiological functions. We show that white seeds have reduced seed dormancy, which likely increased their fitness in agricultural environment. While the color itself might have been under selection due to human preference, we speculate that the actual trait under selection was seed dormancy. This shows that trait changes can be the result of pleiotropic effects of metabolic networks rather than selection on the observed trait.

162T Recreating the mitochondrial endosymbiosis that gave rise to eukaryotes *Cara B. Hull*[1], Shawn Yang[1], Alessandro L. V. Coradini[1], Wan-Zhen Sophie Lin[3], Lucia C. Dalle Ore[3], Noah Malmstadt[2,3,4], Ian M. Ehrenreich[1] 1) Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA; 2) Department of Chemistry, University of Southern California, Los Angeles, CA; 3) Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA; 4) Department of Biomedical Engineering, University of Southern California, Los Angeles, CA

The evolution of mitochondria played a key role in the origin, success, and diversification of eukaryotic life. How this organelle arose and integrated into host cells are fundamental questions in biology. To investigate these questions, we will recreate intermediate steps of nuclear-mitochondrial coevolution. To do this, we have developed methods to clone and genetically engineer intact mitochondrial genomes in yeast. We are now working to transplant these mitochondrial genomes into living cells. I will report on our ongoing progress.

163T Genetic Analysis of Segregating Recessive Variation in the Nematode *Caenorhabditis becei* *Jose Salome-Correa*[1], Solomon Sloat[1], Luke Noble[2], Matthew Rockman[1] 1) New York University; 2) Institut de Biologie École Normale Supérieure

Segregating recessive variants have a substantial influence on biological processes. Outcrossing species tend to accumulate a high number of low-frequency moderately deleterious recessive variants, as their negative effects are masked

by dominant alleles. Recessive variation has remained largely understudied, as low-frequency recessive variants have a small additive contribution, limiting statistical power in quantitative analyses. There is a need for an experimental system in which low-frequency variation can be made more common, recessive alleles can be made more homozygous, and recessive effects can be distinguished from additive effects. Here we present a new experimental evolve-and-resequencing approach to analyze recessive variants through the construction of recombinant inbred line crosses with varying degrees of relatedness and homozygosity. The gonochoristic nematode *Caenorhabditis becei* provides a powerful animal model for the study of recessive deleterious variation due to its ideal properties: Short generation time, the ability to cryopreserve living stocks and replicate genotypes, a small genome with a high-quality chromosome level assembly, and manageable levels of inbreeding depression, allowing us to create the first recombinant inbred panel for any gonochoristic *Caenorhabditis*. We have generated more than 400 recombinant inbred lines via 25 generations of sibmating starting from outbred wild founders. We have sequenced founding individuals, their F1 progeny, and all recombinant inbred lines. We are generating an annotated catalog of segregating variants, phasing haplotypes in the founder genomes, genotyping every recombinant inbred line, and characterizing the recessive deleterious variation exposed in our experimental evolution design.

164V Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes *Ching-Ho Chang*[1,2], Lauren Gregory[1], Kathleen Gordon[3], Colin Meiklejohn[3], Amanda Larracuente[1] 1) Department of Biology, University of Rochester, Rochester, NY 14627; 2) Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; 3) School of Biological Sciences, University of Nebraska-Lincoln, NE 68502

Y chromosomes across diverse species convergently evolve a gene-poor, heterochromatic organization enriched for duplicated genes, LTR retrotransposons, and satellite DNA. Sexual antagonism and a loss of recombination play major roles in the degeneration of young Y chromosomes. However, the processes shaping the evolution of mature, already degenerated Y chromosomes are less well-understood. Because Y chromosomes evolve rapidly, comparisons between closely related species are particularly useful. We generated de novo long read assemblies complemented with cytological validation to reveal Y chromosome organization in three closely related species of the *Drosophila simulans* complex, which diverged only 250,000 years ago and share >98% sequence identity. We find these Y chromosomes are divergent in their organization and repetitive DNA composition and discover new Y-linked gene families whose evolution is driven by both positive selection and gene conversion. These Y chromosomes are also enriched for large deletions, suggesting that the repair of double-strand breaks on Y chromosomes may be biased toward microhomology-mediated end joining over canonical non-homologous end-joining. We propose that this repair mechanism contributes to the convergent evolution of Y chromosome organization across organisms.

165V STR mutation rates do not perfectly track cell divisions but covary with maternal age *Michael E Goldberg*[1], Evan E Eichler[1,2], Kelley Harris[1,3] 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Howard Hughes Medical Institute, University of Washington, Seattle, WA; 3) Computational Biology Division, Fred Hutchinson Cancer Center, Seattle, WA

Short tandem repeats (STRs) are hotspots of genomic variability because of their high mutation rates, which have long been attributed to polymerase slippage during DNA replication. This model suggests that STR mutation rates should scale linearly with the number of cell divisions in male and female germlines. In particular, STR mutation rates are not predicted to scale with the age of the mother at conception, since oocytes spend a mother's reproductive years arrested in meiosis II and undergo a fixed number of cell divisions prior to ovulation that is independent of age. We tested this prediction using de novo mutation calls from the Simons Simplex Collection, a cohort of nearly 2300 human quad families, each consisting of two children plus parents whose ages at conception are known. Contrary to expectations, STR mutation rates covary with maternal age as well as paternal age, implying that some STR mutations are caused by DNA damage in quiescent cells rather than the classical mechanism of polymerase slippage. Our results echo the recent finding that DNA damage in quiescent oocytes is a significant source of de novo SNVs, but are especially surprising in light of the prior belief in replication slippage as the dominant mechanism of STR mutagenesis. We find that homopolymer STRs have a smaller maternal age effect than STRs of longer repeat unit lengths, and that the maternal age effect is not confined to previously discovered hotspots of oocyte mutagenesis. Our results suggest that STR mutagenesis cannot be fully explained by replication slippage, but is influenced by the DNA damage affecting quiescent cells that has recently been shown to generate a significant fraction of point mutations.

166V Parallel Expansion and Divergence of the Hyr/Iff-like (Hil) Adhesin Family in Pathogenic *Candida* Yeasts *Bin He*, Jan Fassler, Lindsey Snyder, Rachel Smoak University of Iowa, Iowa City, IA

Opportunistic yeast pathogens evolved multiple times in the Saccharomycetes class, such as the recently emerged *Candi-*

da auris, a multidrug resistant pathogen associated with multiple hospital outbreaks. Genomic changes shared between independently evolved pathogens could reveal key factors that enable them to infect the host. Yeast adhesins are cell wall proteins that mediate biofilm formation and adherence and are established virulence factors in *Candida* spp. Here we show that homologs of a known adhesin family in *C. albicans*, the Hyr/Iff-like (Hil) family, repeatedly expanded in divergent pathogenic *Candida* lineages including in *C. auris*. The majority of the Hil family genes (~75%) have sequence features consistent with known yeast adhesins. Evolutionary analyses reveal varying levels of selective constraint and a possible role of positive selection on the ligand-binding domain during the expansion of the family in *C. auris*. The repeat-rich central domain evolved rapidly after gene duplication, leading to large variation in protein length and β-aggregation potential, both known to directly affect adhesive functions. Within *C. auris*, isolates from the less virulent Clade II lost five of the eight Hil genes, while other clades show abundant tandem repeat copy number variation. Based on these results, we hypothesize that expansion and diversification of adhesin gene families are a key step towards the evolution of fungal pathogens and their diversification could contribute to variation within and between *Candida* species in their adhesive and virulence properties.

167V The genetic basis of inherited DNA methylation variation in *Arabidopsis thaliana* *Eriko Sasaki*[1,2], Magnus Nordborg[2] 1) Kyushu University; 2) Gregor Mendel Institute of Molecular Plant Biology

The epigenome, in a particular variation of DNA methylation profiles across individuals, has long been of interest as a modifier of the genetic code, with "mutations" reflecting past environments, stochastic events, or genetic regulation. To address this issue, we dissected an inherited epigenetics mark, mCHG methylation, using conditional GWAS approaches in *Arabidopsis thaliana*. Our study revealed the genome-wide mCHG levels largely share the variation with *de novo* methylation and are under the control of major *trans*-modifiers, including the key regulators *CMT2*, *CMT3*, *MIR823A*, and a novel regulator *JMJ26* that specifically regulated RdDM-targeted TEs. These *trans*-modifiers could affect natural variation of transposon activity via interacting with a previously identified modifier of *de novo* methylation.

168V Transformation-mediated chromosome synthesis and replacement in eukaryotic cells *Alessandro L. V. Coradini*, Cara B. Hull, Joshua Roemer, Daniel T. Lusk, Zachary Krieger, Ian M. Ehrenreich University of Southern California, USC, Los Angeles, CA

Genetic manipulation is one of the central strategies that biologists use to investigate the molecular underpinnings of life and its diversity. In recent years, the synthesis of chromosomes, known as synthetic genomics, has emerged as a new form of genetic manipulation. Megabase-sized chromosomes can now be generated from components synthesized *de novo*, obtained from naturally occurring genomes and other existing molecules, or a mixture of the two. Although *de novo* chromosome synthesis starting from oligonucleotides has been used in most synthetic genomics projects to date, this approach is too laborious and expensive to apply to many biological problems. To facilitate more widespread use of synthetic genomics, we developed an efficient method of chromosome synthesis and replacement in yeast based on the capture and reassembly of natural DNA segments. Here, I will describe the method and its application to questions about chromosome architecture and the genetic basis of phenotypic differences between strains, species, and potentially even genera.

169V Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes *Caroline Weisman*[1], Sean Eddy[2], Andrew Murray[2] 1) Princeton University; 2) Harvard University

Comparing the genomes of different species regularly reveals genes that seem to be unique to one or a few related species. These "lineage-specific" genes are often thought to represent genetic novelty, with many potentially interesting consequences; for example, these genes have been proposed to underlie species- or taxon-specific evolutionary innovations. The comparative analyses from which these genes emerge often use genome sequences from which genes have been inferred, or annotated, using a mixture of different methods. Using different methods to annotate different species increases the risk that orthologous DNA sequences of the same coding status, actually both genic or both non-genic, have been erroneously annotated in one species but not in another, merely appearing to be lineage-specific. Here, we quantitatively evaluate the impact of this effect, which we term "annotation heterogeneity," in four case studies. We find that annotation heterogeneity consistently, and often substantially, increases the apparent number of lineage-specific genes, suggesting that it may be a source of substantial artifact.

170V Second time's the charm: adaptive evolution following a prior invasion increases the potential distribution of an invasive weed *Andhika Putra*[1], Paul Battlay[2], Kathryn Hodgins[2], Alexandre Fournier-Level[1] 1) University of Melbourne, Melbourne, VIC; 2) Monash University, Melbourne, VIC

The global spread of invasive species is a major source of ecological and financial concern, but what determines where

invaders end up establishing? We address this question by examining the role of genetic variation on the invasion success of common ragweed *Ambrosia artemisiifolia*. Ragweed is a globally invasive species known for releasing highly allergenic pollen. Native to North America, it was first introduced to Europe before invading Australia's east coast in the 1930s. The Australian distribution of this species is conspicuously narrow compared to North America and Europe, prompting an investigation into the factors driving this discrepancy. Genomic data was used to identify ancestral genomic clusters and build genetically-informed ecological niche models to characterize intraspecific niche differentiation. Genetic clusters showed significant differences in habitat preference with respect to temperature and precipitation, indicating a strong association between genomic and environmental variation. Projecting the habitat preferences of native North American and old invasive European clusters to Australia showed the potential distribution of ragweed on the continent is greater than currently observed, but only when European plants are introduced to the country. North American plants had a surprisingly narrow potential distribution, suggesting it was less adapted to Australian conditions than plants from Europe. These findings highlight the role of genomic variation in successful invasions, and suggest adaptive evolution in previous invasive ranges can produce populations with greater invasive potential than those from the native range.

171V Characterizing Pareto fronts: Trade-offs in the yeast growth cycle constrain adaptation *Jason Tarkington*, Gavin Sherlock, Angelina Chan Stanford University

Adaptive evolution involves optimizing multiple fitness related traits simultaneously. These traits can be projected into a multidimensional space known as trait space. Mapping the trait space accessible by single mutations reveals the constraints imposed on organismal fitness. The fitness of an organism in an environment is often dependent on more than one trait; for yeast these traits include **fermentation**, **respiration**, and **stationary phase** performance. While in some cases it may be possible to optimize multiple fitness related traits independently, certain fitness-related traits can also be constrained due to the pleiotropic effects of other fitness related traits, resulting in a trade-off. Previous work from our lab has shown that following evolution in a glucose containing carbon limited media, pareto fronts, indicative of underlying trade-offs, emerge between stationary phase and respiration, and between respiration and fermentation, though not between stationary phase and fermentation. Here we aim to understand why such trade-offs emerge among some fitness-related traits but not others. We are evolving barcoded yeast in a non-fermentable carbon source with varying amounts of time spent in stationary phase. This experimental design eliminates selection for fermentation entirely and creates varying degrees of selection for performance in respiration and stationary phase e.g., the two-day transfer regime selects primarily for respiration performance, while in the 10-day transfer regime stationary phase performance will be more important. Under these conditions stationary phase performance may be free to increase unconstrained by fermentation performance resulting in the emergence of a pareto front between these components of organismal fitness. Freedom from fermentation performance constraints may also allow yeast to maximize respiration and stationary phase simultaneously. In addition to the phenotypic analysis, we will also characterize the molecular basis of the underlying adaptive mutations that emerge.

172V Tissue-specific regulatory evolution involved in divergent migratory behavior *Matthew Louder*, Hannah Justen, Kira Delmore Texas A&M University

Behavioral traits are critical for adaptation to different environments and can lead to reproductive isolation. Yet, how the regulation of gene expression is involved in the evolution of behavioral traits and reproductive isolation remains unclear. Here we examine gene expression in two subspecies of Swainson's thrush (*Catharus ustulatus*) that take different routes on migration and their hybrids that exhibit intermediate and ecologically-inferior behaviors to pure forms. We use RNAseq data to compare expression patterns between seasons and taxonomic groups, focusing on five brain regions with predicted associations to migratory behavior. We find that the majority of genes are differentially expressed between seasons are specific to a subspecies and brain region, with a large proportion being limited to the cluster N hyperpallium and hypothalamus. Genes that are differentially expressed between subspecies, but not seasonally responsive, are additively expressed in hybrids and exhibit allele specific expression (ASE), indicative of divergence in cis regulatory regions between subspecies. However, seasonally responsive genes that are specific to a subspecies are predominantly regulated by trans mechanisms. Furthermore, genes putatively involved in divergent migratory behavior between subspecies do not exhibit transgressive expression patterns in hybrids. These results indicate that divergence in trans regulation can be important mechanism in the evolution of behavior and reproductive isolation.

173V Selection inference on epigenetic marks: Implications for the evolution of germline mutation rates Leandros Boukas[1], Afrooz Razi[1], Hans Bjornsson[1,2], *Kasper Hansen*[1] 1) Johns Hopkins University; 2) University of Iceland

Natural selection is one of the fundamental forces shaping the evolution of living organisms, and detecting its presence is a major goal of evolutionary biology. So far, the almost singular focus of the field has been on identifying signatures of

selection on DNA sequence. However, natural selection acts on phenotypes, and DNA selection signatures only reflect the extent to which the genomic sequence causally affects phenotype. Since this causal effect of DNA is often mediated via intermediate molecular features, it is of great interest to move beyond DNA and identify signatures of selection directly on such molecular features.

Here, we focus on epigenetic marks (epialleles). We formalize a notion of neutrality for epialleles, and develop a test for selection which also allows for the assessment of possible confounders. Our test captures the known biology of epialleles, including *trans*-regulation. Because we are leveraging the fact that each epiallele occurs multiple times throughout the genome, the test requires neither population-scale data nor data across species. It only depends on knowing genic selection coefficients against heterozygous loss-of-function alleles, which are now known for most genes in human. We apply our test to epigenome-wide data from the human male germline and find: (A) Proximal promoter methylation and gene-body H3K36me3 are under negative and positive - respectively - selection, and this is not entirely explained by their association with gene expression. (B) The size of the hypomethylated region at promoters is under positive selection, while having almost no association with expression. (C) DNA methylation at the transcriptional end site is not under selection.

Both DNA methylation and H3K36me3 affect local mutation rates in the germline; DNA methylation elevates CpG mutation rates, while H3K36me3 lowers exonic mutation rates by recruiting the mismatch repair machinery. We provide evidence suggesting that the residual selective pressure on these marks - after accounting for their relationship with gene expression - may partly be explained by their action as mutation rate modifiers. We show via simulations that such selection can overcome genetic drift because these features affect multiple genes.

Our framework is simple but general, and we anticipate its core idea to be useful for other molecular features.

174V Coevolution is pervasive between unrelated glycosylation pathways and points to potential disease modifiers *Holly Thorpe*, Nathan Clark, Clement Chow University of Utah Department of Human Genetics, Salt Lake City, Utah

Glycosylation is one of the most common post-translational modifications. It is necessary for protein localization, folding, and stability. Defects in glycan biogenesis pathways, such as N-linked glycosylation, O-linked glycosylation, and GPI anchor synthesis, can result in protein folding defects often causing ER stress. These defects lead to rare, multi-systemic disorders classified as Congenital Disorders of Glycosylation (CDG). CDGs typically present with seizures, hypotonia, and developmental delay, but display large clinical variability with symptoms affecting every system in the body. This variability suggests modifier genes affect the phenotypes. I am employing evolutionary approaches to identify modifier genes of CDGs.
Evolutionary Rate Covariation (ERC) relies on the premise that proteins that interact physically or genetically or are functionally related coevolve at similar rates. ERC values are calculated using the correlation coefficient of evolutionary rates of gene pairs in a species tree. I used ERC values to look genome wide for coevolution with CDG genes, specifically genes involved in GPI anchor synthesis.
There was enriched coevolution among GPI anchor synthesis proteins. Unexpectedly, there was also enriched coevolution between GPI anchor synthesis proteins and proteins in other glycosylation pathways, suggesting more overlap between the different pathways than appreciated. Gene Ontology analysis of top genes that coevolve with GPI anchor synthesis proteins showed enrichment in genes involved in RNA modification and mitochondrial gene expression, suggesting interactions between these processes and GPI anchor synthesis. Gene pairs with the highest coevolutionary scores included both *HTT* and *PIGG* and *ATG7* and *PIGG*. *HTT* and *ATG7* are associated with neurodegenerative disorders possible indicating overlap in pathophysiology between the disorders.
To functionally validate these exciting signals, I screened for genetic interactions using the *Drosophila* eye. Many GPI anchor synthesis genes are necessary for *Drosophila* eye development and knockdown of these genes leads to rough and disorganized eyes. By creating double knockdowns of GPI anchor synthesis genes and coevolving genes in the *Drosophila* eye, I identified genetic interactions between genes previously thought to be unrelated. Many of the strongest evolutionary signals validate as interactors in this *in vivo* analysis. Coevolution is an underutilized tool for identifying interactions between unrelated proteins. These connections could indicate the glycosylation proteins have evolved to develop other functions unrelated to glycosylation.

175V Mutualism-enhancing mutations dominate early adaptation in a microbial community *Sandeep Venkataram*[1], Huanyu Kuo[1], Erik Hom[2], Sergey Kryazhimskiy[1] 1) University of California, San Diego; 2) University of Mississippi

From phytoplankton producing the planet's oxygen to wildebeest grazing the Serengeti, each species modifies their

ecosystem. These ecological changes can precipitate adaptive evolution, which in turn can lead to further changes in the ecosystem. Previous studies have shown that this coupling between ecological and evolutionary processes is often driven by interactions between species. While there are a number of case-studies of individual demonstrations of eco-evolutionary feedbacks and the role of species interactions in driving these feedbacks, the details of how they work are not well understood. In particular, we do not know how the addition of a species to a community impacts the distribution of adaptations available to other community members, and how these adaptations in turn affect community ecology.

We address this gap in an experimental microbial community consisting of the yeast Saccharomyces cerevisiae and the alga Chlamydomonas reinhardtii, which have been previously shown to form an obligatory mutualism in certain laboratory conditions (Hom & Murray, Science 2014). We modified these conditions to make this mutualism facultative, which allowed us to measure (1) how addition or removal of one community member (algae) changes the adaptive mutations available to another member (yeast) and (2) how these mutations in turn alter the ecology of the community. We sampled hundreds of adaptive mutations in each condition to gain a quantitative understanding of how a species interaction affects community eco-evolution.

We show that yeast adaptation is highly diverse at both the genetic level and in their effect on community ecology, as some adaptations decrease the yield of one or both species while others increase one or both yields. The community thus has the potential to either enhance or break down the mutualism by the action of natural selection. However, there are no fitness trade-offs for yeast between growing alone or in the presence of algae. Despite this lack of trade-off, algae algae systematically alter the distribution of community impacts by favoring adaptations that increase the yield of both species. This bias can be explained by understanding the dynamics of rapid adaptation in large asexual populations. As a result of this bias, evolution becomes more repeatable at the community level compared to the evolution of yeast in isolation.

176V Species-specific chromatin landscape determines how transposable elements shape genome evolution *Yuheng Huang*, Harsh Shukla, Grace Yuh Chwen Lee UC-Irvine

Transposable elements (TEs) are selfish genomic parasites that increase their copy number at the expense of host fitness. The “success”, or genome-wide abundance, of TEs differs widely between species. Deciphering the causes for this large variety of TEs has remained a central question in evolutionary genomics. We previously proposed that species-specific genomic TE abundance could be driven by the inadvertent consequences of host-direct epigenetic silencing of TEs—spreading of repressive epigenetic marks from silenced TEs into adjacent functional sequences. Here, we compared this TE-mediated epigenetic effect in six species in the *Drosophila melanogaster* subgroup to dissect step-by-step the role of such effect in determining genomic TE abundance. We found that TE-mediated spreading of repressive marks is prevalent and significantly varies within and across species. While this TE-mediated effect alters the epigenetic states of adjacent genes, we surprisingly discovered that the transcription of neighboring genes could reciprocally impact this spreading. Importantly, our multi-species analysis provides the power and appropriate phylogenetic resolution to connect species-specific host chromatin regulation, TE-mediated epigenetic effects, the strength of natural selection against TEs, and genomic TE abundance unique to individual species. Our findings point towards the importance of the host chromatin landscape in shaping genome evolution through the epigenetic effects of a selfish genetic parasite.

177V Structural variation in the 6.5 Gb genome of the flowering plant *Phlox drummondii* *Danielle Khost*[1], Austin Garner[1], Samridhi Chaturvedi[2], Tim Sackton[1], Robin Hopkins[1] 1) Harvard University, Cambridge, MA; 2) University of California Berkeley, Berkeley, CA

Phlox wildflowers are a formative model of plant ecological and evolutionary research. One species, *P. drummondii,* has become a model for studying patterns of evolutionary divergence and reinforcement during the speciation process. In this system, selection against hybridization with the closely related species *P. cuspidata* has driven the evolution of flower color divergence within *P. drummondii* to increased reproductive isolation. While extensive research has characterized the selective landscape surrounding flower color divergence and its role in speciation, progress towards understanding the mutational basis of this trait and a broader understanding of the genomic patterns surrounding the speciation process are limited by genomic resources in this system. Here we present the contiguous, complete assembly of the repeat-dense 6.5 Gbase *Phlox drummondii* genome. To generate this assembly, we used high coverage Oxford Nanopore sequencing, including a fraction of ultra-long reads for scaffolding. This genome is highly complete, as measured by BUSCO score (>98% complete), and reasonably contiguous, comparing favorably to other long-read assemblies of plants with much smaller genomes. The quality of this assembly gives us the opportunity to characterize the repetitive landscape of the genome, which we estimate to be composed of ~90% repetitive elements, mostly interspersed

elements. Using low coverage, long-read population resequencing, we find evidence for widespread structural variation across the *Phlox* genome. We describe patterns of variation both within and between *Phlox* species, including evidence for structural polymorphisms in regions of interest underlying reproductive isolation between the species.

178V Genome-wide Effects of the Y Chromosome on Gene Expression and Genome Architecture in *Drosophila melanogaster* *Matt Metzloff*, Yassi Hafezi, Iskander Said, Asha Jain, Andrew Clark Cornell University

There is a long history of strong trans- genetic effects caused by the Y chromosome. The Y is proposed to cause these effects by acting as a "heterochromatin sink," competing with other regions of heterochromatin to sequester regulatory factors, like HP1a, that are in limited supply. The *Drosophila melanogaster* Y is composed entirely of constitutive heterochromatin; it contains only 16 protein coding genes and no sex-determination locus. Variation and dosage of the *Drosophila* Y are associated with genome-wide expression changes affecting male fertility, sex-specific aging, temperature adaptation of spermatogenesis, geotaxis, and immune response, but the mechanism is unknown. Brown and Bachtrog (2020) identified hundreds of changes in gene expression and the distribution of repressive histone marks (H3K9me2 and H3K9me3) associated with different numbers of Y chromosomes.

Here we further dissected the trans-effects of the *Drosophila* Y by examining the genomic distribution of Y-dosage-induced changes in gene expression and three-dimensional genome folding. We used RNA-seq to profile gene expression in males with zero, one, and two copies of the Y chromosome. To identify important trans effects that may have been masked by genetic background in previous studies, we used a reciprocal crossing scheme to control for effects caused by autosomal and X-linked differences between strains. We identified relevant expression changes through pairwise comparisons between males sharing X chromosome or autosomal backgrounds and using likelihood-ratio tests in a method similar to time series analysis, to compare all lines at once. We discovered a group of genes, mostly on the X, that are suppressed by the presence of one or more Ys. This group includes *pira*, a SUMO protease targeted by Y-linked piRNAs. Another group of autosomal genes had similar expression patterns to Y-linked genes. We performed High-throughput Chromosome Conformation Capture (Hi-C) to investigate how 3D folding of the genome is affected by Y-dosage. If 3D conformation contributes to trans effects of the Y, we expect to see a correlation between changes in gene expression and 3D interactions. Understanding trans interactions between the Y and the rest of the genome may clarify how heterochromatin regulates gene expression, with implications for the dynamics of Y chromosome evolution, the evolution of sexual dimorphism and sexual conflict, and the disease-related consequences of genome dysregulation.

179V Cytonuclear stoichiometry in the wake of genome duplication Raymond Castillo[1], Matheus Fernandes Gyorfy[2], Justin Conover[3], Corrinne Grover[3], Emma Miller[3], Jonathan Wendel[3], Daniel Sloan[2], *Joel Sharbrough*[1] 1) New Mexico Institute of Mining and Technology, Socorro, NM; 2) Colorado State University, Fort Collins, CO; 3) Iowa State University, Ames, IA

The plant genome is partitioned across three distinct cellular compartments: the nucleus, mitochondria, and chloroplasts. Interactions between nuclear-encoded gene products and those of cytoplasmic genomes (*i.e.*, cytonuclear interactions) underlie the essential cellular processes such as respiration and photosynthesis. Whole genome duplication events (WGDs) are a prominent process of diversification in eukaryotes and are expected to perturb cytonuclear interactions in two fundamental ways: altering the genetic stoichiometry of cytonuclear interactions and increasing cell size. Organelle size, organelle genome copy numbers, cytonuclear transcriptomic and proteomic stoichiometry, and ultimately the efficiency of carbon fixation and ATP production might all be altered as a consequence of WGD, but many decades of careful investigation into polyploidy have yet to fully evaluate these predictions. We investigated the relationship between nuclear and organelle genome copy numbers in common-garden-reared diploid and polyploid accessions of both wheat and *Arabidopsis*. Our droplet digital PCR (ddPCR) estimates of nuclear, mitochondrial, and chloroplast genome copy numbers revealed evidence of substantial intra-specific and intra-individual variation for organelle genome copy number, as well as evidence that polyploids exhibit elevated organelle genome copy numbers per cell. To characterize the genomic architecture of cytonuclear stoichiometry in *Arabidopsis,* we used the Arabidopsis 1001 genomes project data to identify genome-wide SNPs associated with variation in mitochondria- and plastid-nuclear ratios. Taken together, our results indicate that polyploids appear to compensate for increased nuclear genome content with increased organelle genome copies in both monocots and dicots, indicating that cytonuclear stoichiometry is an important component of successful interactions between nuclear and cytoplasmic genomes.

180V TMv01, an active mobile element in the genome of the fern, *Marsilea vestita* *Sruthi Srinivasan*[1], Eliana Herman[1], Fay-Wei Li[2], Nasim Rahmatpour[2], Steve Mount[1] 1) Department of Cell Biology and Molecular Genetics, University of Maryland, College Park; 2) Boyce Thompson Institute and Cornell University, Ithaca, New York

We have recently determined the genome sequence of the semi-aquatic, heterosporous fern *Marsilea vestista*. The estimated genome size is 1.27 Gb. and the assembly of 1.03 Gb. has an N50 of 2.97 Mb. Repeatmasker identifies 422 Mb. (41%) of the sequence as repetitive. Because the genomes of ferns have not been extensively analyzed, we have begun an analysis of the structural diversity of repetitive elements that characterize this genome.

The extremely abundant, and apparently active, retroelement TMv01 was selected for detailed analysis. TMv01 is a member of the gypsy LTR superfamily. There are over 2,000 copies of TMv01 in the genome, including 106 copies with identical LTRs (indicating recent transposition). The complete element has a length of 7,500-8,100 base pairs and an open reading frame of 1,485 amino acids with the expected gag, protease, reverse transcriptase, RNaseH, and integrase domains. An additional chromo domain at the C-terminus, although unusual, has been seen in angiosperm elements of the gypsy subfamily. LTRs are 455-458 nucleotides in length. Insertion generates 5 base pair target site duplications. An unusual feature of this element is the presence of a long noncoding region of over 2,900 nucleotides between the 5' LTR and the open reading frame.

181V The effect of crossbreeding on the transcriptome profiling of indigenous cattle populations: a case study *Mohammad Hossein Banabazi*[1,2], Saeed Esmaeilkhanian[1], Ayeh Sadat Sadr[3], Mohammad Reza Attari[4], Hamid Reza Seyedababdi[1], Nematollah Asadi[1], Ikhide G. Imumorin[5] 1) Department of Biotechnology, Animal Science Research Institute of Iran (ASRI), Agricultural Research, Education & Extension Organization (AREEO), Karaj 3146618361, Iran; 2) Department of animal breeding and genetics (HGEN), Centre for Veterinary Medicine and Animal Science (VHC), Swedish University of Agricultural Sciences (SLU), Uppsala 75007, Sweden. ; 3) South of Iran Aquaculture Research Institute, Iranian Fisheries Science Research Institute (IFSRI), Agricultural Research Education and Extension Organization (AREEO), Ahvaz, Iran; 4) Department of Animal Science, Faculty of Agriculture, University of Tehran. Karaj, Iran; 5) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

Crossbreeding is a classic inbreeding strategy to combine the genetic potential of two or more breeds that contribute. It applies in different methods. The crossbred animals slaughter in a terminal-cross scheme. Crossbreeding may continue to grade a breed up. Alternatively, it used to create a new breed after a few generations. But, these methods have not applied to local breeds in the right way. So, crossing with the imported breeds threatens the conservation of native animal diversity. Since the F1 bulls have some benefits of heterosis, most animal breeders would like to disseminate rapidly in the total local populations. But this heterosis will not necessarily be inherited by the next generations. The present research was to study transcriptome profiling and differential gene expression (DGE) among Sistani cattle and its crosses with three imported *Bos taurus* breeds. The Sistani cattle is an Iranian indigenous cattle population (*Bos indicus*) that has crossed with the invasive breeds, including Holstein, Simmental, and Montbéliarde. Whole blood samples were collected from animals reared under the rural production systems and Sistan's climate conditions in the east of Iran. Then, whole transcriptomes were routinely analyzed through an RNA-Seq workflow with the downstream analysis. The DGE analysis was showed that the indigenous breeds are intensively influenced in the transcriptome level by crossbreeding. The gene enrichment revealed that the genes involved in the conservative processes, particularly the major histocompatibility complex (MHC) gene family, were suffering the most changes in transcriptional coverage. It means that the local cattle population may probably lose their compatibility potential. These changes gradually are subsidence and remain. Holstein had fewer differentially up and down-regulated genes with mild changes than two other emerging breeds in that region, Simmental and Montbéliarde. It means that the history of breed importation may be reflected accordingly in the transcriptome profile of their crosses with the local breeds. In conclusion, the out of control crossbreeding of the local populations with the imported invasive breeds should consider as an urgent threat to conserving the indigenous animal populations in the short term and a continuous risk in the long term.

182V Detecting poaching hotspots, trade centers and sex-biased killing from tiger seizures: Implication in effective wildlife conservation *SUDHANSHU MISHRA*[1,2], SUJEET SINGH[2], PUNEET PANDEY[2], SURENDRA PRAKASH GOYAL[2] 1) UTTARANCHAL UNIVERSITY, DEHRADUN, INDIA; 2) WILDLIFE INSTITUTE OF INDIA, DEHRADUN, INDIA

Poaching of the tiger for their body parts is burgeoning due to their demand for global trade, which has become a major challenge for tiger conservation. Increased poaching of Bengal tiger (*Panthera tigris tigris*) has significantly contributed to local extinction in Panna and Sariska Tiger Reserves in India. The most effective way to contain this trade/poaching is to detect poaching hotspots by examining the parts and products seized under wildlife offenses. This enables authorities to execute law enforcement to tiger poaching hotspots and potentially prevents trade before the wildlife is actually killed. Thus, we discuss combined genetic and statistical approaches to detect poaching hotspots of Bengal tigers in India based on unique mitochondrial DNA haplotypes and multi-locus microsatellite genotyping and identify selective sex poaching. Therefore, in the first step, we examined tiger blood (n=20), and scat (n= 263) samples collected from 11 tiger popula-

tions using 9 microsatellite loci to generate a uniform genetic database of tigers from these populations. In the second step, we undertook 48 tiger samples which were seized in different states (Northern, Central, Southern, and North-Eastern) of India by different enforcement agencies. Based on unique mtDNA haplotypes and multilocus genotyping, tiger poaching cases (n=48) were assigned to northern (20%) and central India (51%) tiger populations. Thus, more conservation efforts are needed in central India tiger populations to prevent poaching. Male tigers (71%) were found more prone to poaching as compared to female tigers (29%) among all case samples analyzed so far. We have also identified central and north Indian states as major hubs for tiger trade. This is the first study on detecting poaching hotspots for Bengal tigers in India and the method we used can be applied to trace geographic origins of other tiger sub-species from confiscated tiger parts which would facilitate curtailing global trafficking in tiger trade. This method can be applied to any wild animal species for their better conservation efforts.

183V Heat adaptation in cross-kingdom pathogenic fungus *Fusarium oxysporum* *Dilay Hazal Ayhan*[1], Kaito Hioki[1], Domingo Martínez-Soto[1], Serena Abbondante[2], Michaela Ellen Marshall[2], Cristina López Díaz[3], Neta Shlezinger[4], Eric Pearlman[2], Antonio Di Pietro[3], Li-Jun Ma[1] 1) University of Massachusetts Amherst, MA; 2) University of California Irvine, CA; 3) Universidad de Córdoba, Spain; 4) The Hebrew University of Jerusalem, Israel

F. oxysporum is a cross-kingdom pathogenic fungus that can cause vascular wilt disease in many plant species and can infect animals and cause local or disseminated fusariosis in humans. To be able to infect humans, fungal pathogens must overcome some biotic and abiotic stresses, such as adapting to elevated body temperature.
Focusing on temperature adaptation, we conducted comparative evolution experiments using a plant pathogenic isolate (*F. oxysporum* f. sp. *lycopersici* Fol4287) and a keratitis strain isolated from the cornea of a patient (MRL8996). Both strains were passaged 10 times through media plates with minimal or rich nutrients, at 28°C or 34°C, with 5 independent replicates.
Belonging to the same species complex, Fol4287 and MRL8996 share a core genome with an average 98% nucleotide identity. However, each genome has its own distinct accessory chromosomes (ACs) with different gene ontology enrichments and transposable element (TE) contents. At the start of the short-term experimental evolution study, the human pathogenic strain MRL8996 exhibited higher fitness at elevated temperatures, while the plant pathogen Fol4287 had more tolerance to the osmatic and cell wall stress conditions.
After 10 passages, we observed the most significant phenotypic difference among Fol4287 populations evolved under elevated temperature, showing significant improvement in heat tolerance when compared to the ancestor. As a trade-off, these populations showed a reduced growth rate in some of the other stress conditions, such as oxidative and osmotic stresses.
Sequencing of the final populations revealed signatures of weak selection in the evolved MRL8996 populations with low-frequency TE insertion events by a hyper-active Foxy element. While TEs were also highly active in Fol4287, different patterns and TE families were involved.
Strikingly, an uncharacterized accessory chromosome gene was mutated by a DNA transposon, Hormin, in 8 out of 10 34°C-passaged Fol4287 populations that also showed increased heat tolerance. The mutation site is located in between a gene block that is upregulated at elevated temperature and a gene block that is up-regulated in plant infection.
Overall, our study demonstrated although similar mechanisms were employed in different *F. oxysporum* strains, the adaptation to elevated temperatures was distinct and ACs played an important role.

184V A multivariate approach to understanding the genetic basis of reproductive resource allocation *Joseph Gunn*, Enoch Ng'oma, Elizabeth Jones, Elizabeth King University of Missouri, Columbia, Missouri

Differential resource allocation to major biological functions, including somatic maintenance, storage, and reproduction, directly impacts fitness by dictating individual survival and fecundity. Theory suggests that unpredictable resource availability should favor increased allocation to survival and decreased allocation to storage and reproduction under periods of calorie restriction. Conversely, lower allocation to survival and greater allocation to storage and reproduction should be favored under high-quality dietary regimes. Despite the occurrence of these patterns in some species, responses in other taxa are variable, and we lack a clear understanding of genetic, transcriptomic, and phenotypic changes associated with this plasticity. We used Evolve-and-Resequence (E&R) to assess coordinated changes over time underlying differentiation in resource allocation toward reproduction under three diet regimes in the model organism *Drosophila melanogaster*. Starting with a multiparent base population derived from the Drosophila Synthetic Population Resource (DSPR), we set up three selection regimes: 1) constant high calorie availability, 2) deteriorating availability, and 3) fluctuating availability. At generation 0, 5, 10, 20, and 30 we performed genomic sequencing on pooled samples of females and measured haplotype frequencies, lifespan, and fecundity across this time series. We discuss preliminary results on genomic mechanisms for plasticity in resource allocation patterns in this population. Our study advances our understanding of the

evolutionary mechanisms underlying trade-offs in resource allocation.

185V The response to selection across hundreds of traits in a century-long barley experiment *Jill Marzolino*[1], Jacob Landis[1,2], Astrid Junker[3], Juliane Streit[3], Henning Tschiersch[3], Thomas Altmann[3], Claude Becker[4], Daniel Koenig[1] 1) University of California, Riverside, CA; 2) Cornell University, Ithica, NY; 3) The Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gaterslaben, Germany; 4) Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria

The opportunity to directly observe the process of adaptation is exceptionally rare in multicellular organisms. Here, we examine the tempo and mode of genetic and phenotypic change in a barley evolution experiment, the Composite Cross II (CCII), begun in 1928 and continued throughout the last century. The CCII experiment began with over 10,000 unique recombinant progeny derived from 329 intercrosses between diverse barley varieties. For each cycle, plants were grown with minimal human intervention in Davis, CA, allowed to compete over the growing season, and harvested in bulk to propagate the experiment in the following year. Previous studies of a small number of fitness-correlated traits in early and late generations of the CCII showed significant increases in flowering time, height, and yield. Here, we dramatically increase our understanding of phenotypic evolution in the CCII by describing the evolutionary trajectories of 401 traits throughout the lifecycle of 192 genotypes spanning 58 generations of the experiment. The final dataset includes daily measurements of many traits expected to contribute to plant fitness — including descriptors of plant biomass, architecture, pigmentation, and photosynthetic capacity. Natural selection has played a major role in shaping phenotypic change in the experiment, with dozens of traits showing strong directional shifts. Selection was particularly strong in traits related to plant architecture and pigmentation. Genome wide association studies for 80 traits identified 273,947 loci that controlled variation in the early stages of the experiment. Phenotypic shifts were driven by simultaneous selection across 94,497 loci controlling adaptive phenotypes. Alleles targeted by selection were enriched for color-related traits. Our experiments provide a framework for understanding phenomic shifts in response to selection over the decades of adaptation in a real world environment.

186V Can synergistic pleiotropy explain the low parallelism of temperature mediated gene expression evolution? *Dagny A. Runarsdottir*[1,2], Viola Nolte[1], Christian Schlötterer[1] 1) Institute of population genetics, University of Veterinary Medicine, Vienna, Austria; 2) Vienna graduate school of population genetics, Vienna, Austria

Parallel phenotypic responses to similar environmental stressors have been described not only for different populations, but also between species. Since many of these traits are highly complex, it is not clear to what extent this parallel evolution can be also seen on the molecular level. Here, we address this question by studying the transcriptomic response of polymorphic founder populations from two species, *Drosophila melanogaster* and *D. simulans* to controlled laboratory conditions with fluctuating high (18/28°C) and low (10/20°C) temperature regimes. In both species a strong adaptive response towards temperature in gene expression was observed. 193 genes were differently expressed (DE) between high and low temperature for *D. melanogaster* and 189 for *D. simulans*. Only two (Jaccard index = 0.01) and five (J.I. = 0.03) up- and downregulated genes, respectively, were shared between the species. Additionally, a correlation of the overall transcriptome responses was low (Spearman's *rho* = -0.067). Despite the lack of parallel responses on the level of individual genes, the evolved genes in both species were enriched among midgut and accessory gland specific genes (J.I. = 0.67). Understanding how the experimental thermal adaptation we observe in this study is reflected in natural population is important, especially with temperature being one of the most important environmental stressors. Comparing clinal evolution to experimental thermal evolution of the same species provides a valuable information on this matter. Interestingly, the two species differed in the extent to which the same genes evolved in the laboratory and natural populations. While *D. melanogaster* shared more genes than expected by chance, no shared response was detected for *D. simulans*, which could be explained by the different demographies of the species. Reasoning that pleiotropy may be a major factor determining the adaptive gene expression evolution, we compared the levels of pleiotropy among the significant genes of all four groups. In natural populations, pleiotropy appeared to be higher than in the laboratory, for both species. This might be explained by the more complex environments in the wild (synergistic pleiotropy). We propose that pleiotropy may be an important, but complex, factor contributing to the repeatability of gene expression evolution.

187V Effect of larval crowding on transcriptomic plasticity across populations *Tejashwini Hegde*[1, 2], Viola Nolte[1], Christian Schlötterer[1] 1) Institute of Population Genetics, University of Veterinary Medicine (Vetmeduni), Vienna, Austria; 2) Vienna Graduate School of Population Genetics, Vienna, Austria

Phenotypic plasticity is the phenomenon of genetically identical individuals producing different phenotypes in response to the environment. Plasticity is frequently considered to be beneficial as it helps to survive changes in the environment - particularly in the context of local adaptation. Here, we approach plasticity from a different angle. Rather than asking how plasticity contributes to local adaptation, we study a globally plastic trait and ask to what extent the underlying

mechanisms are shared between populations from three different continents. Density is one such trait and larval density, particularly in holometabolous insects such as *Drosophila,* has been shown to vary with the availability of food and affects many life-history traits, growth rate, body size etc. We used gene expression to study the mechanistic basis of phenotypic plasticity in response to larval crowding in *Drosophila simulans*. As expected, larval density results in highly plastic gene expression pattern in all three populations - more than 1400 genes exhibited significant gene expression differences. 42% of the differentially expressed genes were population specific. Even when we consider functionally similar genes (based on GO category), 78.8% (460) of the differentially expressed population specific genes reflect a functional divergence among local populations. We conclude that plastic traits shared among populations have experienced re-wiring of their mechanistic basis for example by directing one type of stress response towards a different stress. This suggests that the comparison of diverged natural populations may provide an excellent approach to understand how the regulatory network of conserved traits could evolve between species.

188V Genomic signature of sexual reproduction in a Bdelloid rotifer Veronika Laine[1], Timothy Sackton[2], *Matthew Meselson*[2] 1) Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland; 2) Harvard University, Cambridge, MA

Bdelloid rotifers, common freshwater invertebrates of ancient origin and worldwide distribution have long been thought to be entirely asexual, posing a challenge to hypotheses for the evolutionary advantage of sex. That bdelloids do reproduce sexually, although only occasionally, is shown by a study of allele sharing within a mitochondrial clade of the bdelloid species *Macrotrachella quadricornifera*, supporting the view that sexual reproduction is universally essential for long-term evolutionary success in eukaryotes.

189V *In-silico* cross-contamination affects inference of genetic relationships in *Saccharomyces cerevisiae* *Audrey Ward*[1], Eduardo Scopel [2] , Brent Shuman [3] , Michelle Momany [3], Douda Bensasson [1,2,3] 1) University of Georgia; 2) Institute of Bioinformatics, University of Georgia, Athens, GA ; 3) Department of Plant Biology, University of Georgia, Athens, GA

Population genetic analysis depends on the quality of whole genome sequences. Contamination of sequence data may occur *in vitro* prior to sequencing or *in silico* during multiplex sequencing as a result of cross-contamination or barcoding issues. Testing for interspecies contamination is common practice. In contrast, identification and prevention of within-species contamination is more difficult. To test the effects of contamination on genome analyses, we contaminated short read genome data of *Saccharomyces cerevisiae in silico* with genome data from another *S. cerevisiae* strain. We repeated the contamination experiment using strains with varying degrees of relatedness and ploidies, in addition to varying level of cross-contamination along a range from 0 to 50%. Using a standard base calling pipeline, we found that cross contaminated genomes appeared to produce good quality genome-wide data. Past studies estimated relationships among *S. cerevisiae* lineages using maximum likelihood trees inferred from whole-genome data after excluding strains showing recent admixture. We similarly estimated trees that include single simulated cross-contaminated genomes to assess if within-species contamination affects the inference of their genetic relationships. We found that between 5 and 10% contamination is enough to significantly change tree topologies, making contaminated strains look like hybrids in maximum likelihood trees. These results suggest that even low levels of contamination significantly change trees and may lead to misunderstanding of evolutionary relationships within species.

190V Genetic Architecture and Temporal Analysis of Developmental Delay in Intra-species *Caenorhabditis briggsae* Hybrids Jordan Montgomery, Joel Rodriguez, Leonardo Velazco-Cruz, Marisol Lauri, Morgan Montelongo, *Joseph Ross* Cal. State Univ., Fresno

Identifying the alleles that reduce hybrid fitness is a major goal in the study of speciation genetics. It is rare to identify systems in which hybrid incompatibilities with minor phenotypic effects are segregating in genetically diverse populations of the same biological species. Such traits do not themselves cause reproductive isolation, but they might initiate the process. In the nematode *Caenorhabditis briggsae*, a small percent of F2 generation hybrids between two wild isolates suffer from developmental delay, in which adulthood is reached after 33% more time than their wild-type siblings. Prior efforts to identify the genetic basis for this hybrid incompatibility found association with one autosome. Here, we have used F2 hybrids and near-isogenic lines to map one locus to a roughly 1 Mbp region of chromosome III. This architecture agrees with the suggestion that developmental delay is caused by a maternal-effect toxin-antidote element. In order to more specifically define the developmental delay phenotype, we monitored the development rate of F2 hybrids and discovered that delay is not restricted to a particular larval developmental stage, unlike many known developmental pathway mutations. Additional analyses with cytoplasmic-nuclear hybrids suggests that a mitochondrial-nuclear epistatic interaction might also contribute to hybrid developmental delay. Our mapping and refinement of the delay phenotype motivates future efforts to study the genetic architecture of hybrid dysfunction between populations of *C. briggsae*.

191V Genome-wide recombination rate plasticity in response to heat stress in *Drosophila pseudoobscura* *Laurie Stevison*, Ulku Huma Altindag, Madison Watkins, Natalia Rivera-Rincon Auburn University, Auburn, AL

A central goal of evolutionary genetics research is understanding the influence of biotic and abiotic processes on genetic variation in natural populations. While previous work has found extensive variation in the degree of recombination rate plasticity between chromosomal regions, a genome-wide survey of this phenomenon is lacking. Here, we have used a classic genetic crossing scheme to conduct a large-scale backcross between wild type strains with multiple replicates reared at control and heat stress temperatures (21°C and 26°C, respectively). We then isolated DNA and performed low coverage (~0.8x) whole-genome sequencing of a pilot set of 38 F_1 female parents and 730 progeny using plexWell384 kits. We used data from the F_1 parents to validate our predicted set of ancestry informative markers. We then used the software *ancestryinfer* to map reads and identify crossover locations between the two parental genomes in each of the progeny. These results guided the construction of genetic maps in each experimental condition for a fine-scale comparison of recombination rates. Our goal is to identify regions of the chromosome that are more sensitive to these short-term changes in recombination due to heat stress. While the majority of plasticity studies show an increase in recombination in response to stress, this result is not consistent. Therefore, these data will provide insight into how chromosomal features might correspond to specific directional changes in recombination. In concert, we have conducted a thorough re-analysis of 24 published recombination maps of the target species, *Drosophila pseudoobscura*, and close relatives to identify regions of the chromosome that are more sensitive to long-term changes in recombination rate. By comparing these short-term and long-term chromosomal regions that are sensitive to recombination rate changes, our work will determine the role of the environment in driving recombination differences within- and between-species. Moreover, because recombination rates often correlate with population genetic signatures of diversity and divergence, our comprehensive portrait of genome-wide recombination plasticity will inform population genetic studies aimed at explaining the heterogenous genome landscape.

192V Pervasive Under-Dominance in Gene Expression as Unifying Principle of Biomass Hybrid Vigor in *Arabidopsis thaliana* *Wei Yuan*, Fiona Beitel, Thanvi Srikant, Ilja Bezrukov, Detlef Weigel Max Planck Institute for Biology

Heterosis, the generally superior performance in hybrids compared to their inbred parents, is one of the most enigmatic biological phenomena. The many different explanations that have been put forward for heterosis beg the question whether common principles underpinning it do exist at all. We performed a systematic transcriptomic study in *Arabidopsis thaliana* involving 150 random crosses, to search for the general principles, if any, that heterotic hybrids share. Consistent additive expression in F1 hybrids was observed for only about 300 genes enriched for roles in stress response and cell death. Regulatory rare-allele burden affects the expression level of these genes, but does not correlate with heterosis. Non-additive gene expression in F1 hybrids is much more common, with the vast majority of genes (over 90%) being expressed below parental average. These include genes that are quantitatively correlated with biomass accumulation in both parents and F1 hybrids, as well as genes strongly associated with heterosis. Unlike in the additive genes, regulatory rare allele burden in this non-additive gene set is strongly correlated with growth heterosis, even though it does not covary with the expression level of these genes. Together, our study suggests that while additive complementation is an intrinsic property of F1 hybrids, the major driver of growth in hybrids derives from the quantitative nature of non-additive gene expression, especially under-dominance and thus lower expression in hybrids than predicted from the parents.

193V Early stages of butterfly speciation are associated with widespread gene expression divergence in sensory tissues Ningning Wu[1], Steven Van Belleghem[2], Brian Counterman[3], Riccardo Papa[2,4], *Wei Zhang*[1,5] 1) State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China; 2) Department of Biology, University of Puerto Rico, Rio Piedras, San Juan 00925, Puerto Rico, United States of America; 3) Department of Biological Sciences, Auburn University, Auburn, AL 36849, United States of America; 4) Molecular Sciences and Research Center, University of Puerto Rico, San Juan 00907, Puerto Rico, United States of America; 5) Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

Neotropical *Heliconius* butterflies are well-known for their intricate behaviors and multiple examples of incipient speciation. Chemosensing plays a fundamental role in the life history of these groups of butterflies and in the establishment of reproductive isolation. However, chemical communication involves synergistic sensory and accessory functions, and it remains challenging to investigate the molecular underpinnings that result in behavioral differences. Here, we examine the gene expression profiles and genomic divergence of three sensory tissues (antennae, legs, and mouth parts), across sexes (females and males) and life stages (mated and unmated females) in two hybridizing butterflies, *Heliconius melpomene* and *Heliconius cydno*. By integrating comparative transcriptomic and population ge-

nomic approaches, we found evidence of widespread gene expression divergence that supports a crucial role of sensory tissues in the establishment of species barriers. Our work strongly supports the unique chemosensory function of antennae, the importance of the Z chromosome in interspecific divergence, and the nonnegligible role of non-chemosensory genes in the divergence of chemosensory tissues. Collectively, our study contributes to a better understanding of the genetic basis of chemosensory mechanisms promoting reproductive isolation and speciation.

196W Revealing the dynamics of sunflower domestication with archaeological DNA Nathan Wales[1,2], Melis Akman[1,3], Peter Stokes[1], Ray Watson[1,4], Greg Owens[1,5], Bruce Smith[6], Kristen Gremillion[7], M. Thomas Gilbert[8], *Benjamin Blackman*[1] 1) University of California, Berkeley; 2) University of York; 3) California State University, East Bay; 4) University of Virginia; 5) University of Victoria; 6) Smithsonian Institution; 7) Ohio State University; 8) University of Copenhagen

Where and how many times crop plants were domesticated, how strong and how quickly genetic diversity is lost during domestication, and how domestication syndromes are assembled through polygenetic evolutionary change are questions of active research and frequent debate. We have been addressing these questions with archaeological DNA approaches in the common sunflower, *Helianthus annuus*. Native American farmers living ~4000-5000 years ago transformed the common sunflower from a highly branched wild plant with small disks and small seeds into a staple oilseed crop that sports a single large head with large seeds on an unbranched stalk. We have assembled a time series of archaeological samples that spans the majority of this period, and we are using endogenous DNA sequences obtained from these samples to reveal how human cultivation altered genetic diversity through time. My talk will focus on how the genomic libraries obtained from these samples and from ethnographic collections from the historic period are proving fruitful for examining hypotheses about where in North America sunflower was domesticated and for highlighting reductions in sequence diversity at multiple time points in the history of sunflower cultivation. The intriguing patterns of haplotype turnover we observe through time also suggest that shifts in agricultural practices have occurred over this period. In addition, my talk will discuss how we have defined a set of candidate domestication genes through population genomics and transcriptomics approaches with extant germplasm, and how targeted resequencing of these loci from our archaeological times series has revealed the timing and order of selective sweeps and thus insight into how the sunflower domestication syndrome was assembled by Native American farmers through time.

197W Genomic evidence for ancient migration routes along South America's Atlantic coast *Andre Luiz Campelo dos Santos*[1,2], Amanda Owings[3], Henry Socrates Lavalle Sullasi[2], Omer Gokcumen[4], Michael DeGiorgio[1], John Lindo[3] 1) Florida Atlantic University, Boca Raton, FL; 2) Universidade Federal de Pernambuco, Recife, Brazil; 3) Emory University, Atlanta, GA; 4) State University of New York at Buffalo, Buffalo, NY

An increasing body of archaeological and genomic evidence have indicated a complex settlement process of the Americas. Starting from Beringia some 20,000 years before present, ancestral Native Americans (NAs) explored and settled northern North America, where they later diverged into two basal genomic branches: Northern NA (NNA) and Southern NA (SNA). The SNA rapidly dispersed towards South America through the Pacific coast, giving rise to present-day Central and South NAs. During this process, an unsampled population also introduced an Australasian ancestry that can only be found in NAs from western Amazonia and in one archaeological individual from Lagoa Santa (Southeastern Brazil). The easternmost portion of South America, however, remains largely unexplored by archaeogenomic studies. Here we show that newly sequenced archaeological individuals from Northeastern Brazil and Uruguay share strong genomic relationships with Lagoa Santa and ancient Panama. We also found that the Australasian signal and an unexpected high genomic affinity with present-day Onge are representative in ancient individuals along South America's Atlantic coast. Our results provide genomic evidence for ancient migrations along South America's Atlantic coast. Our work unravels the deep demographic history of eastern South America, and it should be a starting point for further fine-scale investigations in the region.

198T Genetic insights into the social organization of 13 Siberian Neanderthals *Laurits Skov*[1,2], Stéphane Peyrégne[2], Divyaratan Popli[2], Leonardo Iasi[2], Thibaut Devièse[8,13], Viviane Slon[2,10,11], Elena Zavala[2], Mateja Hajdinjak[2], Arev Sümer[2], Alba Mesa[2], Daniel Comeskey[8,14], Anatoly Derevianko[4], Aliona Kharevich[4], Sergey Markin[4], Sahra Talamo[15,16], Katerina Douka[9,12,17], Maciej Krajcarz[3], Richard Roberts[6,7], Thomas Higham[8,9,17], Bence Viola[5], Andrey Krivoshapkin[4], Kseniya Kolobova[4], Janet Kelso[2], Matthias Meyer[2], Svante Pääbo[2], Benjamin Peter[2] 1) Dept of Molecular & Cell Biology; 2) Department of Evolutionary Genetics, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany; 3) Institute of Geological Sciences, Polish Academy of Sciences, Warszawa, Poland; 4) Institute of Archaeology and Ethnography, Russian Academy of Sciences, Novosibirsk, Russia; 5) Department of Anthropology, University of Toronto, Toronto, Canada.; 6) Centre for Archaeological Science, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, New South Wales, Australia; 7) Australian Research Council (ARC) Centre of Excellence for Australian Biodiversity and Heritage,

University of Wollongong, Wollongong, New South Wales, Australia; 8) Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford, UK; 9) Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Vienna 1030 Austria; 10) Department of Anatomy and Anthropology and Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, 6997801 Tel Aviv, Israel; 11) The Dan David Center for Human Evolution and Biohistory Research, Sackler Faculty of Medicine, Tel Aviv University, 6997801 Tel Aviv, Israel; 12) Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany; 13) European Centre for Research and Education in Environmental Geosciences (CEREGE), Aix-Marseille University, CNRS, IRD, INRAE, Collège de France, Aix-en-Provence, France; 14) Syft Technologies Ltd, 3 Craft Place, Middleton, PO Box 28 149, Christchurch 8242, New Zealand; 15) Department of Chemistry G. Ciamician, Alma Mater Studiorum, University of Bologna Via Selmi 2, Bologna, 40126, Italy; 16) Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103, Germany; 17) Human Evolution and Archaeological Sciences (HEAS), University of Vienna, A-1030, Vienna, Austria

Genomic analyses of Neanderthals across parts of their geographic range have previously provided insights into their population history and relationship to modern humans, but the social organization of Neanderthal communities remains poorly understood. Here, we present genetic data for 13 Neanderthals from two Middle Palaeolithic sites in the Altai Mountains of southern Siberia: 11 from Chagyrskaya Cave and two from Okladnikov Cave - making this the largest genetic study of a Neanderthal population to date. We used hybridization capture to obtain genome-wide nuclear sequence data, as well as mitochondrial and Y chromosome sequences. All Chagyrskaya individuals are very closely related, including a father-daughter pair and a pair of second-degree relatives, indicating that at least some of the individuals lived at the same time. Up to a third of these individuals' genomes occur in long segments of homozygosity, suggesting that the Chagyrskaya Neanderthals were part of a small community. In addition, the Y chromosome diversity is an order of magnitude lower than the mitochondrial diversity, a pattern that we find is best explained by female migration between communities. These genetic data illuminate the social organization of a late Neanderthal population at the easternmost extent of their known range.

199T Hominin and faunal turnovers identified at Denisova Cave with sediment DNA *Elena Zavala*[1], Zenobia Jacobs[2,3], Benjamin Vernot[1], Michael Shunkov[4], Maxim Kozlikin[4], Anatoly Derevianko[4], Elena Essel[1], Cesare de Fillipo[1], Sarah Nagel[1], Julia Richter[1], Frédéric Romagné[1], Anna Schmidt[1], Bo Li[2,3], Kieran O'Gorman[2], Viviane Slon[1], Janet Kelso[1], Svante Pääbo[1], Richard Roberts[2,3], Matthias Meyer[1] 1) Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; 2) Centre for Archaeological Science, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, New South Wales, Australia; 3) Australian Research Council (ARC) Centre of Excellence for Australian Biodiversity and Heritage, University of Wollongong, Wollongong, New South Wales, Australia; 4) Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Novosibirsk, Russia

The integration of ancient DNA with zooarchaeology and paleoecology allows us to contextualize human evolutionary history within the environments that they lived in. However, these studies are currently limited by the rare discovery of hominin remains and the time and resources required for the classification of faunal remains. The discovery that DNA from Pleistocene mammals, including hominins, can be retrieved from sediments has opened up the possibility of examining the relationship between faunal composition and hominin occupation at archaeological sites. We explore this possibility at Denisova Cave, a site in the Altai Mountains, which is thought to be a contact zone for different faunal and hominin groups. This site is of particular interest for hominin evolutionary history as not only Neandertal, but also Denisovan remains have been identified there. In addition, debates remain as to who were the makers of jewelry and other artefacts found in the Initial Upper Paleolithic layers as no early modern human remains have been found at the site. We tested 728 samples from the cave's Pleistocene layers using a fully automated workflow for DNA extraction, library preparation and hybridization capture, for the presence of both ancient faunal and hominin mitochondrial (mt)DNA. Ancient mammalian mtDNA was identified in 685 (94%) samples and ancient hominin mtDNA was identified in 175 (24%) samples. This rich data set enable us to identify changes in the relative proportions of DNA from various mammalian families and a turnover between the types of bears and mtDNA haplogroups of hyaenas at the site. In addition, we found shifts in the presence of mtDNA from different hominin groups, which appeared to coincide with past climatic changes. This study demonstrates the potential of using sediment DNA for increasing our understanding of past faunal diversity and hominin occupations at archaeological sites.

200W The Genomics of Highly Variable Physiological Response to Temperature *Amand DeLiberto*, Melissa Drown, Marjorie Oleksiak, Douglas Crawford Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL

Biotic and abiotic factors determine organismal survival and success in their habitat. Temperature is one of the most important abiotic factors defining metabolism and species range. It is nearly universally true that ectotherms from warmer environments have evolved lower metabolic rates than those from colder environments. This broad phylogenetic convergence and common physiological and biochemical mechanisms underlying temperature adaptation suggest strong selection that should reduce interindividual variation. Yet, there are few studies characterizing variation in physiological traits, or defining the nucleotide variation driving it. To investigate variation in temperature response, six traits were measured in 250 *Fundulus heteroclitus* individuals collected across a non-clinal thermal mosaic of habitats, with up to an 8°C difference. Traits included whole-animal metabolic rate (MO_2) and Critical Thermal Maximum (CTmax) measured in the same individuals at both 12°C and 28°C, and 4 substrate-specific cardiac metabolic rates measured at either 12°C or 28°C. Here we focus on the thermal sensitivity (defined as the Q_{10}; change in trait value for every 10°C increase in temperature) of MO_2 and CTmax. Among and within populations we observed high phenotypic variation in these traits. Variation in thermal sensitivity ranged from a Q_{10} ~1 (same trait value at 12°C and 28°C) to a Q_{10}>3 (~6-fold difference between temperatures). Much of this variation is within a population, yet up to 10% is associated with habitat temperature. To further understand the genomic basis of variation in these traits, we developed a sequencing technique, Extended Exome Capture sequencing (EXECseq) that targets regions of the genome likely under selection. EXECseq uses expressed RNA probes to capture large genic fragments, including exons and cis-regulatory regions. RNA probes were generated using numerous tissues exposed to heat and oxygen stress to produce a large pool of expressed genes. This method was used to sequence all individuals at 10X coverage across over 10,000 genes at a lower cost than whole-genome sequencing, and greater genome coverage than other reduced representation methods, and >80% of reads fell within genic regions. Using a multivariate analysis of this genomic data and phenotypic data, we can understand how individuals within and among these populations respond to temperature and provide insights into the genetic basis of the high variation in thermal sensitivity.

201W mRNA expression explains metabolic and thermal tolerance trait variation *Melissa Drown*, Marjorie Oleksiak, Douglas Crawford University of Miami, Miami, FL

To understand the molecular mechanisms underlying complex traits in wild populations, we measured metabolic and thermal tolerance traits and related them to mRNA expression. Traits included whole animal metabolic rate (WAM), substrate specific cardiac metabolic rate (CaM; substrates= glucose, fatty acids, lactat+ketones+ethanol, and endogenous), and critical thermal maximum (CT_{max}) among three *Fundulus heteroclitus* populations acclimated to 12°C and 28°C. Populations showed few significant interpopulation differences in these traits due to large inter-individual variation. In contrast, acclimation temperature significantly affected WAM and CT_{max} but not CaM. Within each acclimation temperature we examined which mRNAs were related to each trait using a weighted gene co-expression network analysis. These analyses revealed 9 significant heart ME (first principal component of module expression) and 4 significant brain ME. Heart MEs explain variation in whole animal metabolism (WAM), critical thermal maximum (CT_{max}), and two of the four substrate specific cardiac metabolic rates at 12°C. The only heart trait related to mRNA expression at 28°C was CT_{max}. These patterns at 12°C and 28°C were related to the higher inter-individual variation at 12°C. Brain MEs explains CTmax and WAM at 28°C but not at 12°C. Combining MEs as multiple correlations, 82% of variation in WAM at 12°C was explained by four heart MEs, 80% of variation in FA CaM at 12°C was explained by three heart MEs, and 72% of variation in CT_{max} at 28°C was explained by three brain MEs. MEs significantly correlated with traits were enriched for Kyoto Encyclopedia of Genes and Genomes (KEGG) terms related to specific metabolic pathways known to impact these traits, suggesting that these MEs represent biologically relevant pathways. Together these data suggest that mRNA co-expression explains complex traits and importantly mRNA expression patterns that explain traits are different for different temperatures and tissues.

202T Characterizing *Mimulus guttatus* adaptation to serpentine soil *Allison Gaudinier*[1,2], Jessica Selby[3], Allison Rothrock[1], John Willis[3], Benjamin Blackman[1] 1) University of California, Berkeley, Berkeley, CA; 2) Miller Institute for Basic Research in Science, Berkeley, CA; 3) Duke University, Durham, NC

Yellow monkeyflowers, *Mimulus guttatus*, are particularly adept at overcoming environmental stresses and can be found growing in a wide variety of ecosystems. We are investigating the molecular and phenotypic mechanisms that allow for *M. guttatus* to adapt to serpentine soils, a particularly harsh edaphic environment. Serpentine soils have skewed Mg:Ca ratios, low levels of essential nutrients (N, P, K, Ca), increased levels of heavy metals such as Zn, Ni, Fe, and Cd that are toxic to many plants, and have poor water retention. Previous work suggests that tolerance to the complex stresses of serpentine soils is through polygenic adaptation. We are using hypothesis-generating approaches to determine the genetic mechanisms and traits selected for serpentine adaptation. Population genome resequencing of many serpentine and nonserpentine populations has revealed regions with divergent alleles. This data, coupled with transcriptome

profiling in serpentine and nonserpentine conditions, has revealed strong candidate genes for serpentine adaptation. Allele specific expression analysis of reciprocal F_1 hybrids will determine if differential expression of candidate genes is due to cis-regulatory difference. Further, protein structure prediction has allowed us to identify candidate genes for which serpentine and nonserpentine alleles are predicted to differ functionally. We have targeted several candidate genes for mutant analysis using CRISPR-Cas9 and will examine their functional roles. Accompanying the genomic data, we are performing phenotyping of potentially adaptive traits to define the trait differences between serpentine and nonserpentine populations. Initial results indicate that tolerance of the highly skewed ionic composition of serpentine soils likely derived in part from evolutionary adjustment of ion channel properties.

203T Untangle the quantitative genetics of self and heterospecific pollen rejection during pollen-pistil interactions *Robin Hopkins*[1], Antonio Serrato-Capuchina[1], Charlie Hale[2], Matthew Farnitano[3], Federico Roda[4] 1) Harvard University; 2) Cornell University; 3) University of Georgia; 4) National University of Colombia

Many plant species have evolved pollen-pistil recognition systems to optimize reproductive success. The two most important pistil recognition systems distinguish between self/non-self pollen and between pollen from same/different species, resulting in self-incompatibility and heterospecific incompatibility, respectively. It has long been hypothesized that these recognition systems are mechanistically linked, but little is known about the extent of their genetic overlap across different types of pollen-pistil interactions or whether this overlap constrains the evolution of pistil recognition. Here we untangle the complex genetic basis of variation in the strength of pollen-pistil incompatibility in the Texas wildflower, *Phlox drummondii*. Natural populations of *P. drummondii* exhibit extensive variation in both the strength of the self-incompatibility and the strength of the heterospecific-incompatibility with the closely related species *P. cuspidata*. Furthermore, there is a significant positive correlation between the strength of self-incompatibility and heterospecific-incompatibility across individuals. We use quantitative trait locus mapping and genome-wide association mapping to identify the genomic regions controlling variation in pollen-pistil incompatibilities. We identify the extent to which the two pollen rejection systems are share genetic mechanisms or co-vary due to parallel patterns of selection and demography. Additionally, we reveal if and how the S-locus, which contains the pollen and pistil genes necessary for self-pollen recognition, is involved in the quantitative variation in the strength of self- and heterospecific-incompatibility within *P. drummondii*. Our work sheds new light on the long-standing questions about the genetic mechanisms plants use to choose their mates. By exploring whether self and heterospecific incompatibilities share genetic mechanisms, we can additionally generate predictions for how the evolution of one incompatibility may affect the other.

204W An agent-based model of signaling in Bacillus subtilis biofilms *Obadiah Mulder*[1], Joseph Larkin[2], Michael Edge[1] 1) University of Southern California, Los Angeles, CA; 2) Boston University, Boston, MA

In a microbial biofilm, cells form into an aggregate, producing community goods and providing a variety of benefits to member cells. Cells in a biofilm must solve a coordination problem, in that key nutrients can become locally scarce depending on the feeding behavior of nearby cells. One way Bacillus subtilis biofilms solve this problem is through electrical signaling that regulates feeding. At the individual level, signaling is costly, but a network of signallers that can traverse the biofilm is important for biofilm survival. In this work, we consider signaling from a social evolution perspective, building on previous findings that the proportion of signallers in biofilms seems to hover near a quantity that can be predicted with percolation theory. Here, we extend previous work by considering the direction of biofilm growth and signaling. Individual B. subtilis cells have a long and a short axis, and the biofilms grow in rows, such that there is high relatedness along one axis and low relatedness on the other. Using theory and simulations, we explore how this structure may increase the efficacy of signaling and the resilience of the system to cheater cells.

205W Refining Polygenic Score History Estimation from Reconstructed Ancestral Recombination Graphs *Dandan Peng*, Obadiah Mulder University of Southern California, Los Angeles, CA

A polygenic score (PGS) is a weighted sum of an individual's genotypes used to predict the individual's value for a complex trait. Estimating the evolutionary history of a population's mean value of one or more polygenic scores may rule out or support hypotheses about trait evolution in the population. One of the key components in estimating a PGS history is to estimate the trait-associated alleles' frequencies. In previous work, Edge & Coop (2019) proposed three methods to estimate the historical time course of a population-mean PGS using estimated local coalescent trees encoded by an estimated tree sequence or ancestral recombination graph (ARG). Among Edge & Coop's three estimators, one performs well under neutrality but is biased under natural selection (proportion-of-lineages). The other two are less biased by selection but very noisy (waiting-time and lineages-remaining). Here we report the performance of Edge & Coop's methods using state-of-the-art ARG estimation methods (Relate and tsinfer), and we explore approaches to improving the estimators, including smoothing approaches for the two noisy methods and an expectation-maximization (EM) approach to correct-

ing bias observed under selection for the other method. The improved methods will contribute to more accurate estimates of PGS histories and allow the testing of more specific hypotheses about trait evolution.

206T Genome associations with soil phosphorus availability in sorghum and maize. *Fausto Rodriguez Zapata*[1], Nirwan Tandukar[2], Jung-Ying Tzeng[3], Rubén Rellán-Álvarez[4] 1) Genetics Program, North Carolina State University, Raleigh, NC; 2) Department of Functional Genomics, North Carolina State University, Raleigh, NC; 3) Department of Statistics, North Carolina State University, Raleigh, NC; 4) Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC

Some sorghum and maize populations are adapted to low phosphorus soils. Traditional farmers grow landraces of both in nutrient scarce soils with little to no supplemental phosphorus. In order to discover genetic determinants of local adaptation to low phosphorus soils we made environmental associations between phosphorus availability and the genomes of corn and sorghum. We used georeferenced genotyped datasets of landraces grown around their area of domestication, Latin America for maize, and Africa for sorghum. With the landrace geographical location, we extracted phosphorus availability and soil classification data. We investigated whether phosphorus associated genes were shared orthologs or differed between the two grasses, indicating different ways of adaptation to phosphorus scarcity.

207T Integrative pathway analysis of metabolites reveal genetic architecture of complex traits and disease *Courtney Smith*, Nasa Sinnott-Armstrong, Jonathan Pritchard Stanford University, Stanford, CA

Understanding how genetic variants influence multiple traits is an essential foundational question of statistical genetics. Genome-wide association studies (GWAS) have begun characterizing the genetic architecture of complex traits in numerous species, but the molecular mechanisms connecting GWAS hits to their traits are often unclear without extensive forward genetic analyses. Investigation of molecular traits in related pathways, with well-documented biology jointly impacting their levels, provides an opportunity to uncover putative molecular mechanisms. Here, we perform genome-wide association studies (GWAS; n = 94,464 European ancestry individuals) of 16 metabolites clustered at the intersection of amino acid catabolism, glycolysis, and ketone metabolism and systematically evaluate their shared genetic bases. Among the 213 independent GWAS variants, we find a strong enrichment for genes encoding pathway relevant enzymes (n = 68 variants; 79-fold, P < 2.2e-16) and transporters (n = 46 variants; 13-fold, P < 2.2e-16). We then investigate the joint effects of pleiotropic variants on biologically-related metabolites in the context of their biochemical pathways. This multivariate approach allows for a better understanding of why these variants are associated with and influence the level of these metabolites. We find that variants with effect direction in metabolite pairs opposite the overall genetic correlation of the metabolites, which we define as discordant variants, are more likely to affect enzymes and transporters than other gene types (4.1-fold, P = 0.034). We also find that discordant variants affecting pathway relevant enzymes are more likely to act between, rather than upstream or downstream of, the metabolites for which they are discordant (23-fold, P = 0.0072). Finally, we apply this method to complex trait GWAS hits and identify a coronary artery disease association at PCCB with striking interpretability of effects on disease-relevant pathway metabolites, underscoring the potential of unifying biochemistry with dense metabolomics data to understand the molecular basis of complex traits and diseases.

208W The influence of demographic history and genetic architecture on complex phenotypes via runs of homozygosity Zachary Szpiech Pennsylvania State University

Runs of homozygosity are long stretches of identical-by-descent (IBD) haplotypes inherited from parents with a recent common ancestor. Their abundance and distribution within a population are affected by numerous factors such as population bottlenecks and isolation, founder effects, recent inbreeding, and natural selection, and it has been shown that long ROH are enriched for deleterious homozygotes. Although ROH abundance has been associated with increased risk for various complex traits, it remains unclear the extent to which population history and genetic architecture influence ROH associations with phenotypes. Here we take a simulation approach to characterize the relationship between demographic history, genetic architecture, and a generic complex phenotype. We perform forward-in-time simulations of a three-population human demographic history roughly representing African, European, and Asian continental populations. We simulate a 100 Mb chromosome region with exon structure and a variable recombination map based on the first 100 Mb of human chromosome 1 and allow deleterious mutations with selection coefficients drawn from a gamma distribution. Phenotype is modeled as a function of selection coefficients, with parameters that allow us to vary the relative importance of rare versus common variants in their contribution to the total phenotype. Our results show that demographic history, ROH length, and dominance coefficient are important factors contributing to how much variation in a trait are explained by ROH. We show that ROH influence the simulated phenotype most strongly when alleles are recessive and rare, and that, under these conditions, long ROH (comprised of IBD haplotypes inherited from a very recent ancestor) have more influence than short ROH. These results suggest that the role of ROH in contributing to complex

phenotypes may be largely due to the pairing of rare alleles of recessive effect and that incorporating ROH into disease mapping approaches may help the identification of recessive effects.

209W Experimental Evolution of Hypoxia Tolerance in *Drosophila melanogaster* *Dan Zhou*[1], Gabriel Haddad[1,2] 1) Univ Califonia, San Diego; 2) The Rady Children›s Hospital, San Diego

Hypoxia (lack of oxygen) can be an environmental stress at high altitude or a part of the etiology of many diseases. Understanding the mechanisms regulating hypoxia tolerance or susceptibility is essential for developing novel strategies for prevention or treatment. To do so, we generated 3 populations of hypoxia-adapted flies through experimental evolution over ~300 generations. As compared to the controls, we observed multiple phenotypic changes in flies that evolved under hypoxic conditions. These include a higher oxygen consumption rate under hypoxic condition, a reduced body weight and size, a delayed developmental time that was reversible at later generations growing in the same hypoxic condition, and a reduced level of mitochondrial ROS generation. Furthermore, through whole genomic sequencing and high throughput profiling, we identified genomic regions that were under selection, as well as genes, signaling pathways and metabolic adjustments that play important roles in hypoxia tolerance. We found that the hypoxia tolerant trait is stably maintained in the population stocks even after years of returning to normoxic condition. In addition, we derived a panel of low oxygen tolerant (PLOT) isofemale lines from these hypoxia-adapted populations, which exhibited various levels of hypoxia tolerance under severe, normally lethal, hypoxic conditions. We believe that this unique resource is a powerful tool for identifying and dissecting the mechanisms underlying hypoxia tolerance or susceptibility as well as regulating hypoxic stress-directed evolution, and studying phenotypic plasticity and the genomic dynamics along experimental evolution.

210T Epistatic constraint on RNA secondary structure drives the evolution of SARS-CoV-2 *Mahsa Alemrajabi*, Ksenia Macias Calix, Raquel Assis Florida Atlantic University, Boca Raton, FL

Epistasis is an evolutionary phenomenon whereby the effect of a mutation depends on the genetic background in which it arises. A major source of epistasis in an RNA molecule is its secondary structure, which contains functionally important topological motifs held together by Watson-Crick (WC) base pairs. Here we study epistasis in the secondary structure of the novel RNA coronavirus SARS-CoV-2 by considering population-level frequencies of mutations at ancestral WC base paired sites across the genome. We uncover lower frequencies of mutations at WC than at non-WC base paired sites, supporting the hypothesis that modifications of the SARS-CoV-2 secondary structure are generally deleterious. Further, we find that mutations that convert WC base pairs to G:U “wobble” base pairs are approximately three times more frequent than those that abolish base pairing, suggesting that weak base pairing maintains some integrity of the SARS-CoV-2 secondary structure. Last, we show that WC base paired sites under the strongest epistatic constraint are primarily located in a pseudoknot motif that is involved in programmed ribosomal frameshifting, whereas those under the weakest epistatic constraint are located in a complex motif that is associated with viral pathogenicity. Together, these findings demonstrate the evolutionary importance of the SARS-CoV-2 secondary structure, as well as highlight specific topological motifs and associated viral functions that may be targets of different forms of natural selection.

211T The Evolutionary Patterns of Recombination in North American Gray Wolves (*Canis lupus*) and Domestic Dog (*C. familiaris*) *Christina Del Carpio*[1], Maria Izabel Cavassim[1], Pedro Perez[1], Robert Wayne[1], Kirk Lohmueller[1,2] 1) University of California, Los Angeles, Department of Ecology and Evolutionary Biology; 2) University of California, Los Angeles, David Geffen School of Medicine, Department of Human Genetics

In sexually reproducing species, meiotic recombination enables the proper alignment and segregation of homologous chromosomes while introducing new combinations of alleles into populations. Recombination rates are known to vary on the level of species, populations, sexes, and individuals. Thus, it is a trait that can be acted upon by evolutionary forces such as natural selection and genetic drift. Here we investigated the possible changes in recombination across canids using high-coverage WGS data from a population of North American (NA) gray wolves and one breed of domestic dogs (pugs).We inferred recombination rates from patterns of linkage disequilibrium (LD) using Pyrho, which also models demographic history. Based on this analysis, NW gray wolves (n = 14, mean coverage = ~38X, mean r = 3.1e-9 per bp) have a ~18X higher mean recombination rate than dogs (n = 15, mean coverage = ~48X, mean r = 1.7e-10 per bp). Simulations suggest that this difference in mean recombination rate is not driven solely by differences in demography. Thus, other evolutionary forces may be in play. Future analyses include the identification of recombination hotspots and comparisons of their locations between populations as well as the inclusion of other dog breeds. This work will result in fine-scale recombination maps for NA gray wolves and multiple domesticated dog breeds. The comparison of genetic maps will highlight the possible changes in recombination rate between these sister taxa since the domestication of dogs while also informing about the plasticity of this trait through time.

212W Understanding the heterogeneity in gene regulatory responses to misfolded protein toxicity *Rachel Eder*, Leandra Brettner, Kerry Geiler-Samerotte Center for Mechanisms of Evolution, School of Life Sciences, Arizona State University, Tempe, AZ

Protein misfolding is a problem across all organisms in the tree of life, but the reasons behind misfolded protein toxicity to cells are largely unknown. To better understand toxicity, I investigate if toxicity from misfolded proteins affects all cells equally or affects some cell subpopulations more than others, such as older cells. To define cell subpopulations, I optimized a cutting-edge single-cell RNA sequencing platform (scRNAseq) for yeast, which is a common model organism for investigating protein misfolding. By using scRNAseq in yeast, I study the expression variability of many genes across populations of thousands of cells. I study how the transcriptomes of single cells differ from one another in various conditions: at different stages in the growth phase and with different engineered misfolded proteins. One specific hypothesis I will investigate is whether older cells are more sensitive to misfolded proteins. To do so, I will measure whether cells with gene expression markers related to aging, when challenged by misfolded proteins, express more genes related to stress responses (i.e. chaperones) than younger cells. After identifying subpopulations with a more severe transcriptional response to misfolded proteins, I can study the cells' physiology to gain insights about why that subpopulation is sensitive to misfolded proteins. Thus, understanding the non-uniform distribution of responses to protein misfolding can provide insight into evolutionary biology. For example, it is well known that heterogeneity within microbial populations can be beneficial to their fitness by allowing that population to thrive in diverse environments. This study adds to a growing body of literature documenting and quantifying the degree of non-genetic heterogeneity in nature. Further, it provides insights on the gene regulatory responses associated with misfolded protein toxicity by revealing which type of cells are most sensitive to this intracellular threat.

213W Assessing signatures of selection on transposable elements by accounting for non-uniform transposition rate *Mitra Menon*[1], Robert Horvath[2], Michelle Stitzer[3], Jeff Ross-Ibarra[1] 1) Department of Evolution and Ecology, University of California; 2) Institute for Plant and Microbial Biology, University of Zurich, Zurich, Switzerland ; 3) Institute for Genomic Diversity and Department of Molecular Biology and Genetics, Cornell University

Transposable elements (TEs) make up a majority of flowering plant genomes and contribute towards phenotypic and genotypic diversity. Most earlier studies have assumed TEs experience strong purifying selection due to their ability to move around in the genome and insert into gene coding regions. However, more recent studies have shown that, like any genomic variants, TEs can be neutrally evolving or experience different selection pressures which may include positive and balancing selection. Approaches developed so far to evaluate selection on TEs have relied on the site frequency spectrum (SFS) of TEs and their deviation from a genome wide SFS. But inferring selection based on the SFS alone is problematic due to the non-constant mutational input of TEs. We present an approach to account for variation in TE activity by estimating the age of TE insertions and using an age-adjusted SFS to compare TEs to putatively neutral variants. We evaluated its effectiveness given different demographic histories via simulation and then explore the dynamics of TE variation across 3 different plant species. Additionally, we evaluated selection at the level of individual TE family and TE element by implementing two approaches that look at signatures of selective sweeps and at signatures of balancing selection. In Brachypodium we show that the age-adjusted SFS approach can effectively pick up signals of purifying selection across TEs. However, in Arabis and maize the pattern is complicated by differences in selective pressures amongst TE families. In maize, we show that a number of TE families appear to be under balancing selection, while in Arabis many appear to be neutrally evolving. Overall, we show that the age-adjusted SFS is robust to varying mutation and demographic history and provides a general idea of selection acting across TEs**.**

214T Human populations exhibit correlated abundances and variation of tandem repeat content. *Iskander Said*, Andrew Clark, Daniel Barbash Cornell University

Human genome sequence data, by virtue of the telomere-to-telomere assembly, exceptionally high-quality annotation, and massive sample size, provides an excellent opportunity to investigate population genetic processes at play on repetitive DNA. Some repeats are integral to aspects of cellular and organismal function, such as meiotic segregation and genome regulation, as well as being implicated in complex evolutionary processes, such as speciation and meiotic drive. In humans, population-scale analysis of short tandem repeat polymorphisms (microsatellites) in euchromatin have found high levels of diversity in repeat content and extensive population stratification. Some of this polymorphism has functional consequences and has been implicated in disease etiology. To extend these results to a genome-wide scale, including long tandem arrays that cannot be assembled with short sequence reads, we employ the method *k-seek*, which directly queries unaligned fastq files to discover and quantify tandem repeats consisting of 1-20bp long repeating monomers, without requiring genomic assemblies. We have mined a set of 2,504 high coverage human genomes from the

1,000 Genomes Project to analyze the inter- and intra-population variation of human tandem repeats. We have found over 16,000 distinct tandem repeats, whose expansions and contractions can account for over 10 Mbp of difference in genome size among individuals, with the Y chromosome accounting for a substantial portion of this in males. We see high levels of inter-population divergence, consistent with a high rate of copy number changes. As we have seen in other organisms, there is a striking pattern of correlation in abundances among groups of repeats, whose cause remains a mystery. An exciting future opportunity that will rely on extensive telomere-to-telomere assemblies will be to consider satellite repeat changes localized to specific genomic loci.

215T Nematode genomes reveal a shift in mutation spectrum in the Chernobyl Exclusion Zone *Sophia Tintori*[1], Patrick Ortiz[1], Maxim Ivanenko[2], Igor Chyzhevskyi[3], Timothy Mousseau[4], Matthew Rockman[1] 1) Department of Biology and Center for Genomics & Systems Biology, New York University, NY, NY; 2) Schmalzhausen Institute of Zoology, Kiev, Ukraine; 3) Department of Coordination of International Projects of the State Specialized Enterprise «Ecocentre», Kiev, Ukraine; 4) Department of Biological Sciences, Univeristy of South Carolina, Columbia, SC

Background ionizing radiation is a ubiquitous environmental carcinogen, but it is rarely considered dangerous thanks in part to our DNA repair mechanisms. Like all heritable traits, DNA repair genes can vary between individuals, both in sequence and expression (often considered in the context of individuals' differing genetic predispositions to cancer). Our approach to studying the extent and impact of natural variation in DNA repair components is to compare radiation-tolerant wild strains to radiation-sensitive strains of the same species, interrogating the differences in their genetics and their cellular dynamics in response to chronic radiation exposure.

In search of such radiation-tolerant animals, we have collected nematodes from the fruits and soils of the Chernobyl Exclusion Zone, a landscape altered by contamination from the world's largest nuclear power plant accident 36 years ago. We have recovered and cryopreserved 298 nematode isolates, and are investigating 15 strains of *Oscheius tipulae*, isolated from sites ranging in ambient radiation levels. Long-read sequencing and *de novo* assembly of these 15 animals› genomes, compared to *O. tipulae* from other parts of the world, reveal heritable mutations characteristic of increased double strand break repair.

216W Allelic and array size variation at human centromeres *Carl Veller*[1], Sasha Langley[2], Graham Coop[1], Charles Langley[1] 1) University of California, Davis; 2) University of California, Berkeley

Progress in understanding the evolutionary genetics of centromeres has been hindered by their repetitive sequence content, which has historically precluded their assembly and thus analysis of the genetic variation present at these functionally important genomic sites. Recently, the substantial linkage disequilibrium caused by low recombination in and around centromeres has been exploited to characterize their allelic diversity, without need for assembling the repetitive centromere sequences themselves. Here, we characterize population genetic variation at human centromeres in a large, pedigree-based genomic dataset. We identify major centromere haplotype (cenhap) clades for each chromosome, and confirm that cenhaps typed on this basis show consistent transmission within pedigrees. Next, we develop and validate a read-count based proxy for the total centromere array size per chromosome per individual, and apply a maximum likelihood method to estimate individual centromere array sizes. We find substantial size variation, both within and (especially) between cenhap clades. Examining transmission within family pedigrees, we find no evidence of strong segregation distortion in favor of one cenhap allele over another, or of larger cenhaps over smaller ones. Finally, we estimate full genealogical trees from single-nucleotide variation within cenhaps, and use these trees to evaluate various models for the long-term evolution of centromere array size.

217W Effects of mating system on the molecular evolution and expression of genes in the male reproductive tract of *Peromyscus* mice *Erin Voss*, Michael W. Nachman Museum of Vertebrate Zoology, University of California, Berkeley

Genes involved in reproduction and immunity often evolve rapidly, presumably because they underlie biotic interactions. For example, *Drosophila* accessory gland proteins are a classic case of rapid molecular evolution: male seminal proteins interact with aspects of the female reproductive tract during mating and display strong evidence of rapid evolution driven by sexual selection and conflict. Here, we investigate patterns of molecular evolution and gene expression evolution for genes expressed in three tissues of the male reproductive tract in the context of mating behavior for three species of mice in the genus *Peromyscus*. Postcopulatory sexual selection is mediated by mating behavior: males of promiscuous species face a set of challenges that are different from those in monogamous or polygynous taxa. Males of promiscuous species must take measures to ensure fertilization and exclude the sperm of rival males. We took a comparative approach and asked whether molecular evolution and gene expression patterns differ across species with different mating behaviors. We collected reproductively mature male *Peromyscus californicus*, *Peromyscus boylii*, and *Peromyscus manic-*

ulatus from their native range in the California central coast. These species are monogamous, polygynous, and promiscuous, respectively. We used RNAseq to generate transcriptomes for the testis, the seminal vesicle, and the epididymis, and we describe 1) variation in gene expression and 2) rates and patterns of nonsynonymous changes in protein coding sequence within the context of mating behavior and across different tissues in the male reproductive tract.

218T Seasonal plasticity and adaptive fluctuations of gene expressions of *D. melanogaster* *Yang Yu*, Alan Bergland University of Virginia

Two major mechanisms for populations of short-lived organisms to respond to temporal environmental heterogeneity, such as seasonality, are adaptive tracking and plasticity. Theory predicts that transition from one mechanism to another will have detrimental effects on the populations, thus the genetic architecture between adaptive tracking and plasticity should have limited overlap. Seasonal adaptive tracking can be observed from allele frequency change at individual loci and at genome-wide levels. Although thousands of SNPs have been shown to shift in frequency repeatedly across seasons in *Drosophila melanogaster* populations, we still have limited understanding of the relative strength of evolutionary mechanisms that underlie seasonal adaptation. In this study, we test the hypothesis that there is distinct genetic architecture between adaptive tracking and plasticity by using genome-wide plastic gene expression profiles across the season. We first identify the genes that plastically change in expression levels across the season to gain insight into the functional response to seasonal environmental changes. Second, we will combine publicly available eQTL profiles, genome-wide allele frequency data from multiple seasonal populations, with our seasonal gene expression data to test whether the eQTLs associated with plastic genes are de-enriched for seasonal SNPs. We created and reared a genetically controlled F1 fly population from 24h embryos to adulthood in an experimental orchard across 10 seasonal time points (May to Oct in 2019) to examine the differentially expressed (DE) genes across the season. We extracted whole-tissue RNA from 3-5-day-old adult female flies and prepared pooled libraries using bulk RNA barcoding method (BRBSeq). Our next step is to identify the plastic genes and test our hypothesis that the eQTLs associated with these genes are de-enriched for seasonal SNPs. We hope to provide insight into the general understanding of seasonal adaptation from an expression perspective, and how seasonal environmental heterogeneity maintains functional genetic variation at eQTLs across time.

219T Tracking the origins and rapid rise of two distinct insecticide resistance haplotypes *Jennifer Baltzegar*, Fred Gould North Carolina State University

The mosquito, *Aedes aegypti*, is the primary species that transmits dengue, zika, chikungunya, and yellow fever worldwide. Currently, insecticides (e.g., pyrethroids) remain the most commonly used method to decrease the population size of these pests and reduce disease incidence, but resistance is becoming a widespread problem.

Knockdown resistance (*kdr*) is a specific type of pyrethroid resistance caused by one or more single nucleotide polymorphisms (SNPs) in the voltage-gated sodium channel gene (*vgsc*) that alter the vgsc protein in a way that decreases the effect of pyrethroid on mosquito survival. Many SNPs that contribute toward the *kdr* phenotype have been identified worldwide. Two common SNPs in the Western Hemisphere are F1534C and V1016I.

We previously reported the rapid evolution of two *kdr* haplotypes in a population of mosquitoes from Iquitos, Peru. The rapid rise of *kdr* haplotypes was dominated first by the increase in individuals carrying only the Cys1534 resistance SNP and later by individuals carrying both the Cys1534 and Ile1016 resistance SNPs. Here, we further examine this phenomenon by interrogating polymorphic regions of the *vgsc* gene that are not functionally associated with *kdr* resistance. This information will help to elucidate existing questions about the dynamics that underlie *kdr* resistance in this mosquito. Do resistant individuals arise from one or multiple common ancestors? Are the resistance haplotypes selected from standing genetic variation or from *de novo* mutations? Does population structure of the mosquito within the city impact the spread and interactions between the two resistance SNPs? Gaining a better understanding of the evolutionary dynamics that govern the rise of insecticide resistance will aid in improving resistance management.

220W The genomic basis and repeatability of rapid seasonal evolution *Mark Bitter*[1], Sharon Greenblum[1,2], Seth Rudman[3,4], Subhash Rajpurohit[3,5], Nicolas Betancourt[3,6], Mary Catherine Berner[3], Skyler Berardi[3], Dmitri Petrov[1], Paul Schmidt[3] 1) Department of Biology, Stanford University, Stanford, CA; 2) DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA; 3) Department of Biology, University of Pennsylvania, Philadelphia, PA; 4) School of Biological Sciences, Washington State University, Vancouver, WA ; 5) Division of Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Navrangpura, Ahmedabad; 6) School of Medicine, Stanford University, Stanford, CA

Determining the tempo and repeatability of adaptation in natural populations is of central interest in evolutionary biology. Seasonal environments, which fluctuate cyclically and on a timescale that lends to repeated empirical observation, provide an excellent system within which to approach this venture. For example, Rudman *et al.* (2021) recently leveraged a single spring to fall transition, and a highly replicated, semi-natural mesocosm system, to demonstrate dynamic patterns of phenotypic and genomic evolution in *Drosophila melanogaster*. Notably, the study identified concurrent phenotypic evolution and large allele frequency shifts at numerous independent loci, with changes in the direction of selection throughout the sampling period. Still, inferring the generality of the observed dynamics hinges upon repeated quantification across independent years of sampling. Accordingly, using the same experimental system, I will present results of three additional years of genomics data to show that adaptive tracking of seasonal fluctuations in the environment is indeed a repeatable phenomenon in *D. melanoagster*. I will further discuss the extent to which the genetic basis of seasonal adaptation exhibits parallelism across multiple years of sampling. Finally, I will describe ongoing work aimed at drawing direct links between observed patterns of genetic and phenotypic variation within this system. Specifically, I will present results exploring discrepancies between the large-efect size loci underlying variation in a seasonally evolving trait, pigmentation, and those loci driving the evolution of the trait between the spring and fall.
References: Rudman S.M., *et al.* (2021) *bioRxiv* doi: 10.1101/2021.04.27.441526.

221W When B is shaped like a U: is weak selection on deleterious alleles important? *Vincent Buffalo*, Andrew Kern, Peter Ralph University of Oregon

Selection on new, deleterious mutations affects the evolutionary dynamics of nearby linked loci. While the classic background selection (BGS) model (e.g. Charlesworth et al. 1993, Nordborg et al. 1996) accurately predicts variation under relatively strong selection and weak mutation rates, this model breaks down under weak selection and/or high mutation rates. A number of studies have tried to quantify the impact of BGS using a summary statistic, B, which is the expected reduction in pairwise diversity due to purifying selection (McVicker et al. 2009). B varies along a chromosome as a function of the genetic map and the spatial arrangement of the putative selected sites. B has also been used as the basis for maximum likelihood-based estimators of the strength of selection and deleterious mutation rates. Using forward-time simulations, we show that these commonly-used maximum likelihood-based methods lead to erroneous estimates in the weak selection regime, even under the simplest demographic scenarios. We find weak selection also biases our estimates of levels of diversity in the absence of linked selection, an important quantity used to distinguish the effects of drift from natural selection. Furthermore, we explore alternative, tree-based methods to find regions where BGS models may not be suitable and apply these to inferred tree sequences in humans. Finally, we show that these two selection regimes have different implications for genetic load in a species,

222T The battle of the sexes in humans is highly polygenic *Jared Cole*[1], Peter Golightly[1], Arbel Harpak[1,2], Mark Kirkpatrick[1] 1) Department of Integrative Biology, University of Texas at Austin, Austin, TX; 2) Department of Population Health, University of Texas at Austin, Austin

Sex-specific selection, which occurs when the fitness effects of alleles differ in males and females, has been implicated in the maintenance of genetic variation in natural populations and sex differences in health and disease. Because the sexes mix their autosomal genetic makeup each generation, it has been challenging to quantify the intensity of sex-specific selection using conventional population genetic approaches. Here, we introduce a novel method for estimating the strength of sex-specific selection. Our approach is builds on subtle differentiation in haplotype structure between the sexes that may be generated by viability selection during a single generation. We apply the method to haplotype data from 250K individuals in the UK Biobank. We confirm previous reports showing weak-to-undetectable sex-specific selection at the level of individual loci. At the same time, as we aggregate genome-wide evidence, we estimate a highly polygenic sexually-antagonistic selection. In summary, our results bridge the gap between the underwhelming evidence in human data to date with the longstanding theoretical expectation of pervasive sex-specific selection.

223T Estimation of selection components in a pedigreed population of Florida Scrub-Jays *Elissa Cosgrove*[1], Reed Bowman[2], John Fitzpatrick[1], Andrew Clark[1], Nancy Chen[3] 1) Cornell University, Ithaca, NY, USA; 2) Archbold Biological Station, Venus, FL, USA; 3) University of Rochester, Rochester, NY, USA

Natural selection is a complex process that can impact multiple stages of the life cycle. Exhaustive enumeration of populations, along with construction of the full population pedigree, opens the opportunity to estimate different selection components and evaluate the role of antagonism in maintaining genetic variation for fitness. Here, we infer the action of natural selection on fitness components using a 25-year genomic, phenotypic, and pedigree dataset in the Florida Scrub-Jay (*Aphelocoma coerulescens*), a species in rapid decline due to habitat loss. A population of Florida Scrub-Jays at Archbold Biological Station has been studied since 1969, resulting in annual and lifetime fitness measures for thousands

of individuals on a 14-generation pedigree. We genotyped every individual in our study population over two decades at 15,416 genome-wide SNPs. To test for selection acting on specific life-cycle stages, we modified existing selection component analysis frameworks to take full advantage of exhaustive population sampling. We used generalized linear mixed models that included kinship and considered tens of potential ecological covariates to test for differences in viability, fecundity, mating success, and overall lifetime reproductive success (LRS). We identified 12 loci under viability or fecundity selection and 5 loci with significant differences in LRS among genotypes. In comparing genome-wide effect size estimates between different fitness components, we observed a strong positive correlation between survival and LRS. We also found a negative correlation between effect sizes for male and female LRS, which may reflect a degree of sexual conflict, and between female lifespan and female fecundity, which may reflect the classic life history trade-off between survival and reproduction. This fine-scale dissection of selection components provides important insights into the role of selection in maintaining genetic variation in a natural population.

224W Does adaptation to past viral infections involve changes in protein stability in host virus-interacting proteins? *Chenlu Di*[1], Jesús Murga-Moreno[2], David Enard[1] 1) University of Arizona; 2) Autonomous University of Barcelona

A large proportion of protein adaptation in human evolution was driven by past viral infections. It is not clear however what effects adaptive amino acid changes had on host proteins that were advantageous against viruses. We hypothesize that viral-interacting host proteins adapted in response to viruses through changes in protein thermodynamic stability, or the balance between the folded and unfolded state of proteins. Protein stability is an important property broadly studied in biophysics, protein engineering and drug design. A destabilizing amino acid change can decrease the proportion of folded functional proteins at physiological temperature, and might thus reduce the protein function of a proviral host protein that a virus needs to complete its cycle. Conversely, adaptive amino-acid changes may stabilize and thus increase the amount of a host antiviral protein. We used an Approximate Bayesian Computation version of the McDonald-Kreitman test called ABC-MK to compare the rate of adaption in human proteins between amino acid changes that strongly changed protein stability and changes that do not. Our results suggest that a large amount of protein adaptation to viruses was driven by changes in protein stability, rather than the classic assumption of adaptation at the host-virus contact interface.

225W Relentless Selection: Trait divergence under high gene flow *Moritz Ehrlich*[1], Amanda DeLiberto[1], Melissa Drown[1], Dominique Wagner[2], Marjorie Oleksiak[1], Douglas Crawford[1] 1) University of Miami, Miami, FL; 2) University of Colorado, Boulder, CO

Selection continuously reshapes the genetic and phenotypic composition of populations, yet often adaptive changes are not propagated through time due to e.g. prohibitively high drift or gene flow. Nevertheless, selection may still induce temporary phenotypic divergence at extremely small spatial and short temporal scales. These changes may ultimately be of higher ecological importance than long-term evolutionary trends. Selection on polygenic traits in particular may allow for divergence to occur repeatedly every generation without significant reduction in standing genetic variation.

The teleost *Fundulus heteroclitus* inhabits salt marsh estuaries characterized by high environmental heterogeneity. Population sizes within a marsh are large (>10K) and panmictic breeding results in negligible genetic structure. Yet individual fish demonstrate high site fidelity to distinct microhabitats e.g. tidal ponds or coastal bays, each exhibiting highly disparate temperature and oxygen regimes.

We tagged/recaptured 2000/200 *F. heteroclitus* and confirmed residency in two microhabitats: a cooler, oxygenated coastal basin and hotter, anoxic tidal ponds. After common-gardening basin and pond residents we measured several, fitness-related traits; critical thermal maximum, resting metabolic rate, cardiac metabolic rate and aquatic surface respiration (ASR) latency. We found significant phenotypic divergence among basin and pond residents in resting metabolic rate (5%, p=0.02), cardiac metabolic rate (9%, p=0.004) and ASR latency (15%, p=0.007), suggestive of divergent selection.

We further identified >10,000 genome-wide single nucleotide polymorphisms using a genotyping-by-sequencing (GBS) approach. By sampling each microhabitat at two time points within a single generation we determined allele frequency change over time. Few individual SNPs show significant allele frequency changes beyond that expected by random mortality. However, the proportion of SNPs exhibiting i) allele frequency changes and ii) divergence among microhabitats is significantly elevated over the neutral expectation. These patterns are consistent with selection on polygenic traits where minor allele frequency changes at multiple loci of small-effect may cause significant phenotypic shifts.

Despite high gene flow and negligible demography, *F. heteroclitus* displays surprising phenotypic divergence among microhabitats following common-gardening. This is unlikely due to a plastic response but rather divergent selection on polygenic traits. Given high standing genetic variation and large population sizes, selection may be effective enough to regenerate phenotypic divergence repeatedly every generation. While this does not cause any long-term evolutionary change, such temporary phenotypic heterogeneity may be of high ecological importance.

226T Evolutionary dynamics in the human gut microbiome from infancy through adulthood Daisy Chen[1,2], *Nandita Garud*[1] 1) University of California, Los Angeles; 2) University of California, San Diego

The human gut microbiome is comprised of a complex ecosystem of microbes that reside inside of us and play an important role in our health. While the ecological dynamics of the microbiome have been intensely studied, we currently know very little about the tempo and mode of evolution of gut microbiota and how these dynamics might change over a person's lifetime. We recently quantified the evolutionary dynamics of ~40 prevalent species of gut bacteria from infancy through adulthood. We find evidence for almost 100-fold increase in the rate of evolution and strain turnover in the infant gut compared to healthy adults, with the mother-infant transition at delivery being a particularly dynamic period in which gene loss dominates. Within a few months after birth, these dynamics stabilize, and gene gains become increasingly frequent as the microbiome matures. We furthermore find that evolutionary changes in infants show signatures of being seeded by a mixture of *de novo* mutations and transmissions of pre-evolved lineages from the broader family. Several of these evolutionary changes occur in parallel in multiple infants, highlighting candidate genes that may play important roles in the development of the infant gut microbiome. Our results point to a picture of a volatile infant gut microbiome characterized by rapid evolutionary and ecological change in the early days of life.

227T Selection on gene expression under salinity stress in Rice *Sonal Gupta*[1], Zoe Joly-Lopez[2], Simon Groen[3], Steven Franks[4], Michael Purugganan[1,5] 1) New York University, New York; 2) Université du Québec à Montréal, Quebec; 3) University of California, Riverside; 4) Fordham University, New York ; 5) New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Changes in gene expression is an important aspect of adaptation. Stressful environments can act as selective pressure leading to alterations in relative gene expression levels. Soil salinity is one of the most common stressful environments, especially for rice (*Oryza sativa*) in the arid and semi-arid regions. Here we assayed gene expression for a total of ~1200 individuals from 195 lines of rice exposed to control and salinity stress. Next, using phenotypic selection analysis we estimate the type and strength of selection on the levels of gene expression for over 18,000 transcripts. We found that in contrast to the control conditions, wherein majority of the transcripts are under (nearly) neutral selection or under weak stabilizing selection, the strength of selection increases under salinity stress. Moreover, we found that a higher proportion of transcripts are conditionally neutral (10%) than are antagonistically pleiotropic (0.378%) indicating a lack of trade-off at the expression levels. Interestingly, multivariate selection revealed pathways associated with amino-acid metabolism to be enriched under salinity stress suggesting a role of amino-acid metabolism in response to salinity, as has been suggested in other species. Additional work will provide insights into the molecular underpinnings of response to salinity stress in rice.

228W Genomic basis of climatic adaptation and parallel evolution in house mice from North and South America *Yocelyn Gutierrez-Guerrero*[1], Megan Phifer-Rixley[2], Felipe Martins[1,3], Michael W Nachman[1] 1) Museum of Vertebrate Zoology, Department of Integrative Biology, University of California Berkeley, Berkeley, CA, US; 2) Department of Biology, Monmouth University, West Long Branch, New Jersey, US; 3) Department of Ecology and Evolution, The Australian National University Acton, Canberra, Australia

Parallel evolution in response to similar selection pressures can provide strong evidence of natural selection. For example, *Drosophila melanogaster* shows phenotypic and genotypic clines in opposite directions in the northern and southern hemispheres. House mice (*Mus musculus domesticus*) also provide an excellent opportunity for studying the genetic basis of how species adapt to new environments. Native to Europe, they have been spread around the world in the association with humans in the last few hundred years. Previous studies of house mice from latitudinal transects in eastern and western of North America identified phenotypes and genes that underlie environmental adaptation, including differences in body size and signatures of parallel evolution in genes involved in thermoregulation. Here, we expand the geographic scale to include a latitudinal transect across South America with the aim of identifying unique and shared responses to selection in the northern and southern hemispheres. We sequenced the complete exomes of 76 wild-caught mice spanning 50 degrees of latitude across Brazil and Argentina. Principal components analysis revealed a clear separation between populations from North and South America. Using a Latent Factor Mixed Model (LFMM Ridges), we conducted genome-wide scans for selection by looking for associations between genotypes and environmental

variables separately in North and South America. We detected thousands of SNPs significantly associated with latitude, annual mean temperature, and temperature seasonality (*q-value*≤ 0.001 and *zscore*≥ 2). In the populations from South America, we identified candidate genes involved in lipid metabolism, circadian rhythm, morphogenesis, and regulation of cold thermogenesis. To explore parallel evolution in the northern and southern hemispheres, we identified the subset of genes showing significant associations in each transect. Permutation tests revealed more overlap than expected by chance (*pvalue*≤ 0.001). Of particular note is the discovery that the genes underlying cold and heat sensation (*Trpm8* and *Trpm2)* show clear signatures of selection in both hemispheres. Since mice independently colonized North and South America, these results suggest repeatable evolution driven by adaptation to thermal conditions. Moreover, newly developed wild-derived inbred strains of mice from different latitudes provide opportunities for functional tests of alternative alleles at *Trpm8* and *Trpm2* in the laboratory.

229W Using blood group serology and whole genome sequence data to identify malaria-protective variants introduced through admixture in Oman *Paige E. Haffener*[1], Arwa Z. Al-Riyami[2], Shoaib Al-Zadjali[3], George B. J. Busby[4], Mohammed Al-Rawahi[5], Saif Al Hosni[2], Ali Al Marhoobi[5], Ammar Al Sheriyani[6], Ellen M. Leffler[1] 1) Department of Human Genetics, The University of Utah, Salt Lake City, UT, USA; 2) Department of Hematology, Sultan Qaboos University Hospital, Muscat, Oman ; 3) Sultan Qaboos Comprehensive Cancer Center, Muscat, Oman ; 4) Allelica, London, UK ; 5) Department of Hematology, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman ; 6) Royal Oman Police Hospital, Muscat, Oman

Oman is a country in the southeast corner of the Arabian Peninsula (AP) that historically struggled with malaria endemicity caused by both *Plasmodium falciparum* and *P. vivax*. AP populations in the southern region of the peninsula show genetic admixture with African populations, introduced partly through the Arab slave trade, providing a source for malaria-protective variants that may have been selected for in the Omanis. Indeed, the Duffy null allele at the *DARC* locus, which confers protection from *P. vivax* malaria, shows intermediate frequencies and excess African ancestry in several AP and other admixed populations consistent with post-admixture selection. We hypothesize that variants protective against *P. falciparum* malaria, including several other blood group variants, may also have been introduced by admixture and undergone subsequent positive selection. To test this, we generated whole genome sequence data with paired serological phenotypes for 10 blood groups, including three that have been associated with malaria, from 100 native, healthy Omani blood donors. Concordance between serological and genetically inferred phenotypes was high for all blood groups, with most mismatches explained by rare or novel coding and structural variants. Notably, we identify a distinct Duffy-null allele in Oman that is present across other AP populations but absent elsewhere. We also report the presence of the Dantu variant in the MNS blood group system, which was recently found to be protective against *P. falciparum* malaria and previously reported only in East African populations. Global admixture inference further supports significant admixture with East African populations, consistent with the history of the Omani empire. Local ancestry inference using RFMix2 indicates an excess of African ancestry in Omanis at the *DARC* locus, suggestive of positive selection. Ongoing work includes additional tests for selection at *DARC* and other malaria-protective loci.

230T A two-step adaptive walk rewires nutrient transport in a challenging edaphic environment Emmanuel Tergemina[1], Ahmed F. Elfarargi[1], Paulina Flis[2], Andrea Fulgione[1], Mehmet Göktay[1], Célia Neto[1], Marleen Scholle[3], Pádraic Flood[1], Sophie-Asako Xerri[1], Johan Zicola[1], Nina Döring[1], Herculano Dinis[4,5], Ute Krämer[3], David Salt[2], *Angela M. Hancock*[1] 1) Max Planck Institute for Plant Breeding Research, Cologne, DE; 2) Future Food Beacon of Excellence and the School of Biosciences, University of Nottingham, Sutton Bonington Campus, Nr Loughborough, LE12 5RD, United Kingdom; 3) Faculty of Biology and Biotechnology, Ruhr University Bochum, 44801 Bochum, Germany; 4) Parque Natural do Fogo, Direção Nacional do Ambiente, 115 Chã d'Areia, Praia, Santiago, Cabo Verde; 5) Associação Projecto Vitó, 8234, Xaguate, Cidade de São Filipe, Fogo, Cabo Verde

Most well-characterized cases of adaptation involve single genetic loci. Theory suggests multi-locus adaptive walks should be common, but these are challenging to identify in natural populations. Here, we combine trait mapping with population genetic modeling to show that a two-step process rewired nutrient homeostasis in a population of *Arabidopsis* as it colonized the base of an active stratovolcano characterized by extremely low soil manganese (Mn). First, a variant that disrupted the primary iron (Fe) uptake transporter (IRT1) swept quickly to fixation in a hard selective sweep, increasing Mn but limiting Fe in the leaves. Second, multiple independent tandem duplications occurred at *NRAMP1* and rose to near-fixation in a soft selective sweep, compensating *IRT1* loss by improving iron homeostasis. This study provides a well-resolved case of a two-locus adaptive walk and reveals how adaptive genetic variants reshaped a phenotype and spread over space and time.

231T Contrasting the tempo and mode of adaptation on the X chromosome and the autosomes in *Drosophila*

melanogaster *Mariana Harris*[1], Nandita Garud[2,3] 1) Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA; 3) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA

Adaptation on the X chromosome has attracted significant interest from evolutionary biologists because its dynamics seem to be distinct from that of autosomes. The X chromosome is hemizygous in males, leaving it fully exposed one third of the time to the effects of natural selection and, thus, potentially subject to different evolutionary dynamics than autosomes. Here, we investigate the differences in the mode and tempo of adaptation in the X chromosome and autosomes. Specifically, we test the hypothesis that hard sweeps, in which a single adaptive mutation rises to high frequency, are more common on the X chromosome due to there being a smaller effective population size and a reduction in standing genetic variation resulting from more efficient selection. By contrast, we predict that soft sweeps, in which multiple haplotypes rise to high frequency simultaneously, are more common on the autosomes. We quantify the incidence of hard versus soft sweeps in North American *D. melanogaster* population genomic data with haplotype homozygosity statistics and find an enrichment of hard sweeps on the X chromosome relative to the autosomes, confirming predictions we make from simulations. This suggests that signatures of selection may differ between the X chromosome and the autosomes in *D. melanogaster*. Understanding these differences may enable a deeper understanding of how important phenotypes arise as well as the impact of fundamental evolutionary parameters on adaptation, such as dominance, sex-specific selection, and sex-biased demography.

232W Polygenic adaptation is not a major driver of disparities in disease mortality across global populations *Ujani Hazra*, Joseph Lachance Georgia Institute of Technology

Background and objectives
Health disparities are due to a range of socioeconomic and biological causes, and many common diseases have a genetic basis. Divergent evolutionary histories cause allele frequencies at disease-associated loci to differ across global populations. To what extent are differences in disease risks due to natural selection?

Methodology
Examining a panel of nine global populations, we identified which of the 20 most common causes of death have the largest health disparities. Polygenic risk scores were computed and compared for 11 common diseases for the same nine populations. We then used *PolyGraph* to test whether differences in disease risk can be attributed to polygenic adaptation. Finally, we compared human development index statistics and polygenic risk scores to mortality rates for each population.

Results
Among common causes of death, HIV/AIDS and tuberculosis exhibited the greatest disparities in mortality rates. Focusing on common polygenic diseases, we found that genetic predictions of disease risk varied across global populations (including elevated risks of lung cancer in Europeans). However, polygenic adaptation tests largely yielded negative results when applied to common diseases. Our analyses revealed that natural selection was not a major cause of differences in disease risks across global populations. We also found that correlations between mortality rates and human development index statistics were stronger than correlations between mortality rates and polygenic predictions of disease risks.

Conclusions and implications
Although evolutionary history contributes to differences in disease risks, health disparities are largely due to socioeconomic and other environmental factors.

233W A fundamental constraint on adaptation of a biological module *Minkyu Kim*, Sarah Ardell, Sergey Kryazhimskiy University of California, San Diego

If mutations are highly pleiotropic, that is if each mutation affects many traits, it is difficult for natural selection to improve fitness, a phenomenon known as “the cost of complexity”. One solution to this challenge is modularity, where pleiotropy primarily occurs among genes in the same module but is limited between them. However, it was recently suggested that modularity creates a new challenge termed “evolutionary stalling”. In rapidly evolving populations with limited recombination, evolutionary stalling occurs when beneficial mutations in one module are ‘wasted’ because they cannot successfully compete against those in another module. A recent theoretical work shows that evolutionary stalling of a module depends on its rate of adaptation relative to other modules under selection, where faster modules adapt and slower modules stall. However, as the organism adapts, we expect that the rate of adaptation of the initially faster

modules would generically decline, due to the depletion of the supply of beneficial mutations and diminishing returns epistasis, suggesting that evolutionary stalling may be alleviated during long-term adaptation. Whether such alleviation indeed occurs and under what conditions has not been investigated, and the impacts of long-term adaptation on evolutionary stalling and vice versa are unknown.
Here, we address this question in the Fisher's Geometric Model with two modules. We find that whether or not evolutionary stalling is alleviated over the long run depends on the architecture of the organism and the strength of selection. In particular, if module adaptation slows down solely due to the depletion of the supply of beneficial mutation, stalling persists unabated. In contrast, evolutionary stalling is alleviated over long times if modules are under equal selection pressures and their adaptation slows down solely due to the diminishing returns epistasis. In the more realistic cases where modules are not under equal selection pressures and/or both depleting supply of mutations and epistasis are in effect, we find that the degree of evolutionary stalling changes over time, but it is not necessarily alleviated in the long run. These results suggest that evolutionary stalling imposes a fundamental constraint on the speed of adaptation of individual modules in organisms with limited recombination and that this constraint may not be consistently alleviated by natural selection, even over long time scales.

234T Fitness effects for *Ace* insecticide resistance mutations are determined by ambient temperature *Anna Maria Langmüller*[1,2], Viola Nolte[1], Ruwansha Galagedara[1,2], Rodolphe Poupardin[1,3], Marlies Dolezal[4], Christian Schlötterer[1] 1) Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria; 2) Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria; 3) Paracelsus Medical University Salzburg, Strubergasse 21, 5020 Salzburg, Austria; 4) Plattform Bioinformatik und Biostatistik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

Background: Insect pest control programs often use periods of insecticide treatment with intermittent breaks, to prevent fixing of mutations conferring insecticide resistance. Such mutations are typically costly in an insecticide-free environment, and their frequency is determined by the balance between insecticide treatment and cost of resistance. *Ace*, a key gene in neuronal signaling, is a prominent target of many insecticides and across several species, three amino acid replacements (I161V, G265A, and F330Y) provide resistance against several insecticides. Because temperature disturbs neuronal signaling homeostasis, we reasoned that the cost of insecticide resistance could be modulated by ambient temperature.

Results: Experimental evolution of a natural *Drosophila simulans* population at hot and cold temperature regimes uncovered a surprisingly strong effect of ambient temperature. In the cold temperature regime, the resistance mutations were strongly counter-selected ($s = -0.055$), but in a hot environment, the fitness costs of resistance mutations were reduced by almost 50% ($s = -0.031$). We attribute this unexpected observation to the advantage of the reduced enzymatic activity of resistance mutations in hot environments.

Conclusion: We show that fitness costs of insecticide resistance genes are temperature-dependent and suggest that the duration of insecticide-free periods need to be adjusted for different climatic regions to reflect these costs. We suggest that such environment-dependent fitness effects may be more common than previously assumed and pose a major challenge for modeling climate change.

Keywords: Drosophila, Experimental Evolution, Insecticide Resistance, Temperature

235T The impact of background selection on complex traits *Xinyi Li*[1], John Novembre[2,3], Jeremy Berg[2] 1) Committee on Genetics, Genomics & Systems Biology, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Ecology and Evolution, University of Chicago, Chicago, IL

How different evolutionary processes maintain phenotypic variation is an important question in human genetics. While the importance of background selection in shaping patterns of neutral genetic diversity is well-studied, its influence on the genetic architecture and the prevalence of complex disease is not well understood.

First, we investigated how background selection influences negatively selected mutations. By approximating the effect of background selection as a reduction in effective population size, we derived the predicted genetic diversity reduction of deleterious variants. We found while background selection reduces the genetic diversity of weakly selected sites ($4Ns < 5$), it has minimal effect on strongly selected sites. We performed forward simulation to confirm our theoretical predictions.

Second, as a direct consequence of genetic diversity reduction by background selection, when the causal variants of a complex trait are effectively neutral or weakly selected, background selection reduces heritability of the traits, confirmed by simulations.

Third, to study the disease prevalence, we extended the liability threshold model to include background selection. In the model, mutational pressure increases the liability while selection acts to reduce it, and the equilibrium disease prevalence arises due to a balance between the two forces. We found that when background selection is included, it alters this equilibrium by reducing the genetic variance for liability, which drives an increase in disease prevalence. To validate our theoretical results, we performed forward-time SLIM simulations and measure disease prevalence with varying intensity of background selection. We found scenarios where, for example, when the diversity reduction due to background selection is roughly 20% (as estimated in humans), the disease prevalence increases 10%. In addition, we also examined whether the distortion of site frequency spectrum changes the disease prevalence and found no meaningful effects.

From these investigations, we concluded that background selection reduces genetic diversity of weakly selected variants, but not strongly selected variants. As a result, background selection shifts the distribution of genetic variance more toward rare alleles with large effect than it otherwise would be. In addition, background selection increases disease prevalence primarily through overall levels of diversity. This project addresses the importance of population genetic models in understanding phenotypic variation.

236W Genetic Basis of Lethal Alleles in Nature *Sarah Marion*[1], Brenda Manzano-Winkler[1], Hannah John[2], Iman Hamid[1], Mohamed Noor[1] 1) Duke University, Durham, NC; 2) National Resilience, Inc., Durham, NC

For nearly a century, evolutionary biologists have observed chromosomes which cause lethality when made homozygous persisting at surprisingly high frequencies (>25%) in natural populations of many species. These curious findings provide a challenging question fundamental to understanding natural fitness variation: given the extreme deleterious nature of lethal mutations, why do they appear at such high frequencies? Although most research on this question dates to the 1930's, we still know remarkably little about the genetic basis of naturally occurring lethal mutations or what evolutionary forces create their frequencies. Before even attempting to explain why lethal alleles (lethals) are so common in nature, it is crucial to understand their genetic underpinnings. Decades of research has assumed lethal chromosomes are due to single locus, loss-of-function mutations, but this *has never been directly tested*. Even less is known about the classes of genes in which lethal mutations occur or their distribution throughout the genome.
We are using wild *Drosophila melanogaster* to determine the genetic basis of lethal mutations. *D. melanogaster* were collected from a natural population in Durham, North Carolina, and balancers were used to isolate ~300 independent samples in which chromosome 2 is homozygous lethal. We are in the process of crossing deficiency lines that span over 70% of the second chromosome to map lesions in all lethal isolines as a novel demonstration that most lethal alleles are single-locus and result in loss of function. Fine scale mapping and sequencing results provide novel characterization of naturally occurring lethals, including a nonsense mutation in the *drosha* gene. While this provides evidence that at least some lethal alleles are single locus, we also provide novel evidence that most or all lethal alleles are single locus using Poisson-based mapping expectations.
We present the most extensive mapping study of naturally occurring lethal alleles ever conducted. Our results are the first direct demonstration for a single locus, loss-of-function mode of action of lethal mutations and provide sequence-level characterization of naturally occurring lethal alleles.

237W Balanced Inversions help maintain sexually antagonistic polymorphism *Christopher McAllester*, John Pool UW Madsion

Inversion polymorphisms are well documented across many taxa, despite the potential generation of unfit, unbalanced gametes from inversion heterozygotes. Inversions may fix as a result of linkage with beneficial alleles or due to drift, but many inversions are maintained at intermediate, in some cases clearly balanced frequencies, potentially by linking alleles that share conditional benefit. In African *Drosophila melanogaster*, paracentric inversions are common and many inversions are stably polymorphic throughout diverse African lowland habitats, suggesting the involvement of evolutionary forces beyond local adaptation. We hypothesize that balanced sexually antagonistic selection may be responsible for maintaining the stable polymorphism, in line with the active competition among *D. melanogaster* males and the potential for sexual antagonism. We used a novel forward population simulator with parameters based on *D. melanogaster* life history to model inversion evolution in a population under sexually antagonistic selection at infinite loci and with male reproductive skew. Simulations demonstrated (1) balanced polymorphism involving alleles with a range of antagonistic effects, (2) the persistence of such polymorphic alleles at many loci only under linkage due to competitive

effects, and (3) the rise in frequency and stable persistence of inversions that establish such linkage associations between sets of sexually antagonistic alleles. We followed with an empirical exploration of the selection dynamics on inversions between a pooled Zambian paternal population and their embryo and aged adult offspring to detect correlations between the inversion status, viability and mating fitness. Results demonstrated non-neutral frequency changes, consistent with a complex fitness landscape in which only Inversion 3RK demonstrated a consistent tradeoff between male reproductive success and viability-longevity under the conditions and frequencies tested. This establishes the potential presence of this modeled dynamic in *D. melanogaster* inversions. This model has implications for sex chromosome evolution, as a segregating autosomal antagonistic haplotype would benefit from linkage to a sex determining locus. Further, balancing selection upon epistatic haplotypes, particularly due to sexual or ecological antagonistic selection, may contribute significantly to genetic diversity and ongoing evolution and local adaptation in natural populations.

238T Plasticity in body size in response to diet among wild derived strains of house mice from the Americas *Megan Phifer-Rixey*, Tiffany Longo, Jesse Bragger, Sebastian Vera, Kristi McDonald Biology Department, Monmouth University, West Long Branch, NJ

Body size in house mice (*Mus musculus domesticus*) covaries with latitude, consistent with Bergmann's Rule. There is clear evidence that genetic variation contributes to this pattern in the Americas. However, traits related to body size are complex not only because they are polygenic, but also because they are sensitive to the environment and, in particular, diet. In nature, mice are subject to variation in food availability/quality and latitude is expected to affect seasonality of food resources. New wild-derived strains from regions of the Americas have been developed which vary consistently in aspects of body size in a common laboratory environment, presenting an opportunity to investigate plasticity in response to diet. In this study, male and female mice were fed either a typical breeder diet or a high fat diet after weaning. Body weight, aspects of size, and food intake were measured regularly over twelve weeks. Results suggest not only that diet affected body size, but also that strains differed in response to the high fat diet. Future research will include analysis of differential expression among strains on both diets to help connect regulatory variants to differences in body size.

239T Allele ages reveal signature of balancing selection in human populations *Alyssa Pivirotto*[1,3], Alexander Platt[4], Ravi Patel[2,3], Sudhir Kumar[2,3], Jody Hey[1,3] 1) Center for Computational Genetics and Genomics, Temple University, Philadelphia, PA; 2) Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA; 3) Department of Biology, Temple University, Philadelphia, PA; 4) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

To understand the evolutionary history of a population and expression of phenotypic traits including disease, we first have to understand how selection affects underlying mutations. Here, a new approach is described for assessing the evolutionary history of selection on non-synonymous mutations. The method makes use of the evolutionary probability (EP) approach that leverages a phylogeny with divergence times to assign each amino acid a posterior probability of appearing at a specific loci. We consider the EP of segregating alleles together with estimates of their age and frequency. If a mutation is under selection, we expect it to have a young age and to ultimately and quickly be lost or fixed in a population. Thus, these mutations under selective force would be younger than neutral alleles of the same frequency who reached that frequency due to random genetic drift. We observe in the UK10K sample of 7242 human genomes that derived mutations that have a low EP are younger on average than neutral mutations of the same frequency. This is expected if such alleles are deleterious. By the same reasoning, if derived alleles of high EP are under positive selection, then we expect these to also be younger on average. However, we find that evolutionary derived high EP alleles are older on average than neutral alleles of the same frequency. Our results combine allele age estimates with both evolutionary probability and frequency and reveal unexpected trends that indicate a more complex model of balancing selection on alleles of high evolutionary probability.

240W A heterogeneous landscape of selection and interactions in genes revealed by two-locus statistics Aaron Ragsdale University of Wisconsin-Madison, Madison, WI

Selection impacts patterns of genetic diversity over large regions of the genome. How mutations combine to affect individual fitnesses is actively studied using theoretical, simulation, and empirical approaches. In addition to distorting patterns of linked neutral variation (such as background selection and hitchhiking), selected mutations are known to interfere and interact with each other, affecting probabilities of fixation, allele frequency trajectories, and correlations between pairs of mutations. Linkage disequilibrium (LD) is particularly sensitive to interference and epistasis between selected mutations, and a number of recent studies have used patterns of LD to test for epistatic interactions, with some disagreement over interpreting observations from data.

Despite ongoing interest in learning about selective interactions from observed patterns of LD, there is a lack of analytic or even numerical approaches for expectations for patterns of variation between pairs of loci under the combined effects of selection, dominance, epistasis, and demography. Here, we develop a numerical approach to compute the two-locus sampling distribution under diploid selection with arbitrary dominance and epistasis, recombination, and variable population size. We use this to explore how epistasis and dominance affect expected patterns of LD, including for non-steady-state demography relevant to human populations. Using whole-genome sequencing data from humans, we find that selection strengths and interactions vary across protein-coding genes in a way that correlates with annotated domains and conserved genic elements. Observed positive LD between missense mutations within genes is largely driven by positive correlations between pairs of mutations that fall within the same conserved domain, pointing to compensatory or antagonistic epistatic effects in those domains. The heterogeneous landscape of both mutational fitness effects and selective interactions within protein-coding genes calls for more refined inferences of the joint distribution of fitness and interactive effects.

241W Strong, recent selective sweeps reshape genetic diversity in freshwater bivalve *Megalonaias nervosa* *Rebekah Rogers*[1], Stephanie Grizzard[2], Jeffrey Garner[3] 1) UNC Charlotte; 2) Old Dominon University ; 3) Alabama Department of Conservation

Freshwater Unionid bivalves have recently faced ecological upheaval through pollution, barriers to dispersal, human harvesting, and changes in fish-host prevalence. Currently, over 70% of species are threatened, endangered or extinct. To characterize the genetic response to these recent selective pressures, we collected population genetic data for one successful bivalve species, Megalonaias nervosa. We identify megabase sized regions that are nearly monomorphic across the population, a signal of strong, recent selection reshaping genetic diversity. These signatures of selection encompass a total of 73Mb, greater response to selection than is commonly seen in population genetic models. We observe 102 duplicate genes with high dN/dS on terminal branches among regions with sweeps, suggesting that gene duplication is a causative mechanism of recent adaptation in M. nervosa. Genes in sweeps reflect functional classes known to be important for Unionid survival, including anticoagulation genes important for fish host parasitization, detox genes, mitochondria management, and shell formation. We identify selective sweeps in regions with no known functional impacts, suggesting mechanisms of adaptation that deserve greater attention in future work on species survival. In contrast, polymorphic transposable element insertions appear to be detrimental and underrepresented among regions with sweeps. TE site frequency spectra are skewed toward singleton variants, and TEs among regions with sweeps are present only at low frequency. Our work suggests that duplicate genes are an essential source of genetic novelty that has helped this successful species succeed in environments where others have struggled. These results suggest that gene duplications deserve greater attention in non-model population genomics, especially in species that have recently faced sudden environmental challenges.

242T Genetic constraint of complex traits for drought adaptation in Arabidopsis *Megan Ruffley*[1], Laura Leventhal[1,2], Moises Exposito-Alonso[1,2] 1) Plant Biology, Carnegie Institution for Science; 2) Biology Department, Stanford University

Genetic correlations between traits are thought to be a main constraint to adaptation, as multiple advantageous trait strategies cannot evolve simultaneously. However, comprehensive studies of how genetic correlations persist in natural populations and their ecological relevance when subject to climate-driven natural selection are rare. Through common garden experiments with Arabidopsis thaliana, Exposito-Alonso et al. (2019) showed pervasive antagonistic selection between high and low rainfall environments. Mining and extending the Arapheno database for published ecologically-relevant traits, we quantify direct and total phenotypic selection (Arnold & Lande 1983) under these drought and non-drought environments. We find evidence that the observed antagonistic selection can be explained by a phenotypic trade-off of mutually-exclusive ecological strategies to climate stressors; commonly referred to as drought escape and drought avoidance (Ludlow 1989). In the low precipitation environment, strong total selection favored phenotypes relating to escaping drought by flowering early at the expense of low water use efficiency, a strategy persistent in the warm edge of the species' range, as it reduces mortality. However, when we control for multiple correlated traits related to escape and drought avoidance, we identified a direct positive selection over highwater-use efficiency, which is typically associated with late flowering ecotypes. This reversion in selection highlights that there must be a fitness trade-off between the two strategies. Genome-wide association approaches allowed us to confirm these traits are genetically correlated and, importantly, that hotspots of genetic correlation across these phenotypes map to corresponding regions of the genome that are under antagonistic selection. These results indicate that natural populations of A. thaliana experience a genetic constraint to drought adaptation and importantly, this should be accounted for when making predictions about how species will adapt to a changing climate.

243T Large-scale comparative population genetics identifies repeated targets of natural selection in birds Allison Shultz[1], Cade Mirchandani[2], Sara Wuitchik[3], Brian Arnold[4], Erik Enbody[2], Russell Corbett-Detig[2], *Timothy Sackton*[3] 1) Natural History Museum of Los Angeles County, Los Angeles, CA; 2) University of California, Santa Cruz, Santa Cruz, CA; 3) Harvard University, Cambridge, MA; 4) Princeton University, Princeton, NJ

In the past decade, increasingly large quantities of sequencing data from a diverse range of non-model organisms has been submitted to public databases, but reanalysis and reuse of this rich data has been difficult due to lack of infrastructure and computational batch effects. To facilitate large-scale comparative population genetics research, we have implemented snpArcher, an easy-to-use snakemake pipeline to generate variant calls using bwa/GATK, optimized for use in non-model organisms. Using snpArcher, we reanalyzed public resequencing datasets covering nearly 5000 individuals from over 100 species of non-mammalian vertebrates. This collection of analyzed datasets, all generated with consistent methods and filtering, is publicly available via Globus, as the "Comparative Population Genomics Collection." From this full collection, we selected a subset of species for which an annotated reference genome and appropriate outgroup resequencing data were available for additional analysis. For these species, we used MacDonald-Kreitman tests to identify protein-coding genes with evidence for recent positive selection. By comparing these targets of recent selection across species, we were able to identify a set of commonly adapting proteins. We highlight examples of repeatedly selected proteins, including a number with functions in the immune system, and discuss the implications for convergent molecular adaptation.

244W Allelic gene conversion frequently turns sweeps on single-origin *de novo* mutations into soft sweeps Daniel Schrider University of North Carolina, Chapel Hill, NC

The prominence of positive selection, in which beneficial mutations are favored by natural selection and rapidly increase in frequency, is a subject of intense debate. Positive selection can result in selective sweeps, in which the haplotype(s) bearing the adaptive allele "sweep" through the population, thereby removing much of the genetic diversity from the region surrounding the target of selection. Two models of selective sweeps have been proposed: classical sweeps, or "hard sweeps", in which a single copy of the adaptive allele sweeps to fixation, and "soft sweeps", in which multiple distinct copies of the adaptive allele leave descendants after the sweep. Soft sweeps can occur because of recurrent mutation to the adaptive allele, or the presence of standing genetic variation consisting of multiple copies of the adaptive allele prior to the onset of selection. Importantly, soft sweeps will occur when populations can rapidly adapt to novel selective pressures, either because of a high mutation rate or because adaptive alleles are already present. The prevalence of soft sweeps is especially controversial, and it has been noted that even when multiple copies of the adaptive allele are present at or near the onset of selection, only one copy may end up surviving the sweep. Thus, selection on standing variation or recurrent mutations may not always result in hard sweeps. Here, we show that the inverse is true: selection on single-origin *de novo* mutations may often result in soft sweeps. This is made possible by allelic gene conversion, which "softens" hard sweeps by copying the adaptive allele onto multiple genetic backgrounds. We carried out a simulation study examining the impact of gene conversion on sweeps from a single *de novo* variant in models of human, *Drosophila*, and *Arabidopsis* populations. The fraction of simulations in which gene conversion had produced multiple haplotypes with the adaptive allele upon fixation was appreciable. Indeed, under realistic demographic histories and gene conversion rates, soft sweeps are more likely than hard sweeps, even if selection always acts on a single-origin mutation. Thus, even when the mutation rate is low or there is no standing variation, soft sweeps may be common, and hard sweeps are expected to be the exception rather than the rule in species with high rates of gene conversion. These results also imply that the presence of soft sweeps does not necessarily mean that adaptation has been especially rapid.

245W Ancestry-associated selection signatures in Pacific Islanders *Jan Sokol*[1], Javier Blanco Portillo[2], Obed Garcia[2], Mark Penjueli[3], Alexander Ioannidis[2] 1) Boston University School of Medicine; 2) Stanford University; 3) New York University Abu Dhabi

Using genome-wide array data from over five thousand Hawaiian and Polynesian individuals, we identified several ancestry-specific selection patterns in Pacific Islanders. We analyzed these hits, particularly those within the HLA region, to characterize the putative variants driving this selection. Our findings corroborate recent research that had identified Native American gene flow into Polynesia around AD 1200, and expands upon its functional significance. We suggest that ancestry-specific selection was driven by immunologic challenges to Native Hawaiians and discuss implications for precision-medicine for present-day Hawaiians. Our analysis provides a crucial new piece of information in the genetic study of Polynesian individuals, helping to shift the focus towards populations for whom fewer medical genetic resources exist.

246T A new test of balancing selection and its application to data from humans *Vivak Soni*[1,2], Adam Eyre-Walker[2], Michiel Vos[3] 1) ASU; 2) University of Sussex; 3) University of Exeter

The role that balancing selection plays in the maintenance of genetic diversity remains unresolved. One approach is to consider polymorphisms that are shared between populations or species. However, a major problem with this approach is differentiating selectively maintained polymorphisms from neutral variation inherited from the common ancestor. This problem can be solved by comparing the number of shared polymorphisms at sites that are selected, to those that are neutral. Here we introduce a new test, based on the McDonald-Kreitman test, in which the number of polymorphisms that are shared between populations is contrasted to those that are private at selected and neutral sites. Through simulation we show that this simple test is robust to a variety of demographic changes, and that it can also give a direct estimate of the number of shared polymorphisms that are directly maintained by balancing selection. We apply our method to population genomic data from humans and provide some evidence that hundreds of non-synonymous polymorphisms are subject to balancing selection.

247T Signatures of positive and negative selection in the human gut microbiome *Richard Wolff*, Nandita Garud University of California, Los Angles

The human gut microbiome is a complex ecological community composed of a large number of microbial species which are genetically diverse both within and between hosts. These gut microbes experience selective pressures across multiple timescales, from evolution within individual hosts occurring in a matter of months to longer-term processes occurring over many host generations. But despite the considerable effects of the gut microbiome on human health, our understanding of the evolution of these populations—particularly over timescales greatly exceeding individual hosts' lifetimes—remains limited. Here, we leverage linkage disequilibrium statistics, which measure correlations between sites, to assess evidence for both positive and negative selection across hosts among a broad cohort of species inhabiting the human gut microbiome. With this approach, we find signatures of both purifying selection and adaptation in several gut microbial species. Our findings highlight the heterogeneity of evolutionary outcomes across species, as well as the critical role played by recombination in driving adaptation in these populations.

248W Inferring polygenic selection from GWAS summary statistics for multiple traits and populations *Alexander Xue*[1], Yi-Fei Huang[2], Adam Siepel[1] 1) Cold Spring Harbor Laboratory; 2) Pennsylvania State University

Current approaches for detecting selection from GWAS data are unable to directly estimate the distribution of fitness effects (DFE). To this end, we introduce ASSESS, an inferential method that exploits the Poisson Random Field (PRF) to model selection coefficients from genome-wide allele count data, while jointly conditioning GWAS summary statistics on a latent distribution of phenotypic effect sizes from genotypes. The likelihood function, which is unified under the assumption of an explicit relationship between fitness and trait effect, is optimized using an EM algorithm to yield a trait's DFE. To validate the performance of ASSESS, we conducted several simulation experiments under various data configurations, demographic scenarios, and genomic architectures. We find consistent behavior in accurately recovering the underlying selection history, as well as a high degree of robustness to a range of assumption violations of our conceptual framework. Additionally, we applied ASSESS to publicly available data for an array of human traits in both European and non-European populations. We discover a pattern of polygenicity estimates much higher than in previous investigations, which we attribute to the PRF's sensitivity to weaker selection coefficients. Our *in silico* demonstration as well as the empirical insight gained here illustrate the potential of ASSESS to satisfy an increasing need for powerful yet convenient population genomic inference from GWAS summary statistics.

249W Comparative genomics of *Aspergillus oryzae* genomes from different clades reveals signatures of artificial selection in primary and secondary metabolism in domesticate environments *Katherine Chacon-Vargas*, Colin O. McCarthy, Victoria Donescu, John Gibbons University of Massachusetts Amherst

Human selection of desired traits to enhance their benefits has led to domestication in plants, animals and also microbes (bacteria, yeasts, and molds). Domesticated microbes play an important role in food preservation, nutrition and flavors. *Aspergillus oryzae* is a domesticated filamentous fungal species used during the fermentation of traditional Asian foods and beverages such as sake, soy sauce, and miso. The artisanal practice of continuous passage of *A. oryzae* on food substrates over thousands of years has resulted in adaptation to the food environment along with genetic differentiation from its wild relative *A. flavus*, a toxin producing agricultural pest. Here, we analyzed 300 isolates of *A. oryzae* and *A. flavus* to understand the history of domestication and how this process shaped patterns of genomic variation. Using population structure and phylogenetic analysis we identified 2 major *A. oryzae* populations and two major *A. flavus* lineages. Next, we used two population genomic metrics to identify regions of the *A. oryzae* genome displaying signatures of recent positive selection. We identified 30 candidate selective sweeps, several of which contain genes with functional annotations directly related to fermentation (e.g. an alcohol dehydrogenase, fructose transmembrane transporters, and genes involved in glutathione metabolism). Additionally, we examined differences in gene copy number

variation between *A. oryzae* and *A. flavus*. We found major CN differences in chromosomes 3, 4 and 8 corresponding to to genes involved in primary and secondary metabolism. Most strikingly, we found significantly more copies of the α-amylase encoding genes in *A. oryzae* compared to *A. flavus*, suggesting selection for increased carbohydrate metabolism during domestication. Further, gene absences in *A. oryzae* compared to *A. flavus* were enriched for secondary metabolism function, suggesting selection for loss of toxicity in *A. oryzae*. Taken together, our results show the *A. oryzae* genome was significantly reshaped as a result of domestication to the food environment.

250T Evolution of modifiers of conformity *Kaleda Denton*[1], Yoav Ram[2,3,4], Uri Liberman[5], Marcus Feldman[1] 1) Department of Biology, Stanford University, USA; 2) School of Computer Science, Interdisciplinary Center Herzliya, Israel; 3) School of Zoology, Faculty of Life Sciences, Tel Aviv University, Israel; 4) Sagol School of Neuroscience, Tel Aviv University, Israel; 5) School of Mathematical Sciences, Tel Aviv University, Israel

Conformist transmission entails that the probability of adopting a more common cultural variant exceeds its frequency in the population, whereas anti-conformist transmission occurs if the reverse is true. Because the variant that an individual adopts can affect its biological fitness, there may be gene-culture co-evolution of such transmission biases. First, we re-visit a widely used model of conformity where offspring sample a number *n* of «role models,» which is often assumed to be three. If *n* >= 5, evolutionary dynamics can be more complex than previously assumed: stable frequency cycles or chaos can occur, and new polymorphic equilibria may exist. Second, we investigate the conditions under which a rare genetic modifier of the extent of conformity or number of role models can invade a population. Our results show that near a stable equilibrium with two variants present, the variant that confers higher fitness is more common, so a modifier of conformity invades the population if it decreases individuals' level of anti-conformity. Similarly, a modifier of *n* invades if the level of anti-conformity with the modified *n* is less than that with the resident *n*. Finally, we investigate the case of population subdivision with migration and find that the common claim that conformity can maintain between-group differences is not always true. Therefore, the effect of conformity on the evolution of cooperation by group selection may be more complicated than previously stated. Understanding the evolutionary dynamics under (anti-) conformity may have implications for research on human and nonhuman animal behavior, the evolution of cooperation, and frequency-dependent transmission in general.

251T Do forensic genetic markers compromise medical privacy? *Jhony Zavaleta*[1], Mayra Bañuelos[1,2], Alennie Roldan[1], Rochelle-Jan Reyes[1], Miguel Guardado[1,3], Berenice Chavez Rojas[1], Thet Nyein[1], Ana Rodriguez Vega[1], Maribel Santos[1], Emilia Huerta Sanchez[2], Rori Rohlfs[1] 1) San Francisco State University; 2) Brown University; 3) University of California San Francisco

The FBI's Combined DNA index system (CODIS) contains the genetic profiles of over twenty million people. These profiles are composed of Short Tandem Repeats (STRs) located within non-coding DNA – selected under the stipulation that they did not reveal medical information. However, thirteen of the twenty STR loci were selected before the human genome was sequenced. Recent studies indicate that variation in short tandem repeats may alter gene expression. While it is unknown if CODIS STRs do so as well, many CODIS STRs are very close to genes. In fact, eleven of the twenty CODIS STRs are intronic and others are extremely close to genes and regulatory elements. These nearby genes are implicated in a range of medical conditions including schizophrenia, depression, Perrault syndrome and MELAS syndrome. In this study, we used publicly available data to investigate the relationship between CODIS STRs and the expression levels of neighboring genes. We find that CODIS STRs resemble the genomic characteristics of published gene expression-modifying STRs. Furthermore, we identify five CODIS STRs as being associated with the expression of proximal genes in lymphoblastoid cell lines. Lastly, we explore possible mechanisms for these associations, identifying one of the loci as a potential causal locus and three loci as being in LD with a causal locus. Our results are consistent with the hypothesis that CODIS profiles may reveal gene expression information, thus bringing to question the practices regarding these data.

252W Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection, and mitigation *Runyang Lou*, Nina Overgaard Therkildsen Department of Natural Resources, Cornell University

Over the past decade, there has been an explosion in the amount of publicly available sequencing data. This opens new opportunities for combining data sets to achieve unprecedented sample sizes, spatial coverage, or temporal replication in population genomic studies. However, a common concern is that nonbiological differences between data sets may generate patterns of variation in the data that can confound real biological patterns, a problem known as batch effects. In this project, we compare two batches of low-coverage whole genome sequencing (lcWGS) data generated from the same populations of Atlantic cod (Gadus morhua). First, we show that with a "batch-effect-naive" bioinformatic pipeline, batch effects systematically biased our genetic diversity estimates, population structure inference, and selection scans. We then demonstrate that these batch effects resulted from multiple technical differences between our data sets, including

the sequencing chemistry (four-channel vs. two-channel), sequencing run, read type (single-end vs. paired-end), read length (125 vs. 150 bp), DNA degradation level (degraded vs. well preserved) and sequencing depth (0.8× vs. 0.3× on average). Lastly, we illustrate that a set of simple bioinformatic strategies (such as different read trimming and single nucleotide polymorphism filtering) can be used to detect batch effects in our data and substantially mitigate their impact. We conclude that combining data sets remains a powerful approach as long as batch effects are explicitly accounted for. We focus on lcWGS data in this project, which may be particularly vulnerable to certain causes of batch effects, but many of our conclusions also apply to other sequencing strategies.

253W Genetic ancestors of an admixed population *Lily Agranat-Tamir*, Jazlyn A Mooney, Noah A Rosenberg Stanford University, Stanford, CA

In a genetically admixed population, admixed individuals possess ancestry from the various source groups. Studies of genetic admixture in admixed human populations frequently estimate ancestry components corresponding to fractions of individual genomes that trace to specific ancestral populations. However, the same numerical ancestry fraction can represent a wide array of admixture scenarios. Using a mechanistic model of admixture, we characterize admixture genealogically: how many distinct ancestors from the source populations does the admixture represent? We consider the number of ancestors both in terms of *genealogical* ancestors of an admixed individual chosen at random and *genetic* ancestors, those genealogical ancestors who contribute to the genome of the modern admixed individual. We apply our mathematical results to the African-American population, estimating for individuals chosen at random from this population the numbers of genealogical and genetic ancestors originating from the source populations. The results provide insight both on how many of the genetic ancestors of a typical African-American might have been forcibly displaced to the Americas in the Transatlantic Slave Trade and on how many separate European admixture events might have contributed genetic ancestry to the genome of a typical African-American individual.

254T The Polynesian Settlement of the Hawaiian Archipelago *Javier Blanco-Portillo*[1], Mark Penjueli[2], Feiyang Liu[1], Jan Sokol[1], Charleston Chiang[7], Patrick Kirch[3], Christopher Gignoux[5], Marcus Feldman[1], Keolu Fox[6], Genevieve Wojcik[4], Alexander Ioannidis[1] 1) Stanford University, CA, USA; 2) NYU Abu Dhabi, Abu Dhabi, United Arab Emirates; 3) University of Hawai'i Mānoa, HI, USA; 4) Johns Hopkins University, MD, USA; 5) CU Anschutz, CO, USA; 6) UC San Diego, CA, USA; 7) University of Southern California, CA, USA

Available evidence confirms that the first inhabitants of Hawai'i trace their origins to the Austronesian-speaking founders of Eastern Polynesia. However, the precise identification of the proximate origin and timing of the settlement of Hawai'i are still matters of debate. Using ancestry-specific approaches to explore the genomic variation of modern Hawaiians and individuals from other Polynesian islands, we find evidence of a Tuamotuan origin for the inhabitants of Hawai'i in the 12th century, in contrast with previous theories that posited an origin in the Marquesas or Society Islands. We characterize the founder effects and population bottlenecks that resulted from this migratory process, and describe how they are vital for the development of precision medicine in modern Hawaiians.

255T Drivers of Diversity and Divergence in the Sea Rachel Toczydlowski[1], Reid Brennan[8], Eric Crandall[2], Joanna Kelley[3], Misha Matz[4], Jamie Pringle[5], Cynthia Riginos[6], John Wares[7], *Gideon Bradburd*[1] 1) Michigan State University; 2) Pennsylvania State University; 3) Washington State University; 4) University of Texas at Austin; 5) University of New Hampshire; 6) University of Queensland; 7) University of Georgia; 8) University of Vermont

What are the spatial patterns of genetic diversity within a species, and what drives those patterns? These questions touch on some of the greatest remaining mysteries in ecology and evolutionary biology, but, despite their centrality to fundamental questions in the field, we know little about what factors predict patterns of genetic diversity within a species; indeed, for most species, we have no estimates of genetic diversity at all. Life history traits, such as fecundity, dispersal ability, philopatry are predicted to impact both the overall genetic diversity of a species and its distribution across a species' range. Here, we test hypotheses about the relationship between life history traits and geographic patterns of genetic diversity. We take a data synthesis approach, aggregating publicly available, georeferenced, next-generation sequence datasets for over 100 species across the tree of life. We focus on marine species, for which spatial patterns of diversity are more poorly understood than in terrestrial systems. For each species, we use these data to estimate population genetic quantities, including Wright's neighborhood size and diversity in the collecting phase. We also aggregate biotic trait data for these species and abiotic geographic data for the areas in which they were sampled. Finally, we place these species on a time-calibrated phylogeny. Taken together, these data allow us to integrate across scales both temporally and spatially to test hypotheses about whether biotic and abiotic factors predict genomic diversity and divergence in marine ecosystems.

256W The role of pollinators in shaping plant population genetic structure *Grace Burgin*[1,2], Robin Hopkins[1,2] 1) Arnold Arboretum, Harvard University, Boston, MA; 2) Department of Organismic and Evolutionary Biology, Harvard University, Boston, MA

Plant-pollinator interactions have long been appreciated as a key driver of flowering plant evolution. Notably, much of the existing empirical evidence highlights pollinators as selective agents shaping floral trait adaptation and species diversification. We seek to expand our understanding of the evolutionary significance of plant-pollinator interactions by emphasizing how pollinators can impact plant evolution through control of gene flow via pollen dispersal. We use population genetic tools to test the hypothesis that since pollinators control gene dispersal via pollen flow, plant population genetic structure will reflect variation in pollinator movement. The relationship between the Texas wildflower, *Phlox drummondii*, and its primary pollinator, *Battus philenor*, provide an excellent system to explore this prediction. Field observations indicate *B. philenor* is the most frequent and effective visitor to *P. drummondii*, that no mechanism for long-distance seed dispersal exists, and that *B. philenor* moves broadly across their shared range. Therefore, movement of *B. philenor* is likely a significant factor driving dispersal both within and between populations of *P. drummondii*. To test this expectation, we generated genome-wide molecular markers using restriction-site associated DNA sequencing and characterized patterns of genetic variation in both species. By sampling co-occurring populations of plant and pollinator, we infer patterns of pollinator movement and test pollinator impact on genetic connectivity of plant populations in a spatially explicit context. Our work aims to generate new understanding of the role of pollinators in plant evolution as well as to characterize how pollinator loss under changing climate might impact plant persistence.

257W Population structure and historical demography of a reptile species that has evolved insular dwarfism on the California Channel Islands. *Amanda Clark*, Tonia Schwartz Auburn University

The “island rule” is the worldwide phenomenon of rapid evolution of dwarfism and gigantism on islands. It is unknown to what degree convergence in insular body size is accompanied by convergence at the level of genetics, physiology, and/or life-history traits. We have identified three reptile species—two snakes and one lizard—on the California Channel Islands that have independently evolved insular dwarfism relative to mainland southern California. Island populations of all three reptiles have smaller body size, smaller relative head size, and lower blood glucose relative to mainland populations, though with some variation by sex and year. We further this investigation at the genomic level, presenting a chromosome-level genome assembly for the gopher snake (*Pituophis catenifer*) and whole genome population sequencing across two mainland and two island locations to characterize the genomic divergence across the landscape and specifically in candidate functional genetic pathways. Our data support a single colonization event, with population-level divergence (F^{ST}) estimates between mainland and island sites. Additional results on historical demography of California gopher snakes will be presented and discussed in the context of our ultimate goal—to understand the mechanisms of convergence across the ecology, life-history, physiology, and genetic networks.

258T Genomic diversity and invasion history of *Drosophila suzukii* *Siyuan Feng*[1], Samuel DeGrey[2], Sean Schoville[2], John Pool[1] 1) Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; 2) Department of Entomology, University of Wisconsin-Madison, Madison, WI

Biological invasions are of great research interest as they often carry significant economic and ecological costs, but also constitute natural experiments that allow investigations of evolutionary processes on contemporary timescales. The fruit pest *Drosophila suzukii*, which has rapidly invaded the globe within the past few decades, stands out as an excellent model for studying invasion genomics and local adaptation owing to its occupation of distinct environments. However, despite the recent availability of genomic resources in *D. suzukii*, inference of the invasion route has only been made based on limited microsatellite markers. Here, we investigated genomic diversity and invasion history of *D. suzukii* using whole- genome sequencing data from 29 population samples from four distant continents and three islands. Strong founder event bottlenecks were suggested by the acute drop in nucleotide diversity of invasive populations relative to observed diversity in the native range (including China and Japan), with effectively longer bottlenecks on the X chromosome due to its smaller effective population size. Principal component analysis of allele frequency and matrices of window F_{ST} and D_{XY} recapitulated the expected clustering of distinct native, European and American populations. All other populations were found to have a subset of the genetic diversity present in a sample from southeastern China (Ningbo), consistent with an ancestral or refugial species range in this region of Asia. A hierarchical population structure predicted by maximum-likelihood tree supported that inference, and also suggested separate Asia-sourced invasion events into America and Europe. Further tree-based admixture inference predicted gene flow among continents following the first founder events. To detect genomic targets of local adaptations under specific environmental pressure, we are now performing genome-wide environment association analysis between SNPs and selected uncorrelated environmental

variables at sampled locations. Our findings provide insights into the population history of *D. suzukii*, and have the potential to further our understanding of the genetic architecture underlying this species' invasion success.

259T The genomic status and evolutionary history of *Culex pipiens* mosquito ecotypes *Yuki Haba*[1], Noah Rose[1], Molly Schumer[2], Lindy McBride[1] 1) Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ; 2) Department of Biology, Stanford University, Stanford, CA

The Northern House Mosquito *Culex pipiens sensu stricto* is the most important disease vector mosquito in temperate zones across the northern hemisphere, responsible for the emergence of West Nile Virus and filarial worms over the last few decades. There are two ecologically distinct yet morphologically indistinguishable forms. An aboveground form *pipiens* diapauses in winter and primarily targets birds, while a belowground form *molestus* thrives year round in belowground habitats such as subways or basements, and prefers mammals over birds. Despite their abundance and importance, we know surprisingly little about the genetic and evolutionary history of these two *Cx. pipiens* ecotypes. The origin of the belowground *molestus* is particularly contentious, with iconic populations from the London Underground metro system being held up by evolutionary biologists as a proof-of-principle example of rapid, in situ, urban adaptation and speciation. Furthermore, the two forms hybridize in some zones of contact, which further complicates inferences of evolutionary history and public health efforts. To better understand the genomic status and evolutionary history of the ecotypes, we launched the Culex pipiens Population Genomics Project (PipPop). Here, by sequencing the genomes of 800+ individual mosquitoes collected across the entire global distribution of the species, we provide the first clear picture of genomic variation in this group and infer the timing and geography of divergence within the species. We find that aboveground populations form a latitudinal hybrid gradient stretching from what are known to be traditional *pipiens* populations in northern Europe down to *molestus*-like aboveground populations in the Middle East. We also find that belowground *molestus* from all over the world are genetically most similar to Middle Eastern populations. Together, our data reject the popular in situ evolution hypothesis and instead suggest that underground populations represent recent migrants from the Middle Eastern populations that evolved the anthropophilic ecology over the course of millennia rather than centuries. We test alternative demographic models using the genomic data. Lastly, using the worldwide panel of genomes, we build an eco-geographic model to understand the current and future distributions of the cosmopolitan mosquito.

260W Population genomics of large white-footed mice in the Boston Harbor archipelago *Emma Howell*[1], Peicheng Jing[1], Lauren Nolfo-Clements[2], Bret Payseur[1] 1) Genetics Department, University of Wisconsin-Madison, Madison, WI; 2) Biology Department, Suffolk University, Boston, MA.

Island populations often exhibit departures from their mainland counterparts in key behavioral, morphological, and life-history traits– a phenomenon termed the "island syndrome". Populations of white-footed mice (*Peromyscus leucopus*) inhabiting two islands within the Boston Harbor archipelago provide compelling examples of this rule, measuring between 40 and 50% larger in body weight than mainland mice. In contrast to the ecological and geographical conditions often thought to drive island syndrome phenotypes, the Boston Harbor islands are situated close to the mainland and harbor both predators and interspecific competitors of *P. leucopus*. In addition, these islands have undergone dramatic ecological transformations over the past 400 years owing to a history of human usage that spans agriculture, industry, and recreation. Yet, despite such physical proximity and human-mediated connectivity to the mainland, island populations maintain a significant difference in body size. To understand how such phenotypic differences can persist under these seemingly permeable barriers to gene flow, we performed whole-genome sequencing of wild-caught individuals sampled from Peddocks and Bumpkin Island together with mainland representatives from World's End. Preliminary analyses of genome pairs sampled from each location suggest lower per-site nucleotide diversity in the island populations (Peddocks = 0.00573; Bumpkin = 0.00614) compared to the mainland (World's End = 0.00762), consistent with reduced effective population sizes on the islands, and differences in the extent of genetic differentiation between the populations (World's End vs. Bumpkin Fst = 0.079; World's End vs. Peddocks Fst = 0.107; Peddocks vs. Bumpkin Fst = 0.189). We extend this analysis to larger sample sizes and use the frequency spectra of putatively neutrally evolving variants to reconstruct the colonization history of the islands and estimate migration rates among the island and mainland populations. In the future, we will leverage the wealth of information encoded within whole genome sequences to examine how measures of genetic diversity and differentiation compare across coding, regulatory, and evolutionarily conserved regions of the genome. Together, these genetic comparisons of island and mainland mice will establish important aspects of their demographic history that will be key to interpreting their stark phenotypic differences and identifying the genomic regions that maintain them.

261W Uncovering natural histories of mutator alleles in budding yeast *Pengyao Jiang*[1], Vidha Sudhesh[1], Anja Ollodart[1],

Alan Herr[1], Maitreya Dunham[1], Kelley Harris[1,2] 1) University of Washington; 2) Fred Hutchinson Cancer Research Center

Mutations provide essential raw material for evolutionary change. Genetic factors that increase mutation rates, i.e. mutator alleles, have been observed to arise during experimental evolution, facilitating adaptation to the lab environment. However, little is known about how prevalent mutator alleles are in natural populations and how they have historically contributed to evolution due to the rarity of mutations under normal conditions. To tackle this challenging question, we have established a framework that utilizes the mutation spectrum—the relative frequencies of different types of mutations, calculated from natural polymorphisms—to determine potential historical impacts of *Saccharomyces cerevisiae* mutation rate modifiers that affect certain mutation types disproportionally. We combined this with efficiently accumulating *de novo* mutations in a reporter gene using a modified fluctuation assay of natural isolates to identify *S. cerevisiae* populations that have experienced recent mutation rate and spectrum changes. We discovered a 10-fold range of mutation rate variation among 16 haploid strains from diverse populations. Two strains from the Mosaic beer clade have excess C>A mutations in both *de novo* and rare natural polymorphisms, indicating a recent occurrence of at least one mutator allele common to the two strains. We further identified a mutator allele in *OGG1* that partially explains these strains' mutator phenotype. Our mutation spectrum analysis of polymorphisms also indicates that additional mutators have likely been influencing the evolution of natural populations beyond the Mosaic beer strains. We discovered that strains from the African beer population are even more conspicuous mutation spectrum outliers, and a subset of French dairy strains show intermediate mutation spectra between the African beer and the rest of the strains. Analysis of these French dairy strains suggests that there are likely mutator alleles introgressed from African beer strains affecting the subsequent mutation spectra. We have engineered a reporter into the African beer and French dairy strains for the modified fluctuation assay, and we are measuring *de novo* mutation spectra in these strains to test this hypothesis. In summary, our framework has proven useful in identifying mutator alleles in natural populations of budding yeast, and will reveal how mutator alleles contribute to evolution.

262T Whole genome sequences of 3,000 individuals from India: Insights into South Asian Population History and Disease *Elise Kerdoncuff*[1,2], Laurits Skov[1,2], Wei Zhao[3], Jennifer Smith[3,4], Andrea Ganna[5], Sharon Kardia[3], Aparajit B. Dey[6], Jinkook Lee[7], Priya Moorjani[1,2] 1) Department of Molecular and Cell Biology, University of California, Berkeley, United States of America; 2) Center for Computational Biology, University of California, Berkeley, United States of America; 3) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, United States of America; 4) Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, United States of America; 5) Institute for Molecular Medicine Finland, Helsinki, Finland; 6) Department of Geriatric Medicine, All India Institute of Medical Sciences, New Delhi, India; 7) Department of Economics, University of Southern California, United States of America

Previous genotyping surveys of India have shown that most present-day Indians have ancestry from two divergent ancestral populations: Ancestral North Indian (ANI) related to Central Asians and Iranians, and Ancestral South Indian (ASI) distantly related to the Andamanese group, Onge. This mixture is widespread with most groups deriving ~20-80% ASI ancestry in India. Ancient DNA data from Central Asia and South Asia has further shown that both ANI and ASI were in turn admixed and had ancestry from ancient groups of South Asian hunter-gatherers, Neolithic Iranian-farmers, and Steppe Pastoralists from Eurasia. These mixtures occurred in the past 8,000 years and are associated with the spread of Neolithic farming and Indo-European languages to South Asia. Following the admixture, India experienced a major demographic shift towards endogamy leading to strong founder events more extreme than those in Ashkenazi Jews and Finns, both of which have high rates of recessive disease due to founder events. In this study, we generated ~3,000 whole genome sequences from India. Our study includes diverse ethno-linguistic groups from India, including samples from most geographic regions, speakers of all major language families and tribal and caste groups—providing a comprehensive coverage of genetic variation in India. Using these samples, we provide a detailed reconstruction of the history of India in the past 50,000 years. We examine the distribution of Neanderthal and Denisovan ancestry across regions and in relationship to other worldwide populations. We also study recent events providing new insights about the ANI and ASI mixture and its impact on disease and adaptation on the subcontinent. Finally, we investigate the signatures of founder events in India and uncover key population-specific variants associated to diseases. Together, these analyses provide a detailed view of South Asian history and disease.

263T Effects of isolation by distance on principal components *Lesly Lopez*, Jordan Collignon, Suzanne Sindi, Emily Jane McTavish University of California, Merced

Principal component analysis (PCA) is a popular dimension-reduction technique to summarize patterns of population structure. PCA is regularly applied to genetic data to investigate genetic variation. The expectation of projections on the

principal components can estimated through pairwise coalescent times. We examine how isolation by distance affects coalescent times between samples and, consequently, the estimates of these expected pairwise coalescent times from the projections of the principal components. To simulate genetic variation caused by geographic distance we modified the original Wright Fisher model to violate the assumption of a panmictic population. We use a distance-weighted Wright Fisher model with a modified probability for the location of the parent of each haploid individual.

264W Inference of the demographic history of commensal gut microbes *Jonathan Mah*, Kirk Lohmueller, Nandita Garud University of California, Los Angeles

Human commensal gut microbes play a crucial role in host health, including aiding with the digestion of foods that humans cannot digest themselves. Despite the importance of such microbes to human health, there is little knowledge about the evolutionary history of commensal gut microbes, including their demographic histories and how selective forces shape their genetic variation. In this study, we infer the demographic history for the 27 most highly prevalent commensal gut microbial species in North Americans. Several of these species show evidence of population contractions coincident with the onset of agricultural expansion approximately 10,000 years ago. These results are consistent with reductions in diversity observed at the species and genetic level in commensal microbes sampled from Western populations relative to non-Western rural populations with diets consisting of higher amounts of fiber. However, our present results contrast with the population expansions observed in a similar timeframe by the cavity-causing oral microbe, *Streptococcus mutans*. Taken together, we infer that changes in diet over the course of human history have resulted in effective population size contractions of gut microbiota.

265W The evolutionary history and adaptive divergence of *Daphnia pulex* *Connor Murray*[1], Joaquin Nunez[2], Aakrosh Ratan[3], Alan Bergland[4] 1) University of Virginia; 2) University of Virginia; 3) University of Virginia; 4) University of Virginia

Understanding the mechanisms that drive divergence within and between populations is key to elucidating the genetic basis of local adaptation and speciation. Investigating recently diverged species can help elucidate how disparate regions of the genome accumulate change through distinct selection regimes. Here, we investigate the evolutionary history and adaptive divergence of the *Daphnia pulex* clade, a group of freshwater crustaceans that experiences yearly demographic turnover across the growing season and inhabits the Holarctic. Despite the extensive use of *D. pulex* in ecological and evolutionary genomics, there is a fundamental knowledge gap in the distribution of genetic diversity and divergence history across the species clade. Moreover, we still don't know the extent of genomic divergence between continental *D. pulex* that has led to debate regarding the current species classification. To address these issues, we have assembled a genomic panel of over 1600 North American and European samples from *D. pulex*, *D. pulicaria*, and *D. obtusa* from publicly available data and wild sequenced clones. Our results reveal extensive divergence across the genome of continental *D. pulex* consistent with highly diverged species. Additionally, we show genomic evidence for the distribution of a hybrid complex of North American *D. pulex* and *D. pulicaria* clone across European ponds, highlighting the potential of hybridized lineages to colonize a large habitat range. This work highlights over 30 million years of evolution within the *D. pulex* clade and reveals the putatively adaptive structural and functional gene categories that have resulted in highly disparate genetic diversity across the species range.

266T Drivers of dispersal and genetic variation for bee species in a fragmented tropical habitat *Sevan Suni*, Melissa Hernandez University of San Francisco

Quantifying genetic structure and levels of genetic variation within populations are of fundamental importance to biologists seeking to predict the ability of populations to persist in human-altered landscapes and adapt to future environmental changes. Genetic structure reflects the dispersal of individuals over many generations, which can be constrained by environmental factors or mediated by species-level traits such as body size. Dispersal distances are commonly positively associated with body size and negatively associated with the amount of degraded habitat between habitat fragments, motivating investigation of these two potential drivers of dispersal concomitantly. We quantified genetic structure and levels of genetic variability within populations of nine bee species within the tribe Euglossini. Euglossine bees are important pollinators of over 700 orchid species and many other tropical plants. We tested the following predictions: (1) deforested areas restrict dispersal, (2) forested paths among sites are better predictors of dispersal than Euclidian geographic distances, (3) there is a positive association of dispersal distance and body size, (4) genetic variability is greater in sites surrounded by more intact habitat. We used RADseq and the Stacks pipeline to genotype bees at thousands of SNP loci and estimate population genetic parameters. Body size was a strong predictor of genetic structure, but, surprisingly, larger species showed higher genetic structure than smaller species. The way that deforestation affected genetic structure was not mediated body size, and there was no effect of deforestation or geographic distance on dispersal. There was variability across species in the way that the amount of forest surrounding sites affected genetic variability. These results

challenge the dominant paradigm that individuals of larger species disperse farther, and we discuss potential ecological drivers and implications of these results.

267T Accumulation of hybrid incompatibilities on scale-free networks with purifying selection *Evgeny Brud*, Rafael Guerrero North Carolina State University, Raleigh, NC

Reproductive isolation between divergent populations is often the result of deleterious epistatic interactions between their genetic differences. Not all fixed differences contribute equally to reproductive isolation, however, since loci across the genome vary in their number of gene interactions. We consider how this variation in connectivity affects the accumulation of hybrid incompatibilities between allopatric populations, using computer simulations of substitutions on a scale-free network in which highly connected genes were assigned lower substitution rates (mimicking the effect of selective constraint on hub genes). These results are compared to: (1) uniform substitution rates and (2) varied substitution rates that are random with respect to connectivity. We find that greater variation in negative selection slows the rate of incompatibility accumulation and delays the complexification process by which incompatibilities agglomerate on the gene network.

268W Human-Mediated Admixture in South American Neotropical Cats *Maximilian Genetti*[1], Fernanda de Jesus Trindade[2], Eduardo Eizirik[2], Russell Corbett-Detig[1] 1) University of California Santa Cruz, Santa Cruz, California; 2) Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Admixture is a phenomenon where divergent populations interbred, mixing their genomes, and it is increasingly appreciated as a central process in evolution. The biological phenomenon is thought to drive an array of evolutionary outcomes including shaping patterns of genetic variation, introducing novel adaptive variation, and introducing deleterious mutations. However, when this process is driven by human activity, threatening the genomic integrity of one or more parental species, it becomes a focus of conservation. Increased attention towards the Neotropical cats of the Leopardus genus was prompted after recent studies demonstrated extensive admixture among this species complex and the identification of a new species. In particular, interest in L. geoffroyi and L. guttulus was prompted after genetic studies identified extensive hybridization at their geographic contact zone in southern Brazil, both of which are threatened by human-mediated habitat fragmentation and destruction. This geographic contact zone is where two distinct biomes intersect, Pampas and Atlantic Forest, each of which is associated with one of the two hybridizing species. To understand the age and structure of this hybrid zone, we are using a low-coverage (~1x) whole-genome sequencing approach across ~200 individuals that span a ~2500 km range including the hybrid zone and both parental ranges. This data includes sequencing from 27 L. geoffroyi and 30 L. guttulus, with the remaining individuals representing a gradient of ancestry proportions. We have also sequenced the complete genomes from one pure individual per species at ~20x and ~40x depth. With this data we have identified introgression of genomic regions between species, and distinct signatures of admixture in the autosomes relative to the X chromosome. L. geoffroyi is the major ancestry in ~70% of the admixed individuals and expands southward from the hybrid zone towards the parental distribution of this species. The semi-arid Pampas ecoregion dominating this species range has experienced increased habitat alteration in recent decades. These results are indicative of human-mediated hybridization, and highlight the need for continued study of this complex hybrid zone.

269W The genetic, organismal, and evolutionary origin of color pattern diversity in Phyllobates poison-dart frogs Roberto Márquez University of Michigan

A wide array of organisms use conspicuous signals to warn predators of secondary defenses in order to avoid predation, a strategy known as aposematism. Poison-dart frogs are an ideal system to study this phenomenon, since aposematism has evolved multiple times from cryptic ancestors. Although the ecological and behavioral processes guiding the evolution of aposematism in poison frogs are relatively well known, the underlying evolutionary genetic and biogeographic mechanisms remain largely unknown. In this talk I explore the possible evolutionary genetic mechanisms behind the convergent evolution of bright coloration in Phyllobates poison-dart frogs from population genetic, biogeographic and developmental perspectives. Our results suggest that conspicuous coloration has evolved convergently three times within the past 3 million years through similar developmental mechanisms, yet at least two of these events appear to be underlied by change at different genes. Finally, preliminary work hints that some of these genes may also underlie variation in other traits associated with coloration, such as body size and toxicity. Overall, poison-dart frogs represent a promising emerging model system for the study of integrative evolutionary biology.

270T Genomic islands of speciation have independent rates of molecular evolution across a butterfly hybrid zone *Tianzhu Xiong*[1], Xueyan Li[2], Masaya Yago[3], James Mallet[1] 1) Harvard University, Cambridge, MA, USA; 2) Kunming Institute of Zoology, Chinese Academy of Science, Kunming, China; 3) The University Museum, The University of Tokyo,

Tokyo, Japan

Hybridization is a major evolutionary force eroding genetic differentiation between species, whereas reproductive isolation maintains such differentiation. In studying a hybrid zone between the swallowtail butterflies *Papilio syfanius* and *Papilio maackii*, we made the unexpected discovery that genomic substitution rates are unequal between the parental species. This phenomenon creates a novel process in hybridization, where genomic regions most affected by gene flow evolve at a similar rate across species, while genomic regions with greater reproductive isolation evolve at divergent rates. Thus, hybridization mixes evolutionary rates in a way similar to genetic admixture. Using coalescent theory, we show that the rate-mixing process provides distinct information about levels of gene flow across different parts of genomes, and the maintenance of divergent substitution rates can be predicted quantitatively from relative sequence divergence (F_{ST}) between the hybridizing species at equilibrium. A corollary is that divergent rates will be maintained in regions linked to barrier loci. Overall, we demonstrate that reproductive isolation maintains not only the final outcome of genomic differentiation, but also the rate at which differentiation accumulates. This new information also suggests that the separation of evolutionary rates co-localizes with the separation of gene pools between genomes of incipient species.

271T Tensor decomposition-based feature extraction and classification to detect natural selection from genomic data. *Md Ruhul Amin*, Mahamudul Hasan, Michael DeGiorgio Florida Atlantic University

Inferences of adaptive events are important for learning about traits, such as human digestion of lactose after infancy, the ability of organisms to survive at extreme environments such as high altitudes, and the rapid spread of viral variants. Early efforts toward identifying footprints of natural selection from genomic data involved development of summary statistic and likelihood methods. However, such techniques are typically grounded in simple theoretical models that may limit the complexity of settings that they can explore, running the risk of inaccurate predictions as the summary statistics are hand engineered. Due to the renaissance in artificial intelligence, machine and deep learning methods have taken center stage in recent efforts to detect natural selection, with strategies such as convolutional neural networks applied to images of haplotypes across sampled individuals to simultaneously extract important genomic features and achieve high classification accuracy and power for distinguishing selection from neutrality. Yet, limitations of such techniques include difficulty in estimating the number of model parameters and identification of features without regard to their location within an image. As a complementary approach, we consider an alternative feature extraction method, termed tensor decomposition, which falls within a class of dimensionality reduction techniques to extract features from multidimensional data while preserving the latent structure of the data. We apply tensor decomposition to images of haplotypes across sampled individuals, and then use these extracted features as input to classical linear and non-linear machine learning methods. As a proof of concept, we explore the performance of this pipeline on simulated neutral and selective sweep scenarios, and find that it has high power and accuracy to discriminate sweeps from neutrality, robustness to missing data, and easy visualization of underlying low-dimensional features uncovered by tensor decomposition. Therefore, our approach is a powerful addition to the toolkit for detecting adaptive processes from genomic data.

272W Expanding the Use of Generative Adversarial Networks in Population Genetics to Create Artificial Sequence Alignments *William Booker*, Dan Schrider University of North Carolina at Chapel Hill, Chapel Hill, NC

Research over the last decade has demonstrated that machine learning methods have significant utility in the field of population genetics. Among the numerous potential machine learning tools available for use, generative networks such as Generative Adversarial Networks (GANs) have recently been used in population genetics to generate artificial human genomes and estimate the demographic parameters of populations. The use of GANs in these contexts are largely a modification of the original GAN architecture initially developed to generate fake images such as human faces, but expansions on these networks demonstrate they can be incredibly powerful for a variety of tasks—many of which can be applied to current problems in population genetics and evolution. Here, we further explore the utility of GANs in population genetic research. Broadly, we make the first attempt to use Generative Adversarial Networks in which a neural network generates population genomic sequence alignments—yielding a set of genomic sequences sampled from the same population rather than a single sequence—that resemble those created from sequencing data or simulations. Using modifications from the original GAN architecture including Deep Convolutional and Wasserstein GANs, we demonstrate consistent training success across runs without the need for extensive hyperparameter tuning. We then demonstrate the ability of this network to generate alignments under several demographic scenarios that retain properties of the input alignments relevant to population genetics (e.g. the site frequency spectrum, linkage disequilibrium decay with physical distance, etc.). Overall, this work expands upon the applicability of GANs to population genetics and underlies a framework for significant expansion for the future.

273W Exploiting Genetic Variation to Model Localised Homing Gene Drives *Benjamin Camm*[1,2], Alexandre Fournier-Level[1] 1) University of Melbourne, Parkville, Victoria, Australia; 2) Commonwealth Scientific and Industrial Research Organisation (CSIRO), Victoria, Australia

Gene drives are powerful genetic tools that have the potential to affect entire species. It is vital to design self-limiting gene drives to minimise the chance of unintentional spread, or persistence. Here, we use a genomically-informed gene drive model to demonstrate how natural genetic variation between populations can be exploited to design homing gene drives that are spatially or temporally limited.
In weedy *Lolium rigidum* (Annual Ryegrass) populations from South-Eastern Australia, we found 111 loci with variation that could be exploited to create a localised gene drive. Of these, 17 had more than one polymorphism that met our pairwise allele frequency difference threshold between three populations. The locus with the strongest pairwise allele difference between the three populations was selected to be used in the model. A multi-fasta for that locus was fed into our discrete non-overlapping model where the allele frequencies per population and their respective conversion efficiencies were derived. A slight difference in allele frequency between populations allowed control of the gene drive. The weighted average conversion efficiency of the target population was 0.556, with the off-target populations having 0.535 and 0.417. The model showed that the drive was able to fixate in the target population (>0.997) while in the off-target populations, the drive peaked at frequencies of 0.315 and 0.0502.
A localised gene drive simulation was achieved by using sequence information to infer the allele frequencies and conversion efficiencies. This modelling allows for population specific gene drive modelling, building confidence in the predicted outcomes.

274T From pattern to function: eco-evolutionary representations of complex spatial structure for the new era of spatial biology Oana Carja Carnegie Mellon University

Through innovations in both imaging techniques and the ability to process these images at scale, high-resolution imaging is transforming the field of molecular biology, yet its power has yet to be fully utilized for asking questions in evolutionary biology. Just as demographic surveys can reveal more or less densely populated areas where, for example, a contagious disease may spread at different rates, these imaging datasets can help us quantify cellular and molecular patterns of spatial variation and understand how this variation affects rates of evolution, by impeding or accelerating the spread of new variants through the population. What are spatial topologies that act to amplify the selective advantage of new mutations in the population, versus structures that dampen the force of selection and slow down rates of evolution?

While classic models in population genetic theory have been extraordinarily important for producing initial testable predictions about the role of space and structure in evolution, most previous modeling approaches represent spatial structure as a small number of distinct patches, symmetrically connected by migration corridors. This makes these models analytically tractable, but it also makes their predictions hard to use for the wealth of these emerging spatially-resolved datasets with large amounts of spatial heterogeneity and complex patterns of cellular co-localization and interaction. The challenge is that studying more complex spatial topologies and deriving intuitive analytic results with predictive power is a much harder mathematical problem. Solving this problem lies in finding the right spatial geometric representations that can both capture the complexity of spatial structure in emerging datasets, but also allow for mathematical modeling tractability, in other words finding the mathematical representations that inform on the spatial characteristics driving functional, evolutionary design.

I will discuss how we can build a general theory of evolutionary dynamics for populations with complex spatial structure. Using tools from network theory and algebraic topology, I will present how we can derive the relevant selective parameters as a function of the statistical properties of the population spatial structure. I will also present several relevant applications of our theory, including recent work where we have used recent microscopy datasets to build the cellular spatial networks of the stem cell niches of the bone marrow and to ask whether the spatial arrangement of these cellular collectives acts to amplify or suppress the spread of variants in the cellular population.

275T The effect of consanguinity on X-chromosomal and autosomal genomic sharing *Daniel Cotter*[1], Alissa Severson[1], Shai Carmi[2], Noah Rosenberg[3] 1) Depatment of Genetics, Stanford University, Stanford, CA; 2) Braun School of Public Health and Community Medicine, Hebrew University of Jerusalem, Ein Kerem, Israel; 3) Department of Biology, Stanford University, Stanford, CA

By providing additional opportunities for coalescence within families, the presence of consanguineous unions in a population reduces coalescence times relative to non-consanguineous populations. First-cousin consanguinity can take one of

four forms differing in the sexes of the siblings whose offspring join in a consanguineous union: patrilateral parallel, patrilateral cross, matrilateral parallel, and matrilateral cross. Considering populations with each of the four types of consanguinity individually, as well as a population with a mixture of the four types, we examine coalescent models of consanguinity. We previously computed, for first-cousin consanguinity models, the mean coalescence time for X-chromosomal loci and the limiting distribution of coalescence times for autosomal loci. Here, we use the separation-of-time-scales approach to obtain the limiting distribution of coalescence times for X-chromosomal loci. This limiting distribution has an instantaneous coalescence probability that depends on the probability that a union is consanguineous; lineages that do not coalesce instantaneously coalesce according to an exponential distribution. The computations of coalescence time distributions are useful for understanding features of runs of homozygosity (ROH) and identity by descent (IBD): as coalescence times decrease in a population, ROH and IBD segments increase in length. We apply our calculations for X-chromosomal coalescence times alongside the analogous calculation on the autosomes to develop theory for the expected fractions of the X chromosome and of the autosomal genome that lie within ROH and IBD, expressing these fractions as functions of the rate and type of consanguinity. The results can inform the understanding of haplotype sharing patterns in highly consanguineous populations.

276W Efficient analysis of allele frequency variation from whole-genome pool-sequencing data *Lucas Czech*[1], Yunru Peng[1], Jeffrey Spence[2], Patricia Lang[3], Tatiana Bellagio[1,3], Julia Hildebrandt[4], Katrin Fritschi[4], Rebecca Schwab[4], Beth Rowan[4], Detlef Weigel[4], J.F. Scheepens[5], François Vasseur[3,6], Moises Exposito-Alonso[1,3,4,7], GrENE-net consortium 1) Department of Plant Biology, Carnegie Institution for Science, Stanford, USA; 2) Department of Genetics, Stanford University, Stanford, USA; 3) Department of Biology, Stanford University, Stanford, USA; 4) Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany; 5) Faculty of Biological Sciences, Goethe University, Frankfurt, Germany; 6) Centre d'Écologie Fonctionnelle et Évolutive (CEFE), University of Montpellier, Montpellier, France; 7) Department of Global Ecology, Carnegie Institution for Science, Stanford, USA

In recent decades, so-called Evolve-and-Resequence (E&R) experiments have become a popular approach to survey rapid evolution of populations over multiple generations. These experiments allow us to measure shifts in the allele frequencies of a population in response to new or shifting environmental conditions, such as a changing climate.

Pool-sequencing of several individuals at once is a cost-effective and efficient tool to obtain reliable allele frequencies from a population of thousands to hundreds of thousands of individuals, and is often used in E&R experiments. However, specialized tools to efficiently analyze these data that take sampling biases stemming from the pool-sequencing approach into account were lacking. We developed two software tools to overcome statistical and bioinformatic challenges arising in this context.

First, we present grenepipe, a workflow from raw sequencing data of individuals or pooled populations to genotypes (variant calling) and population allele frequencies. The pipeline automates trimming, mapping, variant calling, and quality control, with a selection of popular software tools in each of these steps, and produces variant calls and frequency tables. While generally applicable to individual sample data, it offers specialized steps for pool-sequencing. With a single command line call, our software downloads all dependencies and runs all steps automatically, parallelizes processing for computer cluster environments, and recovers from any failing steps.

Second, to enable inferences of evolutionary signatures from frequency data, we created grenedalf, a C++ command line tool to compute population genetic statistics. It computes unbiased statistics of Fst, Pi, Tajima's D with pool-sequencing data, far outperforming alternative tools. Further it offers novel data exploration tools such as windowed allele frequency spectrum visualizations and PCA and MDS on the allele frequencies, and built-in data filters and manipulations.

These tools are designed for scalability and ease-of-use with contemporary file formats, which we showcase using the GrENE-net.org project, a large-scale Evolve-and-Resequence experiment with Arabidopsis thaliana from across the world.

277W Neural ADMIXTURE: rapid population clustering with autoencoders *Albert Dominguez Mantes*[1,2], Daniel Mas Montserrat[1], Xavier Giró-i-Nieto[2], Carlos Bustamante[1], Alexander Ioannidis[1] 1) Stanford University, Stanford, United States; 2) Universitat Politècnica de Catalunya, Barcelona, Spain

Characterizing the genetic substructure of large cohorts has become increasingly important as genetic association studies are extended to massive, increasingly diverse, biobanks. ADMIXTURE and STRUCTURE are widely used unsupervised clustering algorithms for characterizing such ancestral genetic structure. These methods decompose individual genomes into fractional cluster assignments with each cluster representing a vector of DNA marker frequencies. The assignments,

and clusters, provide an interpretable representation for geneticists to describe population substructure at the sample level. However, with the rapidly increasing size of population biobanks and the growing numbers of variants genotyped (or sequenced) per sample, such traditional methods become computationally intractable. Multiple runs with different hyperparameters are required to properly depict population clustering using these traditional methods further increasing the computational burden, leading to days of compute. In this work we present Neural ADMIXTURE, a neural network autoencoder that follows the same modeling assumptions as the current standard algorithm, ADMIXTURE, providing similar (or better) clustering, while reducing the compute time by orders of magnitude. Indeed, the equivalent of one month of continuous compute using the current standard algorithm (ADMIXTURE) can be reduced to just hours with Neural ADMIXTURE. In addition, by using a multi-head approach Neural ADMIXTURE can include multiple clustering outputs, providing results equivalent to running standard algorithms many times with different numbers of clusters. Our models can also be stored, allowing later cluster assignment on new data to be performed with a linear computational time and without needing to share the training data. The software implementation of Neural ADMIXTURE can be found at https://github.com/ai-sandbox/neural-admixture.

278T Evolution of Evolvability In Rapidly Adapting Asexual Populations *James Ferrare*, Benjamin Good Stanford University

Mutations can affect both the immediate fitness of an organism and the rates and benefits of future mutations. While numerous examples of these evolvability modifiers have been identified, their favorability has been difficult to predict in rapidly evolving asexual populations, where large numbers of mutations compete simultaneously. Here we calculate the fixation probability of a mutation that modifies the rates and benefits of future mutations, demonstrating how this quantity depends on parameters like the population size and amount of concurrent genetic diversity in the population. We show that competition between linked mutations can exponentially influence the fixation probability of modifiers to the magnitude of fitness effects of beneficial mutations, providing a comparatively larger contribution to the fixation probability than modifiers to the rate of beneficial mutations. We find that evolvability modifiers that increase or decrease the magnitude of fitness effects of beneficial mutations respectively suppress or amplify an associated direct fitness cost. By considering deleterious mutations, we also demonstrate that modifiers to the magnitude of fitness effects of beneficial mutations can be exponentially favored to fix in a population even when they reduce the rate of adaptation by orders of magnitude.

279T An EM algorithm for detecting general diploid selection in time series allele frequency data *Adam Fine*[1], Matthias Steinruecken[2,3] 1) Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL; 2) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Chicago, Chicago, IL

Detecting selection and quantifying its strength is a fundamental problem in evolutionary biology, with applications ranging from finding mutations critical to early hominid evolution to identifying genes responsible for rapid adaptation in experimental evolution. Since selection is a multi-generational process, many approaches for quantifying selection use genetic data sampled from two or more time points. A common class of models for analyzing such time series data are Hidden Markov models (HMMs), in which the underlying population allele frequency trajectory is modelled as a Wright-Fisher or Moran process, and the sampled allele frequencies are the observed variables. Mathieson and McVean (2013) introduced an expectation-maximization (EM) algorithm based on a maximum likelihood estimator for the haploid selection coefficient derived by Watterson (1982) that they use to efficiently estimate selection coefficients under genic selection. Here, we introduce DIESELFUEL (DIploid Estimates of SELection Forces Using Expectation-maximization and Lagrange multipliers), a novel method for estimating selection coefficients under general diploid selection scenarios. To this end, we derive a diploid version of Watterson's maximum likelihood estimators and extend the EM algorithm of Mathieson and McVean to infer general diploid selection. Moreover, we introduce an approach based on Lagrange multipliers to efficiently explore subspaces of the full parameter space. We use this approach to introduce a new framework for choosing the best model across both one-parameter (genic, dominance, recessive, over-/under-dominance) and full diploid selection scenarios. We also explore introducing a simplified version of Mathieson and McVean's spatially structured population model, in which all migration rates are identical. On simulated data, we find that DIESELFUEL accurately and efficiently estimates selection coefficients and chooses the correct mode of selection across a range of parameter values and scenarios. Lastly, we apply DIESELFUEL to the human lactase and MHC loci and characterize selection of these loci more generally than previously possible to showcase the potential of our method.

280W Emergent evolutionary forces in spatial models of microbial growth in the human gut microbiota *Olivia Ghosh*, Benjamin Good Stanford University, Stanford, CA

The genetic composition of the gut microbiota is constantly reshaped by ecological and evolutionary forces. These strain-level dynamics can be challenging to understand because they emerge from complex spatial growth processes that take place within a host. We introduce a general population genetic framework to predict how stochastic evolutionary forces emerge from simple models of microbial growth in spatially extended environments like the intestinal lumen. Our framework shows how fluid flow and longitudinal variation in growth rate combine to shape the frequencies of genetic variants in sequenced fecal samples, yielding analytical expressions for the effective generation times, selection coefficients, and rates of genetic drift. We find that the emergent evolutionary dynamics can often be captured by well-mixed models that lack explicit spatial structure, even when there is substantial spatial variation in species-level composition. By applying these results to the human colon, we find that continuous fluid flow and simple forms of wall growth are unlikely to create sufficient bottlenecks to allow large fluctuations in mutant frequencies within a host. We also find that the effective generation times may be significantly shorter than expected from traditional average growth rate estimates. Our results provide a starting point for quantifying genetic turnover in spatially extended settings like the gut microbiota, and may be relevant for other microbial ecosystems where unidirectional fluid flow plays an important role.

281W SLiM 4: Multispecies eco-evolutionary modeling *Benjamin Haller*, Philipp Messer Cornell University, Ithaca, NY

SLiM, a free open-source software framework for forward genetic simulation, is widely used because of its power, flexibility, and speed. Version 2.0, released in 2016, added the Eidos scripting language, the SLiMgui interactive modeling environment, and many other new features. Version 3.0, released in 2018, added support for non-Wright-Fisher models and tree-sequence recording. We are now at version 3.7, and there's a lot of new stuff to talk about since 3.0! This poster will present some of the major advances since version 3.0, including: (1) Nucleotide-based models including sequence-dependent mutation rates, biased gene conversion, and more; (2) The addRecombinant() method and other improvements for simulating different mating systems, from clonal bacteria to haplodiploids to alternation of generations; (3) mutation() callbacks, for influencing the generation of new mutations; (4) survival() callbacks, for influencing mortality (and migration); (5) New Eidos features including the Image, Dictionary, DataFrame, and LogFile classes, and lots of new utility functions; (6) The new cross-platform SLiMgui for macOS, Linux, and Windows, with some exciting new features; (7) Lots of new example "recipes" in the manual; and (8) A new free online SLiM workshop that will teach you how to use SLiM, SLiMgui, and Eidos from the ground up! We will also foreshadow some of our plans for version 4 of SLiM.

282T Digital image processing using alpha-molecules to detect selective sweeps *Mahmudul Hasan*, Md Ruhul Amin, Michael DeGiorgio Florida Atlantic University

In recent years, advances in image processing and machine learning have fueled a paradigm shift in detecting genomic regions under natural selection. Early machine learning techniques employed population-genetic summary statistics as features, which focus on specific genomic patterns expected by adaptive and neutral processes. Though such engineered features are important when training data are limited, the ease at which synthetic data can be generated in modern times has led to the recent development of approaches that take in images of haplotype alignments and automatically extract important features with deep learning techniques. Alpha-molecules are a class of techniques for multi- scale representation of objects that can extract a diverse set of features from images. One such method, termed wavelet decomposition, lends greater control over high-frequency components of images. Another method, curvelet decomposition, is an extension of the wavelet concept that considers events occurring along curves within images. We show that application of these alpha-molecule techniques to extract features from images of haplotypes yields high power and accuracy to detect selective sweep signatures from genomic data with both linear and non-linear machine learning classifiers. Moreover, we find that such models are easy to visualize and interpret, with performance rivaling those of contemporary deep learning approaches for detecting selective sweeps.

283T Statistical inference in population genomics *Parul Johri*[1], Charles Aquadro[2], Mark Beaumont[3], Brian Charlesworth[4], Laurent Excoffier[5], Adam Eyre-Walker[6], Peter Keightley[4], Michael Lynch[1], Gil McVean[7], Bret Payseur[8], Susanne Pfeifer[1], Wolfgang Stephan[9], Jeffrey Jensen[1] 1) School of Life Sciences, Arizona State University, Tempe, AZ, US; 2) Department of Molecular Biology and Genetics, Cornell University, Ithaca, US; 3) School of Biological Sciences, University of Bristol, Bristol, UK; 4) Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK; 5) Institute of Ecology and Evolution, University of Berne, Berne, CH; 6) School of Life Sciences, University of Sussex, Brighton, UK; 7) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 8) Laboratory of Genetics, University of Wisconsin-Madison, Madison, US; 9) Leibniz-Institute for Evolution and Biodiversity Science, Berlin, DE

The field of population genomics has grown rapidly with the recent advent of affordable, large scale sequencing technologies. As opposed to the situation during the majority of the 20th century, in which the development of theoretical and

statistical population-genetic insights out-paced the generation of data to which they could be applied, genomic data are now being produced at a far greater rate than they can be meaningfully analyzed and interpreted. With this wealth of data has come a tendency to focus on fitting specific (and often rather idiosyncratic) models to data, at the expense of a careful exploration of the range of possible underlying evolutionary processes. For example, the approach of directly investigating models of adaptive evolution in each newly sequenced population or species often neglects the fact that a thorough characterization of ubiquitous non-adaptive processes is a prerequisite for accurate inference. We describe the perils of these tendencies and demonstrate how multiple incorrect models can often be fit equally well to population-genomic data. We also demonstrate how confounders such as background selection effects, unmodelled population history, mutation and recombination rate heterogeneity, SNP ascertainment bias, and progeny skew, when not accounted for, can bias the inference of both selection and demography. Thereby, we argue for the importance of defining a biologically relevant baseline model tuned to the details of each new analysis, of skepticism and scrutiny in interpreting model-fitting results, and of carefully defining addressable hypotheses and underlying uncertainties. Finally, I would like to close by addressing current best practices in population genomic data analysis and highlighting areas of statistical inference and theory that are in need of further attention.

284W Modeling alignment cost in mixed-membership unsupervised genetic clustering *Xiran Liu*[1], Naama Kopelman[2], Noah Rosenberg[1] 1) Stanford University, Stanford, CA; 2) Holon Institute of Technology, Holon, Israel

Mixed-membership unsupervised clustering is widely used to extract informative patterns from data in many application areas. In population genetics, unsupervised clustering methods such as ADMIXTURE and STRUCTURE have been widely used to infer population structure and ancestry proportions from genetic data. For a shared data set, clustering results produced by different algorithms, or even multiple runs of the same algorithm, can be difficult to compute, as outcomes can differ owing to permutation of the cluster labels, meaningful differences in clustering results, or both. Here, we study the cost of misalignment of mixed-membership unsupervised clustering replicates under a theoretical model of cluster memberships. Using Dirichlet distributions to model membership coefficient vectors, we provide theoretical results quantifying the alignment cost as a function of the Dirichlet parameters and the Hamming permutation difference between replicates. For fixed Dirichlet parameters, the alignment cost is seen to increase with the Hamming distance between permutations. Data sets with low variance across individuals of membership coefficients for specific clusters generally produce high misalignment costs---so that a single optimal permutation has far lower cost than suboptimal permutations. Higher variability in data, as represented by greater variance of membership coefficients, generally results in alignment costs that are similar between the optimal permutation and suboptimal permutations. We demonstrate the application of the theoretical results to data simulated under the Dirichlet model, as well as to membership estimates from inference of human-genetic ancestry. The results can contribute to improving cluster alignment algorithms that seek to find optimal permutations of replicates.

285W Robust supervised machine learning for population genetic inference with domain adaptation *Ziyi Mo*[1, 2], Adam Siepel[2] 1) Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

A series of supervised machine learning methods have recently been proposed to address a range of problems in population genetics. Despite their much-improved performance over traditional statistical methods, model mis-specification remains the Achilles' heel of this new paradigm. Here, we propose that domain adaptation techniques can be a powerful tool to mitigate the effect of model mis-specification on the performance of these methods. We demonstrated that the Deep Reconstruction-Classification Network (DRCN) helped our previously proposed SIA model achieve a better performance in identifying selective sweeps when the parameters used for simulating training data mismatch those underlying the generation of the data at test time. We anticipate that this approach will be widely applicable and can be an important tool to build confidence in the results when adopting supervised machine learning methods for inference.

286T FSTruct: An F_{ST}-based tool for quantifying ancestry variability *Maike Morrison*[1], Nicolas Alcala[2], Noah Rosenberg[1] 1) Stanford University; 2) International Agency for Research on Cancer

How variable are the ancestry vectors in a population structure plot? In other words, how much variation in estimated membership coefficient vectors exists across individuals in a population? Population structure inference methods such as STRUCTURE and ADMIXTURE are among the most widely used tools in modern population genetics, yet the variability of a population's estimated ancestry is often determined only visually. We present a new method and R package, FSTruct, which computes a normalized variability statistic that equals 0 when the population's ancestry is minimally variable and 1 when it is maximally variable. This property means that the method can be used to compare the ancestry variabilities of multiple populations, even populations with different numbers of individuals or substantially different mean ances-

try vectors. The method is based on the population-genetic statistic F_{ST}, relying on an analogy between the variability of ancestry vectors across individuals and the variability of allele frequency vectors across populations. We demonstrate the method in a theoretical model of cluster memberships and in comparisons involving admixed populations, ancient genomics, and multiple data types, and we also introduce a bootstrap test for equivalence of two or more populations in their levels of ancestry variability. The R package is available on GitHub at github.com/MaikeMorrison/FSTruct.

287T The impact of sexually antagonistic selection on polygenic traits *Pavitra Muralidhar*, Graham Coop University of California-Davis

Genetic conflict between the sexes – sexual antagonism – is predicted to occur in all species with distinct males and females. Sexual antagonism arises because males and females will often have different fitness optima for many phenotypes, but are hindered in evolving to these optima because they share the majority of their genome. A 'tug-of-war' results, in which genetic variants that are beneficial in one sex but detrimental in the other may be held at intermediate equilibrium frequencies, instead of spreading throughout the population or being eliminated. Sexual antagonism could thus be a powerful force maintaining genetic and phenotypic variation in a population, but we know little about how this form of selection acts on traits with complex genetic architectures.

Here, we investigate how sexually antagonistic selection acts on polygenic traits – traits encoded by many loci of varying effect size scattered throughout the genome. Using a combination of mathematical modelling and whole-genome simulations, we examine the scenario where a polygenic trait is under stabilizing selection for different fitness optima in males and females, and characterize the mutation-selection dynamics for this case. We characterize genetic correlations between the sexes across the allele frequency spectrum, and find that sexually antagonistic selection can generate the transient appearance of sex-specific control of a polygenic trait. Finally, we study how sexually antagonistic selection manifests in genomic regions with sex-biased transmission (such as sex chromosomes), and suggest new approaches to detecting sexually antagonistic selection in population genomic data.

288W Insights into *D. melanogaster* and *D. simulans* transcriptome evolution and complexity using transcript distance (*TranD*) *Adalena Nanni*[1,2], James Titus-McQuillan[3], Oleksandr Moskalenko[4], Francisco Pardo-Palacios[5], Sarah Signor[6], Srna Vlaho[7], Zihao Liu[1,2], Ana Conesa[2,5,8], Rebekah Rogers[3], Lauren McIntyre[1,2] 1) Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL; 2) University of Florida Genetics Institute, University of Florida, Gainesville, FL; 3) University of North Carolina Department of Bioinformatics, Charlotte, NC; 4) University of Florida Research Computing, University of Florida, Gainesville, FL; 5) Dept. of Applied Statistics and Operational Research, and Quality, Polytechnical University of Valencia, Spain; 6) Department of Biological Sciences, North Dakota State University, Fargo, ND; 7) Department of Biological Sciences, University of Southern California, Los Angeles, CA; 8) Institute for Integrative Systems Biology, Spanish National Research Council (CSIC), Paterna, Spain

Alternative splicing is a critical component of evolution. As species branches deepen, tracing transcript orthologs is complex. Yet, comparing transcripts between and within species is an important first step toward understanding questions about how evolution of transcript structure changes between species and contributes to sub-functionalization. These questions are complicated by issues of data quality and availability, with the amount and quality of data differing widely among species and between tissues. The recent explosion of affordable long-read sequencing of mRNA has enabled the study of transcriptional variation. There is a clear need for additional straightforward, reproducible metrics that compare transcripts. In addition to total transcript length, structural phenotypes (intron retention, donor/acceptor variation, alternative exon cassettes, alternative 5' and 3' UTRs) and nucleotide-level distance metrics to compare transcripts are suggested and then implemented in *TranD,* a PyPi package and available in the open-source web-based Galaxy platform. We use *TranD* to enumerate variation in long reads compared to the annotation and to compare methods for estimating transcripts from long reads (FLAIR and IsoSeq3). Additional illustrations of the utility of this approach are given in comparisons of competing annotation and in identifying isoform variation between male and female *D. melanogaster*.

289W Inferring mechanisms of population-wide phenotypic shifts in longitudinal single-cell RNA-sequencing experiments *Chibuikem Nwizu*[1], Ava Soleimany[2], Lorin Crawford[1,2] 1) Brown University Providence, RI; 2) Microsoft Research New England, Cambridge, MA

The recent innovations in single-cell sequencing technology have allowed researchers to uncover biological mechanisms that drive phenotypic variation between both healthy and disease populations. These data can be collected at different time points, enabling the profiling of pertinent disease and developmental processes under various micro-environmental perturbations. Many statistical methods have been developed for modeling time-series data with the key assumption that repeated measurements are taken from the same unit over different time points. However, this assumption breaks

down in single-cell analyses due to the practical inability of sequencing the same unit (cell) twice. In this project, we explore a novel strategy to overcome this limitation by taking a population genetic perspective on longitudinal single-cell RNA-sequencing (scRNA-seq) studies. Methodologically, we present a novel dynamic, Bayesian hierarchical Dirichlet process model (SCoOP) where, rather than making statistical inferences about individual cell trajectories, we seek to model the trajectories of phenotypic clusters that are present among the profiled cellular populations. By adopting this approach, we can both identify evolving cell population-specific signatures of disease and capture significant mechanistic shifts. To demonstrate the utility of our approach, we study a real data set of time-series scRNA-seq collected from patient-derived organoids of pancreatic ductal adenocarcinoma. Here, we highlight our ability to observe and estimate the strength of phenotypic selection on cellular clusters over time as a response to micro-environmental perturbations.

290T Under low dispersal, local competition can cause populations in continuous space to divide into discrete clusters *Gilia Patterson*, Peter Ralph University of Oregon, Eugene, OR

Explicitly simulating individuals, their locations in 2-d space, and their interactions is useful for studying how populations and genes are distributed across the landscape. But, these simulations can lead to surprising results. A continuous population spontaneously arranges itself into a grid of isolated clusters solely through negative density-dependent reproduction when dispersal distance is less than the interaction distance for density dependence. This is because individuals located between clusters must interact with the individuals in all of the surrounding clusters, whereas individuals in the middle of a cluster only interact with the individuals in that cluster. With low dispersal, individuals rarely disperse into the area between the clusters and the pattern is maintained. I explore the consequences of this pattern for isolation-by-distance.

291T Leveraging Ancestral and Derived Allele Sharing to Infer the Admixture Proportion *David Peede*[1, 2], Diego Ortega-Del Vecchyo[3], Emilia Huerta-Sánchez[1, 2] 1) Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI; 2) Center for Computational Molecular Biology, Brown University, Providence, RI; 3) Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Juriquilla, Querétaro, México

The evolution of species has traditionally been viewed as a bifurcating process. However, the recent influx of genomic studies has challenged this notion, demonstrating that hybridization is not only common in nature but is also a powerful force in shaping the evolution of genomes and patterns of genetic variation. Introgression's role in the evolution of species has most often been inferred from whole-genome sequence data using summary statistics that measure the ratio of discordant phylogenetic relationships between gene trees and the assumed species tree by using site pattern frequencies as a proxy for gene tree frequencies. These methods mostly focus on measuring levels of shared derived alleles between the recipient and donor lineages to quantify the amount of introgression in the genome—also known as the admixture proportion. Introgression, however, also re-introduces ancestral alleles, and not including sites where the recipient and donor lineages share ancestral alleles leaves out sites that can be informative of introgression.
Here, we leverage both ancestral and derived allele sharing to 1) derive the analytical expectations of all possible topologies for a three-taxon species tree and subsequently derive the analytical expectations for existing estimators of the admixture proportion f_{HOM}, d_f, and D_p as a function of the mutation rate, effective population size, divergence times, the timing of introgression, and the admixture proportion; 2) define new summary statistics for quantifying the admixture proportion that leverage both ancestral and derived allele sharing and; 3) perform a simulation study where we vary the direction, timing, and amount of introgression, to assess if incorporating both patterns of derived and ancestral allele sharing improves estimates of the admixture proportion. We show that quantifying the admixture proportion is context-dependent, but in the absence of no *a priori* information about the direction, timing, or amount of introgression, our new estimator of the admixture proportion (P_D)—which considers both patterns of derived and ancestral allele sharing—performs the best 57% of the time across our 180 different demographic scenarios.

292W A geometric relationship of F_2, F_3 and F_4-statistics with Principal Component Analysis Benjamin Peter MPI Evolutionary Anthropology

Principal Component Analysis (PCA) and F-statistics *sensu* Patterson are two of the most widely used population genetic tools to study human genetic variation. Here, I derive explicit connections between the two approaches and show that these two methods are closely related. F-statistics have a simple geometrical interpretation in the context of PCA, and orthogonal projections are a key concept to establish this link. I show that negative F_3 corresponds to a circle on a PCA plot, and that F_4 is related to an angle measurement. I illustrate my results on two examples, one of Western Eurasian, and one of global human diversity. In both examples, I find that the first few PCs are sufficient to approximate most F-statistics, and that PCA-plots are effective at predicting F-statistics. My results extend F-statistics to more continuous

population models, moving towards a more complete descriptions of human genetic variation.

293W Simulating neutral genetic diversity in *P. vivax* parasite populations *Krista Pipho*, Shyamalika Gopalan, Jillian Grassia, Amy Goldberg Duke University, Durham NC

Plasmodium vivax disease monitoring and control is increasingly guided by analysis of genetic data. This process is hindered by poor understanding of the expected neutral genetic variation in *Plasmodium*. The parasite's complex lifecycle and particular epidemiology limit the utility of common population-genetic models. Here, we develop a whole-genome simulation of *P. vivax* that incorporates multiple realistic aspects of its lifecycle and epidemiology in a single framework. Using SLiM forward genetic simulations, we include a whole-genome simulation across 14 chromosomes incorporating the recurring bottlenecks and population expansions of the parasite's life history, as well as asexual reproduction within hosts and sexual reproduction within vectors. A stochastic Ross-MacDonald model determines the transmission of infections between vectors and hosts. Under this model, we investigate the effect of changes in parameters such as transmission intensity and population size on the genetic diversity of parasite populations. We find that the severity of bottlenecks during transmission from mosquito to human has a greater impact on measures of variation (such as pairwise diversity or the site frequency spectrum) than the carrying capacity of parasites within a human host. This result is consistent with population-genetic theory, which holds that the historical minimum of a population's size disproportionately impacts the effective population size and genetic diversity. Moving forward, our model can be used to predict the impact of proposed disease interventions by making relevant changes to model parameters, or to infer population history. More generally, this work improves our understanding of the processes shaping neutral genetic diversity in *P. vivax*, which in turn helps inform interpretation of parasite relatedness and response to selective pressures.

294T Location, location, location: Dissecting errors in machine learning prediction of geography *Clara Rehmann*, CJ Battey, Peter Ralph, Andrew Kern Institute of Ecology and Evolution, University of Oregon, Eugene, OR

The geographic history of a population is encoded within its genetic variation. Recently, we introduced Locator, a deep learning-based method for individual-level prediction of geographic location based on genotypic variation (Battey et al. 2020), that we demonstrated to be efficient and accurate using both simulated and empirical datasets. Further, when applied to empirical datasets, Locator's residuals appear to reflect known patterns of geographic ancestry: in human populations, prediction errors correspond to known instances of migration, and in *Anopheles* and *Plasmodium* populations, predictions are potentially biased towards corridors of gene flow.
In order to assess how ancestral migration patterns are reflected in prediction errors, we explore the use of Locator to predict the locations of individuals in simulated populations undergoing increasing degrees of anisotropic dispersal. Our results confirm that residuals from Locator predictions align along the axes of biased dispersal, and we demonstrate a relationship between error magnitude, dispersal distance, and degree of dispersal bias. Additionally, we investigate the potential for spatial imbalance in the training set to bias predictions towards densely-sampled areas of the landscape and offer solutions to reduce this overfitting for empirical applications. Finally, we show that the magnitude and direction of errors in geographic prediction are strongly correlated between *Anopheles* and *Plasmodium* datasets, suggesting that we are capturing a coupled migration pattern of host and parasite.

295T Medea elements are on ancient haplotypes but not for the reason you'd think Matthew Rockman New York University

Medea alleles act maternally to kill embryos that don't inherit them. Recent work has shown that polymorphic Medea loci, long known from flour beetles, are particularly abundant in the genomes of *Caenorhabditis* nematode species whose mating systems involve high rates of selfing. In several cases, loci harbor multiple antagonistic Medea alleles, such that each allele kills embyros homozygous for the alternate allele. This superficially looks like overdominance, and the elements occur on ancient haplotypes, suggestive of balancing selection. At the same time, these species have genomes that are mosaics of hyper-polymorphic and nearly monomorphic regions, and it's unclear whether and how Medeas contribute to this broader pattern of heterogeneous genetic diversity.
I derived analytical results for dynamical models of antagonistic Medea elements and their paternal-effect counterparts, Peel elements. The evolutionary behavior of these alleles is profoundly affected by partial selfing, and by details of the mating system (e.g., monoecy vs androdioecy). The major finding is that partial selfing generates positive frequency dependence, such that a weakly penetrant allele that is at high frequency in a population can prevent invasion and displacement by a much stronger allele. The positive frequency dependence precludes overdominance as an explanation for the ancient haplotypes as there are no stable internal allele frequency equilibria for a single population. Instead, ancient haplotypes in *Caenorhabditis* genomes may result from abundant weak Medeas that act as barriers to gene flow among populations. Antagonistic Medeas effectively behave as local adaptation loci, locally fixed and resisting homogenization

in the face of gene flow, despite the lack of phenotypic benefits for their bearers.

296W Fast Multinomial Clustering of multiallelic genotypes to infer genetic population structure Arun Sethuraman San Diego State University

Identifying population structure from multilocus genotype data is key to downstream population genetic analyses in a variety of fields, including conservation, evolutionary genetics, Genome Wide Association Studies (GWAS), and pedigree reconstruction for quantitative genetics. Several methods have been proposed to estimate population structure, but issues with speed of computation, reproducibility, and accuracy of estimation remain, particularly with Bayesian MCMC based methods to perform inference on the `admixture' model. Here I develop a likelihood based approach to infer population structure under the admixture model while handling polyploid, multi-allelic (e.g. SNP, STR, allozyme) loci to infer genetic admixture proportions and ancestral allele frequencies. I present three separate algorithms to perform inference - (1) Expectation Maximization, (2) Block Relaxation, and (3) Quasi Newton and SQUAREM acceleration which are implemented into the MULTICLUST framework. Comparative analyses with both simulated and empirical data with MULTICLUST and STRUCTURE indicate considerable improvements in time of computation for comparable inference, fast, reproducible, and accurately estimated parameters.

297W The contribution of admixture, selection, and genetic drift to allele frequency change in time series genomic data. *Alexis Simon*[1,2], Graham Coop[1,2] 1) Department of Evolution and Ecology, University of California, Davis; 2) Center for Population Biology, University of California, Davis

Genomic time series offer us a chance to observe the interplay of various evolutionary forces, such datasets are increasingly common in both experimental evolution studies and ancient DNA datasets. The total variance in allele frequency change between two time points can be decomposed into the contributions of admixture, genetic drift, and linked selection. One promising use of such datasets is to obtain a genome-wide view of linked selection. In closed populations, the contribution of linked selection is identifiable because it creates positive covariances between time intervals, and genetic drift does not. However, natural populations are more complex than closed laboratory experiments and can be subject to another evolutionary force: migration. Indeed, repeated admixture between populations can also produce positive covariances between allele frequency changes in different time intervals. Here we lay out how to decompose the total variance in allele frequency change due to drift, linked selection, and admixture. We show how to accurately estimate the fraction of variance in allele frequency change due to these processes in the focal population experiencing gene flow from other source populations. Our approach should be widely applicable to the growing number of temporal population datasets. We apply these methods to ancient DNA datasets from Europe to characterize the contribution of these processes to allele frequency over thousands of years.

298T Inferring spatial population genetic parameters using deep learning *Chris Smith*, Peter Ralph, Andrew Kern University of Oregon

Most organisms disperse a limited distance from their birth location. This limited dispersal shapes patterns of genetic variation over a landscape, such that individuals that are close to one another in space will be more closely related than those that are spatially distant. This pattern, sometimes called isolation by distance, in turn can be used to infer spatial population genetic parameters. Here we demonstrate how deep neural networks can be used in combination with geographically-referenced genotype data to estimate a critical spatial population genetic parameter, σ, the mean per-generation dispersal distance. Using extensive simulation, we show that our deep learning approach is competitive with or outperforms state-of-the-art methods, particularly at small sample sizes (e.g., n=10). Whereas competing methods depend on accurate identification of identity-by-descent tracts or information about local population density as input, our method uses only single nucleotide variants and spatial coordinates as input. These features make our method, which we call *dspp*, a potentially valuable new tool for estimating dispersal in non-model systems. We demonstrate *dspp* on two such datasets with publicly available data, *Anopheles gambiae* and *Arabidopsis lyrata*. Finally, we consider the future utility of deep learning methods for spatial population genetic inference.

299T Inferring demographic history from allele frequency spectra with multi-layer perceptron regressors *Linh Tran*, Connie Sun, Mathews Sajan, Ryan Gutenkunst University of Arizona, Tucson, AZ

Previously, our group had developed dadi, a software for inferring demographic history using the diffusion approximation and composite likelihood. Inferring demography with dadi requires considerable understanding of the software and can be computationally expensive. In this work, we aim to improve the ease of use and lower the computational burden for dadi users with supervised machine learning. For each dadi-supported demographic model, we use dadi to simulate the expected allele frequency spectrum (AFS) under different demographic parameter values and train the scikit-learn

Multi-layer Perceptron Regressor (MLPR) algorithm to infer these parameters from input AFS. We demonstrate that the trained MLPRs can infer the population-size-change parameters very well (ρ≈0.98) and other parameters such as migration rate and time of demographic event fairly well (ρ≈0.6-0.7). The trained MLPRs also make good predictions when tested on AFS generated by the msprime simulator, which includes linkage in its simulations. Importantly, our trained MLPRs provide parameter predictions instantaneously from input AFS, with accuracy comparable to parameters inferred by dadi's likelihood optimization while bypassing its long and computationally intensive evaluation process. We also implement an accompanying method for quantifying the uncertainty of the point estimates output by the trained regressors, using a scikit-learn-compatible package, MAPIE (Model Agnostic Prediction Interval Estimator). We show that this method provides much better coverage for all demographic parameters tested compared to traditional bootstrapping.

300W Timesweeper: Detecting positive selection using genomic time series *Logan Whitehouse*, Dan Schrider University of North Carolina at Chapel Hill

Despite decades of research, identifying selective sweeps, the genomic footprints of positive selection, remains a core problem in population genetics. Of the myriad methods that have been made towards this goal, few have are designed to take advantage of the potential of genomic time-series data. This is because in most population genetic analyses only a single, brief period of time can be sampled for a study. Recent advancements in sequencing technology, including improvements in extracting and sequencing ancient DNA, have made repeated samplings of a population possible, allowing for serialized genomic datasets. With these advances in mind, here we present Timesweeper, a fast convolutional neural network-based tool for identifying selective sweeps in data consisting of multiple genomic samplings of a population over time. Timesweeper utilizes this serialized sampling of a population by first simulating data under a demographic model appropriate for the data of interest, training a 1D Convolutional Neural Network on said model, and inferring which SNPs in this serialized dataset were the direct target of a completed or ongoing selective sweep. We show that Timesweeper is powerful under multiple simulated demographic and sampling scenarios, works with low coverage anthropological data, identifies selected variants with impressive resolution, and has high tolerance for demographic model misspecification. In sum, we show that dramatically more accurate inferences about natural selection are possible when genomic time-series data are available, as will increasingly be the case in coming years due to both the sequencing of ancient samples and repeated samplings of extant populations with faster generation times. Methods like Timesweeper thus have the potential to help resolve the controversy over the role of positive selection in the genome. We provide Timesweeper both as a Python package and Snakemake workflow for use by the community.

301V Weak pleiotropic effects for natural variation in *Drosophila* *Christian Schloetterer*, Eirini Christodoulaki, Wei-Yun Lai, Viola Nolte Vetmeduni Vienna

Pleiotropy is the phenomenon that a gene affects multiple phenotypes. The extent of pleiotropy is still disputed, mainly because of power issues. A further challenge is that empirical tests of pleiotropy are restricted to a small subset of all possible phenotypes. To overcome these limitations, we propose a new measurement of pleiotropy, which integrates across many phenotypes and multiple generations to improve power. We infer pleiotropy from the fitness cost imposed by frequency changes of pleiotropic loci. Mixing *Drosophila simulans* populations, which adapted independently to the same new environment using different sets of genes we show that the adaptive frequency changes have been accompanied by measurable fitness costs. Unlike previous studies characterizing the molecular basis of pleiotropy, we show that many loci, each with weak effects, contribute to genome-wide pleiotropy. We propose that the costs of pleiotropy were reduced by the modular architecture of gene expression, which facilitated adaptive gene expression changes with low impact on other functions.

302V Enrichment analyses identify shared associations for 25 quantitative traits in over 600,000 individuals from seven diverse ancestries *Samuel Pattillo Smith*[1,2], Sahar Shahamatdar[1,2], Wei Cheng[1,2], Selena Zhang[1], Misa Graff[3], Christopher Haiman[4], T.C. Matise[5], Kari E. North[3], Ulrike Peters[6], Eimear Kenny[7,8,9,10], Chris Gignoux[11], Genevieve Wojcik[12], Lorin Crawford[1,13,14], Sohini Ramachandran[1,2] 1) Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Department of Ecology and Evolution, Brown University, Providence, RI; 3) Department of Epidemiology, University of North Carolina - Chapel Hill, Chapel Hill, NC; 4) Department of Preventative Medicine, University of Southern California, Los Angeles CA; 5) Department of Genetics, Rutgers University, Piscataway NJ ; 6) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle WA; 7) The Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York City NY; 8) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York City NY ; 9) Department of Medicine, Icahn School of Medicine at Mount Sinai, New York City NY; 10) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai,22New York City NY ; 11)

Division of Biomedical Informatics and Personalized Medicine, University of Colorado, Denver CO; 12) Department of Epidemiology, Johns Hopkins University, Baltimore MD ; 13) Department of Biostatistics, Brown University, Providence RI ; 14) Microsoft Research New England, Cambridge MA

Since 2005, genome-wide association (GWA) datasets have been largely biased toward sampling European ancestry individuals, and recent studies have shown that GWA results estimated from self-identified European individuals are not transferable to non-European individuals due to various confounding challenges. Here, we demonstrate that enrichment analyses which aggregate SNP-level association statistics at multiple genomic scales—from genes to genomic regions and pathways—have been underutilized in the GWA era and can generate biologically interpretable hypotheses regarding the genetic basis of complex trait architecture. We illustrate examples of the robust associations generated by enrichment analyses while studying 25 continuous traits assayed in 566,786 individuals from seven diverse self-identified human ancestries in the UK Biobank and the Biobank Japan, as well as 44,348 admixed individuals from the PAGE consortium including cohorts of African-American, Hispanic and Latin American, Native Hawaiian, and American Indian/Alaska Native individuals. We identify 1,000 gene-level associations that are genome-wide significant in at least two ancestry cohorts across these 25 traits, as well as highly conserved pathway associations with triglyceride levels in European, East Asian, and Native Hawaiian cohorts

303V Seeking for autoimmunity risk variants with a strong functional effect by pinpointing targets of natural selection Vasili Pankratov[1], Milyausha Yunusbaeva[2], Sergei Ryakhovsky[2], Maksym Zarodniuk[3], *Bayazit Yunusbayev*[1,2], Estonian Biobank Research Team 1) Institute of Genomics University of Tartu; 2) SCAMT Institute, ITMO University, Saint-Petersburg, Russia; 3) Institute of Bio- and Translational Medicine, University of Tartu, Tartu, Estonia

Causal variants for inflammatory diseases might have been under pathogen-driven natural selection. Such variants are promising for functional experiments since they likely strongly affect the immune response. While this hypothesis has important implications for biomedicine, its application in practice has been hindered by challenges with pinpointing the targets of selection (mutations). We attempted to approach this challenge using Biobank-scale sequence data and a new class of methods based on local tree inference. We focused on 593 risk loci associated with 21 autoimmune disorders. Altogether, 4838 candidate SNPs were analyzed across these loci using Relate-inferred local trees and likelihood-based selection tests using CLUES. We found that 204 out of 593 risk loci contain at least one candidate SNP with evidence for natural selection (logLR>1.59). Inferred selection coefficients suggest that these SNPs were likely under weak and moderate selective sweep. Such sweeps can leave some flanking variation, making it possible to fine-map the target of selection. We were able to fine-map likely targets of natural selection among candidate risk SNPs (57 loci out of 204), distinguish neutral hitchhikers and identify more complex scenarios. Risk SNPs with adaptive history are promising targets for functional analyses since natural selection picks mutations with a tangible effect on the phenotype.

304V Antigenic variation in *Plasmodium falciparum* is maintained on extrachromosomal DNA *Emily R. Ebel*[1,2], Bernard Y. Kim[2], Marina McDew-White[3], Elizabeth S. Egan[1], Timothy J.C. Anderson[3], Dmitri A. Petrov[2] 1) Stanford University School of Medicine, Stanford, CA; 2) Stanford University, Stanford, CA; 3) Texas Biomedical Research Institute, San Antonio, TX

Antigenic variation in the malaria parasite *Plasmodium falciparum* is mediated by the *var* gene family, named for its *var*iability. The high diversity of *var* types in natural populations is thought to result from frequent structural mutations—especially recombination events—that generate novel, chimeric coding sequences. To better understand the mechanisms by which *P. falciparum* generates *var* diversity, we performed long-read sequencing of clonal, haploid parasite populations derived from a 6-month-long mutation accumulation experiment.

We used PacBio reads (7-10 kb in length) to produce high-quality genome assemblies for each clonal lineage of *P. falciparum*. Surprisingly, each assembly contained multiple contigs with inconsistent haplotypes at certain *var* loci. To clarify this result, we generated much longer Nanopore reads (30-400 kb) and developed a homology-based method to visualize gene organization on individual reads. This analysis revealed extensive genetic diversity at four *var* loci within most clonal samples. Individual reads displayed one of several possible haplotypes per locus, which varied from one another by large indels that appeared to create novel, chimeric *var* genes. Notably, the limited set of possible haplotypes at each locus was largely shared across the 17 samples; and the minor haplotype frequencies within each sample were fairly high (3-45%). These observations are inconsistent with the premise that *de novo* mutations generate diversity in each clonal lineage independently. Instead, we hypothesized that multiple *var* haplotypes per locus could be maintained in single *P. falciparum* cells through extrachromosomal DNA (ecDNA), as has recently been shown for a drug-resistance locus. Using quantitative ddPCR, we observed that a *var* locus on chr12 is present at ~1.3 times the chromosomal copy number, con-

sistent with its presence on ecDNA. Furthermore, by combining ddPCR with an enzyme that digests only linear DNA, we found that circular DNA constructs are strongly enriched for the same *var* locus. This discovery that antigenic variation in *P. falciparum* can be maintained on ecDNA has major implications for our understanding of mutation, selection, and the evolution of immune evasion in malaria parasites.

305V Origins and evolution of epigenome-mediated mutation bias Grey Monroe UC Davis, Davis, CA

Mutations are the ultimate source of genetic variation, shaping and fueling evolution. We recently found that mutation rates are lower in gene bodies and essential genes in *Arabidopsis thaliana*. Analyses of synonymous mutations confirm that cryptic selection in mutation accumulation experiments are unlikely to explain these patterns. Instead, these observations support an emerging mechanistic understanding of epigenome-mediated mutation bias. Our findings, like those of others, suggest that the epigenome functions as a scaffold on which DNA repair processes interact. By leveraging the regulated distribution of particular histone modifications, functionally distinct regions of the genome receive more or less DNA repair, and thus less or more mutation, respectively. Here we examine how this mechanistic view fits in theoretical models of mutational modifiers showing that genetic drift imposes a significant constraint on the evolution of adaptive mutation biases. We also explore the evolutionary origins of adaptive mutation bias by examining natural diversity in specific mechanisms and patterns of mutation across diverse organisms.

306V A novel high-throughput approach to measure the fitness effects of protein misfolding mutations *Natalie Quan*, Yuichi Eguchi, Kerry Geiler-Samerotte Arizona State University

Protein misfolding is a common intracellular occurrence that has devastating effects. Misfolded proteins are cytotoxic and their accumulation within the cell can impose a fitness cost that usually manifests as a growth rate reduction. Therefore, isolating and measuring their effects on fitness are integral to understanding the significance of misfolded protein toxicity. There are also basic questions about protein misfolding that remain unanswered, such as, what mutations cause protein misfolding, how does the harm that misfolded proteins cause scale with their abundance within the cell, and why misfolded proteins are toxic. Here, we present a high-throughput system to accurately measure the fitness effects of thousands of misfolded protein variants and to understand the mechanistic basis of their toxicity. Through massively parallel genome editing, competitive growth assays, and next generation sequencing, we study over 2,000 barcoded unique misfolding mutations to either yellow fluorescent protein (YFP) or superoxide dismutase 1 (Sod1) in *Saccharomyces cerevisiae.* We express these proteins such that observed fitness costs are likely due to the gain of cytotoxic misfolded proteins, rather than the loss of any beneficial protein function. Through this system, we measure the fitness defect associated with each mutation. We also devise a novel technique that we name Intra-*FCY1* to estimate how much misfolding is caused by each of our thousands of mutations. Unlike western blots, this orders-of-magnitude higher-throughput technique uses relative growth rates to measure the amount of misfolded protein within each barcoded yeast strain. Our comprehensive screen of mutations in these two proteins reveals which mutations cause the most severe misfolding, and shows how the fitness effects of protein misfolding scale as misfolding becomes more severe. Our Intra-*FCY1* system, surveying 2,000 distinct mutants per protein, represents a notable advancement in investigating the effects of misfolded proteins on cells. This advancement will inform questions in the field of evolutionary biology about the distribution of fitness effects (DFE) for new mutations, since the majority of mutations in protein-coding sequences tend to increase the chance of the resulting protein being misfolded.

307V What makes a generalist? Using whole genome amplification with whole genome sequencing to quantify host-associated genetic structure in root knot nematodes *McCall Calvert*, Linda Wu, Corlett Wood Department of Biology, University of Pennsylvania, Philadelphia, PA

Generalists are species capable of surviving in a variety of niches and are present in most ecosystems. Despite their ubiquity, very little is known about how generalism evolves and is maintained. In this study, we will test two hypotheses for the maintenance of generalism using the obligate parasite and common crop pest, the root knot nematode *Meloidogyne hapla*. *M. hapla* has over 500 recorded host plants and the extant genetic variation for generalism across populations has yet to be described. We will test whether 1) Populations possess large amounts of standing genetic variation, maintained by spatial variation in in host species, that contributes to the expression of generalism at the species level (e.g. individuals and populations specialize on different host plant species) or 2) All individuals in the species possess a conserved set of "virulence" genes that allows individuals to infect most hosts they could potentially encounter. We predict that these two distinct mechanisms should give rise to unique population genetic signatures. For the first hypothesis, population genetic structure should be strongly associated with host plant use, while for hypothesis two, population genetic structure will *not* be associated with host plant use and instead be influenced by geographic distance. To test these hypotheses, we collected root knot nematodes from four host plant species at multiple locations across the Eastern United States.

We sequenced individuals using a combined whole genome amplification and whole genome sequencing approach. To our knowledge this is the first population genetic study using *individual* root knot nematodes. In addition to exploring the evolutionary maintenance of generalism, our work will also quantify dispersal among populations. Dispersal distance is a central component to many models that predict the spread of root knot nematodes, but it has yet to be empirically estimated.

308V Strength of stabilizing selection is associated with the amount of non-additive variance in gene expression Margarita Takou[1,5], Daniel Balick[3], Kim Steige[1,2], Hannes Dittberner[1], Ulrike Göbel[1], Holger Schielzeth[4], *Juliette de Meaux*[1] 1) University of Cologne; 2) University of Hohenheim; 3) Harvard Medical School; 4) University of Jena; 5) Penn State University

The distinction between additive and non-additive components of genetic variance is crucial for predicting the adaptive potential of a variable trait. Yet we know surprisingly little about the genomic and evolutionary factors that influence non-additive genetic variance in natural populations. Here, we use a quantitative genetic breeding design to decompose the additive and non-additive components of variance of 17,657 gene transcripts in the outcrossing plant *Arabidopsis lyrata*. We find that the major fraction of gene expression variance corresponds to non-additive variance in our data. Transcripts showing the highest levels of non-additive variance tend to be longer, cluster in larger groups of co-expressed genes, and are particularly enriched among genes involved in cellular differentiation and epigenetic reprogramming of gene expression. Furthermore, we find that genes with the highest fraction of non-additive variation are exposed to stronger stabilizing selection than genes with higher additive variance. As a source of many evolutionary novelties, gene expression variation plays a key role in adaptive evolution, yet the amount of non-additive genetic variation will limit its capacity to respond to selection. **Our study is the first to show that stabilizing selection on amino-acid sequences associates with a decrease in a population's potential for adapting gene expression levels.**

309V Inference of the proportion of recessive lethal mutations in humans and Drosophila Chris Kyriazis[1], Emma Wade[1,2], Maria Izabel Cavassim[1], *Kirk Lohmueller*[1] 1) UCLA, Los Angeles, CA; 2) Mississippi State University, Starkville, MS

Early genetic studies indicated that some mutations, when homozygous, will result in lethality or sterility. While the presence of such recessive lethal mutations is now clearly established, quantifying what proportion of new mutations are recessive lethal has been much more challenging because population genetic studies of the distribution of fitness effects (DFE) may be underpowered to detect such mutations. To overcome this challenge, here we present an alternative approach for quantifying the fraction of new mutations that are recessive lethal. Specifically, we use mutation-selection-drift balance models in concert with previously published estimates of segregating recessive lethals in humans and Drosophila. Our approach shows a very small fraction of new nonsynonymous mutations (<1%) in Drosophila and humans are likely to be recessive lethal. We then we use forward-in-time simulations of human coding sequence with DFEs, mutation rates, and demographic models inferred from the literature, combined with a range of plausible dominance coefficients to validate our model. We find that, under nearly all parameter combinations, existing DFEs of nonsynonymous mutations with an additional 1% of mutations being recessive lethal can explain observed levels of inbreeding depression in humans. Our work validates molecular population genetic estimates of the DFE by showing that these DFEs can recapitulate reductions in fitness measured in an orthogonal manner. Further, this analysis places limits on the proportion of recessive lethal mutations, which has implications for studying inbreeding depression in a variety of species.

310V Polygenic signals of sexually antagonistic selection in contemporary human genomes *Filip Ruzicka*[1], Luke Holman[2,3], Tim Connallon[1] 1) Monash University; 2) University of Melbourne; 3) Edinburgh Napier University

Mutations that increase fitness in one sex may decrease fitness in the other. Such "sexually antagonistic" (SA) genetic variants can constrain adaptation and increase variability for fitness components—including survival, fertility, and disease susceptibility. However, detecting SA selection in genomes has proven to be immensely challenging, requiring prohibitively large datasets that combine genomic sequences with individual fitness measurements. Here, we use genotypic and reproductive success data from ~250,000 UK Biobank individuals to comprehensively assess the extent of SA genetic variation in humans. We first develop new theoretical models for signals of SA selection spanning a full generational life cycle—including SA polymorphisms affecting survival, reproductive success and overall fitness. Comparing our models with UK Biobank data, we uncover multiple empirical signals of polygenic SA selection. These signals include sex-differential effects of genetic variants on each fitness component, and positive correlations between sex-differential effects and minor allele frequencies. We show that these signals cannot be explained by sex differences in purifying selection, or by potential confounders such as population structure and sequence mapping errors. We further show that candidate SA sites disproportionately affect functional genomic regions, including polymorphisms associated with quantitative traits (e.g., height) and disease. Finally, we examine historical evolutionary processes affecting candidate SA sites, which are

consistent with the drift-dominated dynamics predicted by previous theory. Overall, our results reveal the first robust signals of SA selection in human genomes, and highlight its functional and evolutionary consequences.

311V Predicting the Genetic Signatures of Aestivation and Dry-Season Persistence of Malaria-Transmitting Mosquitoes *Tin-Yu Hui*[1], Rita Mwima[2], Austin Burt[1] 1) Department of Life Sciences, Silwood Park Campus, Imperial College London; 2) Uganda Virus Research Institute

Sub-Saharan malaria-transmitting mosquito populations exhibit strong seasonal dynamics with huge temporal population size variation. Very often these populations re-establish and expand rapidly with the first rain after a prolonged period of arid conditions. It remains a great challenge to understand their survival during the dry season. One popular hypothesis is that aestivation, a form of dormancy, helps them persist in the drier months. Such hypothesis, with little direct evidence, is as controversial as the other hypotheses proposed. Utilising the indirect genetic approach we aim to establish the expected genetic signatures under different hypotheses, including aestivation, continuously-reproducing, and migration models. These models are developed by customising the classical Wright-Fisher model. Genetic signals such as linkage disequilibrium, change in allele frequency, heterozygosity, relatedness, are explored. In particular, the discrepancy between longer and shorter term effective population size estimates may be one of the consequences of aestivation. These quantitative findings provide valuable resources to support future population monitoring, parameter estimation, and experimental design.

312V Genetic population description of a liverwort *Cheilolejeunea rigidula* in the Amazon region *Astrid Munoz-Ortiz*, Laura Campos Universidad de La Salle

We studied the genetic structure between populations of the liverwort *Cheilolejeunea rigidula* (Lejeuneaceae), which occurred four lowland rain forest sites from Colombian Amazon (Amazonas, Caquetá, Putumayo, and Vaupés). In addition to 65 successfully sequenced samples from the study sites, we included individuals of *C. rigidula* from Guiana and Brazil (Manaus and Tapajos) to investigate the connectivity and genetic structure of this species across the Amazon region. Each site in Colombia, Brazil and Guiana was considered a subpopulation. The sequenced chloroplast markers (partial *atpB* gene, partial *psbA* gene/*psbA-trnH* spacer) showed little variation across the Amazon. The nuclear marker (ITS) showed a spatial structure indicating genetic differentiation of subpopulations across the Amazon, but again little genetic structure along trees. The gradient across the Amazon shows a relationship between genetic distance and geographic distance, indicating limitations not only in dispersal ($P<0.001$). Our results suggests that dispersal can have a dominant effect on populations and communities at local and regional scales, increasing connectivity.

313V A population dynamics tipping point for aging as a cause of adult death *Andrea Scharf*, Kerry Kornfeld Washington University School of Medicine in Saint Louis

Populations are a fundamental level of biological organization that poses major challenges for analysis. Individual traits that influence development, diapause, reproduction, aging, and lifespan interact in complex ways to determine birth and death. Birth and death drive population dynamics and determine whether a population survives or is doomed for extinction. However, we lack a deep understanding of the relationships between individual traits and population dynamics. Therefore, we developed a laboratory ecosystem using the model organism *C. elegans* and a computational simulation that realistically models the laboratory ecosystem. We used these platforms to investigate the conditions that permit animals in the population to die of old age, a critical step in understanding the role of aging in population dynamics. Old age as a cause of death was influenced by three conditions: maximum lifespan, rate of adult culling, and progeny number/food stability. Remarkably, populations displayed a tipping point for aging as the primary cause of adult death. With high numbers of progeny, almost all adults in a population died young, whereas a slight decrease in progeny number caused a dramatic shift in the population, and almost all adults died of old age. The conditions defined here establish a conceptual framework for understanding why certain animals die of old age in the wild, such as mayflies and elephants.

Andrea Scharf, Josh Mitteldorf, Brinda Armstead, Daniel Schneider, He Jin, Zuzana Kocsisova, Chieh-Hsiang Tan, Francesca Sanchez, Brian Brady, Natasha Ram, Gabriel B. DiAntonio, Andrea M. Wilson, and Kerry Kornfeld. A laboratory and simulation platform to integrate individual life history traits and population dynamics. *Nature Comp Sci* in press

314V Frequency and distribution of some human phenotypes among a population in Akwa Ibom State, Nigeria Ime E. Etim Akwa Ibom State University

Phenotypes are expressed morphogenetic traits that are passed on from parents to their progenies using one of the numerous inheritance patterns. An inheritance pattern responsible for a trait may result in variation within a popula-

tion and the resulting expressions create diversity among populations. This preliminary study aimed to determine the frequency and distribution of some human morphogenetic features among 500 undergraduate students of Akwa Ibom State University in Nigeria. The features observed were fingerprint patterns, hair colour, handedness, eye colour and height. With respect to fingerprint pattern, loop was more prevalent among males (52.28 %) and arch among females (58.40 %). Black hair colour and right handedness were predominant in the population. White was sparsely seen among young adult and the frequency of left handedness was higher in the female (75 %) than in males. Brown eye colour was predominant whereas blue, green and Amber were not observed in the sampled population. The height of the tallest individuals in the population was between 1.71 metres (m) and 1.85 m with most individuals, in the population, in this range. Height of females in the population ranged from 1.55 m to 1.70 m. Further association studies between these traits will reveal the significant use in forensic studies.

315V A pan-genome view of complex trait dissection in *Eucalyptus* *Alexander Myburg*, Julia Candotti, Anneri Lotter, Melissa Reynolds, Nanette Christie, Marja Mostert-O'Neill, Forest Molecular Genetics Programme, University of Pretoria, Pretoria, South Africa Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

Eucalyptus is an iconic genus of woody perennials that harbours vast genetic, chemical and morphological diversity and occupies diverse ecological niches in Australia and islands to its north. A handful of fast-growing eucalypt species and their interspecific hybrids form the genetic basis for the most widely planted hardwood fibre crop globally (>20 mha). Breeding material from these species remains essentially undomesticated, or is undergoing early domestication. Population genomics studies using genome-wide SNP markers have provided the first insights into patterns of genome diversity in the natural range and how this diversity is reshaped by early domestication, selective breeding and interspecific hybridization in breeding populations. Complex trait dissection in these trees is however hampered by high levels of genetic load and long generation times. Furthermore, haplotype and structural variation, key components of pan-genome variation, are relatively unexplored as sources of complex trait variation in these outbred organisms. To address this, we are characterizing haplotype and structural variant (SV) diversity in *Eucalyptus grandis* and *E. urophylla,* the parents of the most widely planted eucalypt hybrid combination in subtropical zones. Using Flex-seq technology (RAPiD Genomics), we developed a haplotype mining panel using over 10,000 oligonucleotide probe sets targeting coding and/or non-coding sequences at 5000 genes with multiple evidence lines supporting their involvement in growth, wood properties and biotic and abiotic interactions. We also produced ~200X long read (Oxford Nanopore) coverage and ~100X short-read (Illumina) coverage of three F_1 progeny and used a trio-binning approach to assemble haplogenomes inherited from their *E. grandis* pollen parents and *E. urophylla* seed parents, and then identified SVs within and between the parental species. Linkage maps of the parental genomes were derived from SNP72K chip genotypes for approx. 384 F1 hybrid progeny from each of a series of interconnected full-sib families for QTL detection and nested association mapping. We aim to impute haplotypes and SVs segregating in the population as a first step towards assessing and understanding how these two types of pan-genome variation may affect quantitative trait variation in *Eucalyptus* tree species.

316V Understanding epistasis in the Hsp90 network *Gaurav Bilolikar*, Kerry Geiler-Samerotte Arizona State University

Biological complexity is built on different types of molecular interactions, such as those in protein-protein interaction (PPI) networks. The highly connected 'hubs' in a PPI network interact with many of the less-connected proteins and often exhibit different types of epistasis. For example, Heat shock protein 90 (Hsp90), a hub and conserved molecular chaperone, can sometimes buffer mutations in other proteins (i.e., dampen their phenotypic effects) but in other cases can potentiate (i.e., enhance) these effects. The mechanisms behind these differential epistatic interactions are not well understood. Understanding mechanisms of epistatic interactions will result in better predictions about when a mutation will be buffered vs. potentiated, and an improved understanding of how epistasis contributes to the genotype-phenotype map.

Previous high throughput studies of epistasis focus on gene knockouts and therefore may fail to describe and understand the mechanisms underlying epistasis involving single-nucleotide mutations. To study epistasis in a high throughput manner, we used 'Cas9 retron precise parallel editing via homology' (CRISPEY). to generate ~4982 unique single-nucleotide mutations. These mutations were a mixture of randomly chosen variants and standing genetic variation among natural populations of S.cerevisiae in the genomic region of 92 proteins with known or predicted interactions with Hsp90.

This project focuses on studying the impact of Hsp90 on mutations in its PPI network, including the Ras/PKA signaling pathway of S.cerevisiae. Genes in the Ras/PKA signaling pathway among other pathways in the Hsp90 network are hotspots for rapid adaptation in yeasts and diseases such as cancers. Understanding how mutations in these pathways in-

teract could be particularly useful for evolutionary forecasting. Using a barcode-lineage tracking approach, the fitness of these mutations will be measured relative to unmutated control strains in two conditions, one in which Hsp90 is inhibited and one in which it is not. This study will reveal the prevalence of epistasis within the Hsp90 PPI network, and whether certain types of mutations demonstrate predictable patterns of epistasis, for example, whether mutations within the same gene tend to have the same type of epistasis with Hsp90. By looking for predictable patterns of epistasis, this study may suggest underlying mechanisms that cause epistasis in PPI networks. Ultimately, a better understanding of epistasis will help us make better clinical and evolutionary predictions about how mutations will interact.

317V The relationship between the distribution of fitness effects and the distribution of mutation rates *David Castellano*, Ryan Gutenkunst University of Arizona

The distribution of fitness effects (DFE) of new mutations is a key parameter in molecular evolution. The DFE has been explored across different species and across different genes within a species. Recent work has compared the DFE for sites with different levels of conservation across a multispecies alignment finding that the more conserved a site, the more likely mutations occurring at this site are deleterious. In this work, we compare the DFE of sites with different mutation rates in humans based on their 3mer sequence context. For each of the 64 mutation types, we compute their mutation rate using de novo mutations from trio data and their DFE using the distribution of allele frequencies of synonymous and non-synonymous mutations occurring at the same sequence context. We correlate the mutation rate and the mean selective effect of the 64 mutations, compute the unweighted DFE (that is an artificial DFE where all sites have the same mutation rate) and compare it with the observed DFE under the germline and somatic mutation spectra. Our work highlights the importance of mutation bias across sites to explain the observed DFE in humans and the unappreciated tumorogenic role of the somatic mutation spectrum.

318V Genome-wide comparison of artificially induced mutations and natural variations in *Brachypodium distachyon* *Li Lei*[1], Joel Martin[1], Mingqin Shao Shao[1], Jie Guo[1], Amy Cartwright[1], Jeremy Philips[1], Lesly Fang[1,2], Jacob Espinosa[1,2], Corey Carter[4], Skylar Wyant[4], Chaochih Liu[4], Peter Morrell[4], David Goodstein[1], Richard Sibout[5], Debbie Laudencia-Chingcuanco5[3], John Vogel[1,6] 1) DOE Joint Genome Institute; 2) The University of California, Merced, CA; 3) USDA-ARS Western Regional Research Center, Albany, CA, USA; 4) Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN USA 55108; 5) Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000, Versailles, France; 6) UC Berkeley Department of Plant and Microbial Biology, Berkeley, CA, USA

Natural variation has been extensively used during the domestication and improvement of plants and animals. However, the spectrum of polymorphisms present in natural variation is limited due to the low spontaneous mutation rate (5.9 - 7 x 10-9 per base pair per generation) and natural selection against variants that may be beneficial in an agricultural setting but deleterious in a natural environment. To overcome these limitations, artificially induced mutations have been to accelerate the development of improved cultivars. However, the nature of specific mutations induced the molecular and physiological basis for the improved characteristics in the varieties has been poorly understood.[vv1] In this study, we created a powerful reverse and forward genetic tool by sequencing a collection of 2,000 chemical and radiation mutants in the model grass *B. distachyon*. In total, we identified ~2 million mutations and predicted which mutations, including nonsynonymous mutations, may alter protein function. We also compared the induced mutations with the natural variation from 116 diverse *B. distachyon* lines. The[vv2] chemical mutagens ethyl methanesulfonate and sodium azide tend to induce more transition mutations, particularly C->T or G ->A transitions than radiation and natural variation. Furthermore, induced mutations are predicted to alter protein function (with a higher proportion of deleterious mutations) than natural polymorphisms. We used a pan-genomic approach to determine that induced mutations induce deleterious mutations evenly in all genes whereas potentially deleterious natural variants are less likely to occur in conserved (core) genes than in variable genes. By predicting the potential deleterious effects of nonsynonymous mutations we increased the utility of the mutation collection to study all aspects of grass biology.

319V The male ejaculate proteins and their lineage-specific variation in *Apis mellifera* *Bahar Patlar*, Kathleen A. Dogantzis, Amro Zayed York University, Toronto, ON, Canada

Male reproductive genes, especially ones coding seminal fluid proteins are known amongst the most rapidly evolving genes. Thus, they likely accumulate genetic differences between populations at a high rate that may eventually contribute to speciation. However, our knowledge is limited on the amount of molecular variation in seminal fluid genes among the different genetic populations within a species. Honeybee, *Apis mellifera* consists of at least seven genetically distinct evolutionary lineages (Africa: A and L lineages, Asia: Y and O lineages, Europe: C lineage, and Eurasia: M lineage, Madagascar island: U lineage) that occur naturally over the vast and varied geographical areas across the world. These lineages are ideal to examine the level of molecular variation in seminal fluid genes between populations and within

species. Therefore, in this study, we first combined the list of drone seminal fluid and sperm proteins which were reported in previous studies to consolidate current knowledge on the identity of the honeybee reproductive genes. Next, we estimated F_{ST}, a relative measure of differentiation, for each ejaculate protein-coding gene to identify genes that may have been subject to selection and contributing lineage diversity. Preliminary results have shown that the drone ejaculate contains in total 549 unique proteins, among them, 209 and 263 proteins were identified as in seminal fluid or sperm-specific, respectively. The rest of the proteins couldn't be classified as belonging to one or another group, however, they could belong to both. We are currently performing F_{ST} analyses to understand the amount and patterns of differentiation in ejaculate proteins among genetically distinct lineages of honeybees.

320V Investigating patterns of methylation associated with hypoxia in lowland and highland *Peromyscus maniculatus* populations *Dhriti Tandon*[1], Shane Campebell-Staton[1], Zac Cheviron[2], Bridgett vonHoldt[1] 1) Princeton University, Princeton, NJ; 2) University of Montana, Missoula, MT

Epigenetic modifications such as DNA methylation act as means for phenotypic plasticity, as these changes drive quick and reversible phenotypic responses in comparison to sequence-level genetic changes. *Peromyscus maniculatus* populations can span altitudinal gradients, and consequently harbor plastic physiological mechanisms to adapt to low oxygen conditions typically found in high altitudes. However, studying DNA methylation changes specific to oxygen metabolism can be challenging, as wild mice can harbor non-specific methylation differences due to a large number of environmental variables and lab mice can accumulate epigenetic changes that could make results less representative of mechanisms driven by natural environmental variation. We used reduced representation bi-sulfite sequencing of left ventricle tissue to identify sites that are differentially methylated in response to hypobaric hypoxia. We conducted two statistically controlled analyses of these data. In the first, we compared lab-reared mice exposed experimental hypoxia to those housed under normoxia, and in the second, we compared wild mice sampled in lowland (450 m a.s.l.) and highland conditions (4350 m a.s.l). We identified common differentially methylated sites across the two analyses, with methylation differences in the same direction for both low-oxygen conditions. Some of these sites were present within genes *Zfp142*, *Egln3*, *P pt2* and *Stard13*, and have at least 25% methylation difference between the low and high oxygen conditions in both analyses. Strikingly, *Egln3* encodes the Egl-9 Family Hypoxia Inducible Factor 3. *Egln3* is involved in the regulation of hypoxia inducible factor, a transcription factor that mediates physiological responses to hypoxic conditions and plays is a critical role in maintaining oxygen homeostasis. We also found *Egln3*, *Ppt2* and *Stard13* are connected through a common gene network, which is significantly enriched for gene ontology terms such as "protein hydroxylation", "response to decreased oxygen levels" and "L-ascorbic acid binding". These results underscore the prevalence of environmentally induced epigenetic modifications, possibly causing intra-specific physiological differences in deer mice populations spanning an altitudinal gradient.

321V Genome-wide association mapping of transcriptome variation in *Mimulus guttatus* indicates differing patterns of selection on *cis*- versus *trans*-acting mutations *Keely Brown*[1], John Kelly[2] 1) University of California, Riverside, Riverside CA; 2) University of Kansas, Lawrence KS

We measured the floral bud transcriptome of 151 fully sequenced lines of *Mimulus guttatus* from one natural population. Thousands of single nucleotide polymorphisms (SNPs) are implicated as transcription regulators, but there is a striking difference in the allele frequency spectrum of *cis*-acting and *trans*-acting mutations. *Cis*-SNPs have intermediate frequencies (consistent with balancing selection) while *trans*-SNPs exhibit a rare-alleles model (consistent with purifying selection). This pattern only becomes clear when transcript variation is normalized on a gene-to-gene basis. If a global normalization is applied, as is typically in RNAseq experiments, asymmetric transcript distributions combined with "rarity disequilibrium" produce a superabundance of false positives for *trans*-acting SNPs. To explore the cause of purifying selection on *trans*-acting mutations, we identified gene expression modules as sets of coexpressed genes. The extent to which *trans*-acting mutations influence modules is a strong predictor of allele frequency. Mutations altering expression of genes with high "connectedness" (those that are highly predictive of the representative module expression value) have the lowest allele frequency. The expression modules can also predict whole-plant traits such as flower size. We find that a substantial portion of the genetic (co)variance among traits can be described as an emergent property of genetic effects on expression modules.

322V Using Tree-Based Identity-By-Descent Segments to Evaluate the Effect of Directional Selection on the Estimation of Recent Effective Population Size and Population Structure in *Plasmodium falciparum* *Bing Guo*[1,2], Shannon Takala-Harrison[2], Timothy O'Connor[1] 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; 2) Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD

Intensive malaria elimination efforts to counter the spread of multidrug-resistant *Plasmodium falciparum* (*P.f.*) have led

to a dramatic decrease in malaria cases in the Greater Mekong Subregion over the past decade, with a corresponding reduced parasite effective population size (Ne) and increased population structure. Estimates of parasite population demography are critical for monitoring sources and sinks of malaria transmission and the impact of targeted interventions in reducing malaria. Identity-by-descent (IBD) has been widely used in human studies to estimate Ne and population structure in recent time scales. However, the strong directional selection related to antimalarial drug resistance in *P.f.* may violate the assumptions made in human-oriented IBD-based tools, potentially leading to inaccurate estimation of parasite population demography. Here, we combined the power of coalescent (*msprime*) and forward simulators (*SLiM*) to simulate the genealogical trees with varying directional selection parameters, implemented a tree-based true IBD-estimation algorithm (which circumvents the bias introduced by sequence-based IBD inference), and directly evaluated how selection affects IBD-based demography estimation in *P.f.*. With high-quality IBD as input, we found that selection duration, strength, and selected allele frequency all affect IBD coverage and length distribution, and thus the IBD-based estimation of parasite Ne (IBDNe). The effects are further aggravated when reductions in Ne occur in a recent time frame. We found that removing IBD segments within the high-IBD region associated with selection and splitting the chromosome abolishes selection-induced bias in Ne estimation, especially for the most recent 30 generations. We also designed a model of multiple subpopulations with various migration parameters to test the effect of directional selection on IBD-based population structure estimation. We hypothesize that strong directional selection could increase within and inter-population IBD sharing and hide the underlying population structure. However, using filtered IBD segments could recalibrate the inference and thus allow identification of finer-scale, more accurate population structure in low transmission settings with active migration of drug-resistant parasites. Ongoing work will include genealogy inference, tree-based IBD finding, and recalibrated IBD-based estimation of *P.f.* demography using whole-genome sequencing data from field isolates.

323V Fitness contributions of the *Responder* satellite in *Drosophila melanogaster* *Matthew Lindsay*, Danna Eickbush, Xiaolu Wei, Amanda Larracuente University of Rochester

Large blocks of repetitive, non-coding satellite DNA are a major component of eukaryotic genomes, but their exact functions or contributions remains unknown. While satellite DNA is typically seen as 'junk' DNA that accumulates in genomes, some also have roles in chromosome segregation and nuclear organization. However, few specific functions or fitness effects have been assigned to satellite loci. In order to study whether and how satellite DNA contributes to fitness, we used CRISPR to modify the *Responder (Rsp)* satellite in *Drosophila melanogaster* to investigate its specific fitness contributions. *Rsp* is the target of the *Segregation Distorter (SD)* meiotic driver: chromosomes with large *Rsp* loci (many copies of the *Rsp* element) are susceptible to destruction by *SD*, and heterozygous *SD* males only produce sperm bearing *Rsp*-lacking *SD* chromosomes. While complete *Rsp* deletions are viable and fertile, large, drive-susceptible *Rsp* loci are common in wild populations. In a previous study, a large deletion including *Rsp* significantly reduced fitness, suggesting *Rsp* may serve an unknown but important function. However, this large deletion removes many sequences, making it difficult to attribute fitness effects to *Rsp* specifically. Here we used CRISPR to make precise modifications of the *Rsp* locus, and used PacBio HiFi sequencing to investigate the modified loci. We show that *Rsp* deletions significantly reduce sensitivity to *SD* in previously sensitive backgrounds, indicating that we successfully altered the phenotype associated with the *Rsp* satellite. We are conducting relative fitness assays to determine the effect of the *Rsp* deletion on fitness relative to ancestral controls, and are generating *Rsp* deletions in different lines to determine whether the fitness effects are dependent on genetic background. In addition to the experimental work, we are using genomic data from a diverse set of world-wide natural populations of *D. melanogaster* to determine how the frequency of *SD* affects *Rsp* copy number in the wild. Using simulations and modelling, we will also detect any deviations of observed *Rsp* copy number from simulated neutral expectations, which may imply that natural selection acts on *Rsp* copy number. These experiments will help explain why drive-sensitive alleles of *Rsp* persist in natural populations- leaving them vulnerable to the selfish *SD* system- and could more broadly help explain why satellite DNA makes up such a large proportion of eukaryotic genomes.

324V Substitution load imposes a mild constraint on adaptation, with a high proportion of deaths in A. thaliana being selective *Joseph Matheson*[1], Moises Exposito-Alonso[2], Joanna Masel[1] 1) University of Arizona, Tucson, AZ; 2) Carnegie Institution for Science, Stanford University, Stanford, California

Haldane (1957) argued that the need for a minimum number of "selective deaths" over the course of an adaptive substitution acts as a limit to the speed of adaptation. He estimated that mammalian species could only sustain one adaptive substitution every 300 generations. The fact that substitutions occur much faster than this limit was the original argument put forward in favor of neutral theory. However, while most population geneticists today no longer consider Haldane's argument to imply a serious limit to adaptation, there is no agreement as to the reasons why, nor what the true limit is. Many follow Maynard Smith (1968), attributing the resolution of the problem to an extreme form of syn-

ergistic epistasis, which there is now enough data to exclude. More plausible is that early load arguments mistakenly compared average fitness to the fitness of a perfectly adapted population (Ewens 1970), a mistake also corrected as part of Maynard Smith's model. Here we disentangle multiple distinct lines of reasoning about the ambiguous term 'substitution load', including selective deaths, reproductive excess, and load sensu strictu. Importantly, while a variety of considerations do not meaningfully limit adaptation, reasoning based on selective deaths still can. We therefore applied the concepts of load and selective deaths to survival and fecundity data on 517 different genotypes of Arabidopsis thaliana grown in eight different environmental conditions. We estimate highly permissive limits to the speed of adaptation in all environmental conditions. While harsher environmental conditions decrease reproductive excess and hence the potential for selective deaths, this is compensated by higher proportions of deaths being selective. Selective deaths are more common than anticipated during historical discussions of speed limits. Less fecund species than Arabidopsis could nevertheless face meaningful limits to the speed of adaptation, especially in harsher environments where adaptation is most important.

325V Mixture Density Regression reveals frequent recent adaptation in the human genome *Diego F. Salazar-Tortosa*[1,2], Yi-Fei Huang[3,4], David Enard[1] 1) Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA.; 2) PROFITH 'PROmoting FITness and Health through physical activity' research group, Sport and Health University Research Institute (iMUDS), University of Granada, Granada, Spain; 3) Department of Biology, Pennsylvania State University, University Park, PA, USA.; 4) Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA.

How much genome differences between species reflect neutral or adaptive evolution is a central question in evolutionary genomics. In humans and other mammals, the prevalence of adaptive versus neutral genomic evolution has proven particularly difficult to quantify. The difficulty notably stems from the highly heterogenous organization of mammalian genomes at multiple levels (functional sequence density, recombination, etc.) that complicates the interpretation and distinction of adaptive vs. neutral evolution signals. Here, we introduce Mixture Density Regressions (MDRs) for the study of the determinants of recent adaptation in the human genome. MDRs provide a flexible regression model based on multiple Gaussian distributions. We use MDRs to model the association between recent selection signals and multiple genomic factors likely to affect positive selection, if the latter was common enough in the first place to generate these associations. We find that a MDR model with two Gaussian distributions provides an excellent fit to the genome-wide distribution of a common sweep summary statistic (iHS), with one of the two distributions likely capturing the positively selected component of the genome. We further find several factors associated with recent adaptation, including the recombination rate, the density of regulatory elements in immune cells and testis, GC-content, gene expression in immune cells, the density of mammal-wide conserved elements, and the distance to the nearest virus-interacting gene. These results support that strong positive selection was relatively common in recent human evolution and highlight MDRs as a powerful tool to make sense of signals of recent genomic adaptation.

326V Comparing accuracy of forensic DNA mixture analysis across populations with varying genetic diversity *Cara Ly*[1], Kamillah Felix[1], Evan Ho[1], Chris Godek[1], Niquo Ceberio[1], Maria Flores[2], Hannah Thorner[3], Carina Kalaydjian[2], Matt Paunovich[1], Rori Rohlfs[1] 1) San Francisco State University; 2) University of California, Los Angeles; 3) George Washington University

Police are increasingly using trace amounts of DNA in investigations, often when samples contain DNA from multiple contributors. The reliability of interpreting these DNA mixtures is complex. A likelihood ratio (LR) is a value used to assess whether genetic evidence at the crime scene supports the defense hypothesis (the suspect did not contribute to the DNA mixture), or the prosecutor's hypothesis (the suspect did contribute to the mixture). The LR is calculated based on the allele frequency distribution of the assumed population group that the suspect belongs to, which will be referred to as the reference population. When the reference population's allele frequency is different from the suspect's true genetic background, the accuracy of the LR results can be impacted. We hypothesized that there will be higher error rates when the reference population is incorrectly assumed, and even higher when the true population has low genetic diversity. We also expected to see that as the number of contributors in a mixture increases, the amount of false positives increases. To test our hypothesis, we used Forensim – a free open source R package – to simulate individual genotypes, forensic DNA mixtures, and calculate the LR, which allows us to compare the probability of the data for when the suspect does and does not contribute to the mixture. We observed that when there are more individuals in the DNA mixture, there is less reliability in assessing the evidence that a suspect did not contribute. With the correct reference population, there was an increase in the false positive rate for populations with low genetic diversity compared to populations with high genetic diversity. We identified more false positives with an incorrect reference population and an even higher false positive rate when the true population of the suspect has low genetic diversity and the reference population has much

higher genetic diversity. Our results indicate that forensic DNA mixture analysis tools used today may falsely identify individuals with certain genetic backgrounds due to differences in both genetic diversity and the appropriateness of reference populations.

327V Pedigree reconstruction in the era of many thousands of samples *Daniel Seidman*, Ryan O'Hern, Amy L. Williams Cornell University Graduate School

As modern genetic datasets grow in size, the fraction of samples with one or more close relatives in a given dataset increases. These relationships could allow for the construction of massive numbers of pedigrees, but scalable and accurate pedigree reconstruction methods rare. Pedigrees have wide-spread utility as they can improve the quality of phasing and imputation, help trace the origin of alleles, and yield enhanced heritability and linkage studies.
We propose PELICAN, PEdigree reconstruction from LIkelihoods and ConstrAiNts, an algorithm that can rapidly and accurately reconstruct pedigrees using the latent relatives in large datasets. Using likelihoods for specific relationship types of both first and second degree relatives, the algorithm creates a list of sorted potential edges for a pedigree graph. First degree relationship likelihoods, either full-sibling or parent-child, are calculated from identity-by-descent (IBD) regions shared by pairs of individuals. Our algorithm receives likelihoods for second degree relationship types, such as grandparent/grandchild or half-sibling, from a separate algorithm called CREST (Qiao, Sannerud, et al. 2021). It combines those likelihoods with additional likelihoods generated from a kernel density estimator (KDE) trained on simulated relatives to form composite likelihoods for the relationships. PELICAN then adds relationships, in order of likelihood, to the pedigree graph. We impose restrictions on what relationships constitute valid additions to the graph, and the algorithm backtracks from situations where: a new relationship results in implied inbreeding in the last two generations, sets of individuals form a biologically impossible combination of relationships, or the partial pedigree cannot form pedigrees of equal or higher likelihood than those already found. With these restrictions in place, the algorithm investigates all possible pedigrees, but does so without performing the time-consuming process of generating and comparing each one explicitly. In this way, PELICAN is guaranteed to find the maximum composite likelihood pedigree.
As a proof of concept, we applied PELICAN to the UK Biobank dataset's ~500,000 samples to demonstrate the algorithm's scalability and provide these pedigrees as a resource to the community. PELICAN performed its analysis in ~3 hours, inferring 11,529 separate pedigrees of more than two samples in size.

328V Genetic variation of Scots pine in Europe and Asia – traces of glacial refugia and human activities *Weronika Barbara Żukowska*[1], Błażej Wójkiewicz[1], Andrzej Lewandowski[1], Witold Wachowiak[1,2] 1) Institute of Dendrology, Polish Academy of Sciences, Kórnik, Poland; 2) Institute of Environmental Biology, Adam Mickiewicz University in Poznań, Poland

Scots pine (*Pinus sylvestris* L.) is the most widespread representative of conifers, native to Eurasia. It is one of the most important forest-forming tree species in the northern hemisphere. Genetic conservation of Scots pine seems to be of low priority because of its very wide distribution spanning several climatic zones. However, the effects of over century-long cultivation and increased mortality due to climate changes make it necessary to describe the genetic resources of Scots pine in more detail.

Using 16 nuclear microsatellite markers, we investigated the genetic variation of 62 Scots pine populations (1,289 individuals) located in Europe and Asia. We confirmed the high overall genetic variation and low interpopulation differentiation (F_{ST} = 0.029, R_{ST} = 0.042; p < 0.001) of this species. The average observed heterozygosity was equal to the average expected heterozygosity ($H_O = H_E$ = 0.525). Lower genetic variation was found in the British Isles, a few mountain populations from Poland, two Italian stands located south of the Alps, and most locations from Eastern Siberia. Private alleles were found mostly in southern regions that are considered the glacial refugia of the species. Central and Northern Europe seems to have been recolonized mainly by Balkan migrants. It appears, however, that other so-called cryptorefugia from higher latitudes, as well as populations from south-eastern regions, could have also contributed to the genetic variation observed in Europe. On the other hand, the populations far east can be largely divided into two groups with the lowest genetic variation in Eastern Siberia.

We detected a significant phylogeographic structure among populations and geographic regions. This means that stepwise-like mutations have contributed to the genetic differentiation of Scots pine. Bayesian assignment and the Principal Coordinates Analysis (PCoA) showed clear differences between populations from Europe and Asia. Only the sites located west of the Urals were genetically more similar to Central and Northern European populations. The test for the deficiency in M-Ratio revealed that most stands have suffered from the past genetic bottleneck. We hypothesize that the lack of genetic structure among Northern and Central Europe, including some western stands, is at least partially the result of

past human activities related to the transfer of germplasm in the 19th and in the early 20th century.

The research was financed by the Institute of Dendrology of the Polish Academy of Sciences and the Polish National Science Centre (grant no. 2016/21/N/NZ9/01499).

329V Genome reassembly in *Chlamydomonas reinhardtii:* A novel approach *Nolan Shelley*, Thamali Kariyawasam, Sunjoo Joo, Jae-Hyeok Lee University of British Columbia

Chlamydomonas reinhardtii is a single-celled green alga and model organism of the plant kingdom exhibiting high intraspecies diversity with a small, relatively well-characterized genome ~110 Mb in size. These features, combined with its haplontic, easily manipulable life cycle and existing mating and tetrad separation protocols, make it an optimal species in which to study meiotic recombination. After mating a pair of divergent parental lines, genome-wide recombination patterns were analyzed across a set of ~1.5 million variant positions in ~500 strains from ~200 tetrads. This analysis led to the identification of chromosomal regions of suspected misassembly in the reference genome. This discovery, in combination with the existence of several dozen known unmapped scaffolds, prompted us to develop a novel genome reassembly protocol using linkage analysis to properly reconstruct the *C. reinhardtii* reference genome. Reassembly procedures that use linkage analysis traditionally involve using a set of independent samples that are each genotyped with high confidence at the same set of loci. This approach, however, is rather limited because sequencing depth, contamination rates and genotyping error rates can vary greatly depending on genomic content, budgetary constraints, and sequencing protocols. We developed a procedure that borrows information from neighboring genotyped variant positions to smooth out uncertainties in genotype caused by low sequencing depth, non-negligible contamination, and other sources of error, and we additionally accounted for the correlation in genotype between samples from the same tetrad. Using this robust approach, we then chose optimal locations to break existing chromosomes/scaffolds into new contigs that were reassembled into groups of ordered sets, making up new chromosomes. The accuracy of these new chromosomes was then confirmed and improved upon using discordant reads of various template sizes. In total, our linkage analysis yielded several dozen large-scale changes to the genome, the vast majority of which were strongly supported by the presence of discordant read connections spanning contig boundaries. Over the coming months, we will be explicitly comparing our new reference genome to the recently released JGI *C. reinhardtii* genome assembled using third-generation sequencing. With this new JGI genome, we will generate the recombination landscape to document the genome-wide profile of crossover and gene conversion events for the first time in unicellular algae.

330V Deriving Biological Insight from Genome Scans: a Tissue Enrichment Method for Noisy Gene Lists *Lauren Sugden*[1], Arthur Sugden[1,2] 1) Duquesne University; 2) Behaivior

Many genomic studies such as selection scans and GWA studies result in lists of genes (hereafter defined as “genes of interest”) determined to contain some statistical signature signifying that these genes are undergoing selection, or are important for understanding a particular disease or phenotype. Often, some (but not all) of these genes may be involved in a common process or pathway that could provide some biological insight, but making these kinds of inferences robustly remains a significant challenge.

A common approach to addressing this compares the genes of interest to various curated gene sets, looking for an overall signature of “enrichment” where genes of interest are overrepresented compared to what one would expect by chance. One consequence of this approach is that a true signal involving a subset of genes of interest can be swamped by genes of interest that are unrelated to the signal. In addition, users can quickly run into multiple testing problems, potentially resulting in a high false discovery rate.

Here, we introduce a method that detects enrichment of tissue-specific genes using gene expression data from GTEx. We first build a bipartite graph of genes and tissues with weighted edges carefully calibrated to account for the vast range of gene specialization vs generality across the genome, as well as the divergent gene-expression profiles across tissues in the database. We then generate a “tissue score profile” for genes of interest as genes are dropped one by one from the list, allowing us to be sensitive to situations in which only a subset of genes of interest are driving a signal. This approach avoids multiple testing problems while still allowing us to observe multiple subsets of the data as we remove genes that introduce noise. Our method allows for genes of interest to have associated weights or probabilities, as might be generated from a genome-wide scan, and generates permutation-based empirical p-values for enrichment scores.

331V Estimation of ibd probabilities for pairs of inbred individuals. *Bruce Weir*[1], Jerome Goudet[2] 1) Univ Washington; 2) Univ Lausanne

Inbred pairs of individuals can share two, three, four or two pairs of alleles identical by descent (ibd) at a locus. Although there are 15 possible ibd states, it is generally sufficient to combine these into nine states with probabilities adding to one, or eight summary states with eight probabilities. For SNP data, some authors claim only five states have identifiable probabilities, while other authors give estimates for probabilities of nine states. In any event the higher-order ibd probabilities for a target pair of individuals cannot be estimated by single-SNP methods unless allele probabilities are known in the population to which ibd refers, and this is unlikely to be the case. Nor is it appropriate to use sample allele frequencies in place of allele probabilities because powers of these frequencies have expectations that depend on the inbreeding and relatedness of all individuals in the study sample.

A remedy is to estimate within-population measures of ibd. Instead of estimating the probability an indivdual carries two ibd alleles, for example, that probability can be estimated relative to the average coancestry of pairs of individuals in the study sample: this is the individual-specific version of Wright's F_{IS}. It is F_{IS} rather than F_{IT} that is estimable with data from a single population. We work with observed proportions of allele pairs that match within or between individuals, and employ ratios of functions of these proportions for which the unknown allele probabilities cancel out of the expected values. With a large number of SNPs, these methods give good estimates of within-population inbreeding and kinship coefficients for pairs of individuals. We extend this approach for within-population estimators of functions of ibd probabiliities for three, four or two pairs of alleles for pairs of inbred individuals.

As an example, we estimate the probability that there is no ibd among any three alleles from two individuals, two alleles from one and one allele from the other, relative to the probability there is no ibd among any three alleles taken randomly, one from each of three individuals in the study sample. This can be compared to the corresponding quantity for two alleles to establish the extent of ibd at sets of three alleles. These measures are needed to express the trait covariance for pairs of inbred relatives.

The advanatage of working with allele-matching, or allele-sharing, is that the estimates are rank-invariant: one pair of individuals has its ibd probabilities less than those for another pair, for example, regardless of the scope of the study including those individuals. Invariance is not guaranteed for estimators that make explicxit use of sample allele frequencies.

336W Simulating systemic effects of expression quantitative trait loci across gene regulatory networks *Matthew Aguirre*[1], Guy Sella[2,3], Jonathan Pritchard[4,5] 1) Department of Biomedical Data Science, Stanford University, Stanford, CA; 2) Department of Biological Sciences, Columbia University, New York, NY; 3) Program for Mathematical Genomics, Columbia University, New York, NY; 4) Department of Genetics, Stanford University, Stanford, CA; 5) Department of Biology, Stanford University, Stanford, CA

Gene regulatory networks (GRNs) govern many of the core developmental and biological processes which give rise to complex traits. Even as genome-wide transcriptomic resources approach population scale, it remains challenging to interpret how the structure of GRNs impacts the distribution of genetic effects on gene expression. Learning the genetic architecture of gene expression traits is a key aim in quantitative genomics, but there is still an unmet need for theoretical modeling which places the statistical basis of these traits in the mechanistic context of GRNs. Here, we propose a simple approach to model and simulate the structure and function of GRNs, making use of techniques from small world network theory and dynamical systems models of gene regulation. Specifically, we model gene expression regulation using a stochastic differential equation with terms for endogenous transcription and contributions from parent nodes (transcription factors) in a GRN. This formulation permits variation in regulatory parameters which naturally corresponds to effects from expression quantitative trait loci (eQTLs) or intracellular interventions (e.g., targeted gene knock-down or knock-outs). We use this model to generate synthetic samples of gene expression data, and probe the system-wide effects of perturbing key regulators in the GRN. We further this analysis by simulating a population of individuals which harbor a substantial burden of eQTLs, investigating the distribution of gene expression heritability introduced by these factors in the network. We conclude by discussing implications of our work towards understanding the architecture of heritable genetic variation in complex traits, even beyond gene expression.

337W Assessing the impacts of single-end and paired-end RNA-seq on gene expression estimates and eQTL detection. *Sam Ardery*[1], Selcan Aydin[1], Daniel A. Skelly[1], Matthew Pankratz[2], Devin K. Porter[2], Ted Choi[2], Laura G. Reinholdt[1], Christopher L. Baker[1], Gary A. Churchill[1], Steven C. Munger[1] 1) The Jackson Laboratory, Bar Harbor, ME; 2) Predictive Biology, Inc., Carlsbad, CA

Short read RNA sequencing (RNA-seq) has become the prevailing method for quantifying gene expression levels genome-wide. The most common RNA-seq platforms sequence 75-100bp reads from one or both ends of RNA fragments, termed single-end (SE) or paired-end (PE), respectively. While SE and PE reads yield similar expression estimates for most genes in isogenic samples, for the subset of genes that do differ substantially, PE reads have generally been found to yield

more accurate estimates owing to better alignment specificity. However, this finding has not been verified in genetically diverse samples nor is it clear how these differences affect our ability to detect expression quantitative trait loci (eQTLs). To address these questions, we analyzed a large 2x75bp PE RNA-seq dataset from 185 genetically diverse mouse embryonic stem cell lines (mESCs). We compared gene-level estimates of transcript abundance from aligning just the forward reads (SE) to those from the full paired read (PE) alignments, and then used both expression values along with mESC genotyping data to map eQTLs. We identified nearly 1,500 genes as expressed in one of the analyses but not the other (1,065 in SE, 427 in PE), and gene annotation showed that the SE list was overrepresented for pseudogenes while the PE list was overrepresented for protein-coding genes. Analysis of uniquely aligning reads in the SE data show that many are likely transcribed from protein coding genes but misalign to pseudogenes, a problem exacerbated by the high genetic diversity in the mESC lines. These alignment errors appear to affect eQTL detection in two related ways by causing spurious genetic signals (false positive) and missing real genetic signals (false negative). Importantly, by limiting spurious read alignment to pseudogenes and correctly assigning more reads to protein coding genes, PE sequencing results in fewer false positive eQTLs for pseudogenes and fewer false negative eQTLs for protein coding genes, and in so doing provides a more accurate understanding of gene regulatory variation compared to SE RNA-seq. Future efforts will be focused on improving SE expression estimates using alternative read aligners or alignment strategies. We recommend that researchers use PE RNA-seq in eQTL mapping studies, especially when using samples with high levels of genetic variance from the reference genome used for alignment.

338T Genetic dissection of the pluripotent proteome through multi-omics data integration *Selcan Aydin*[1], Tian Zhang[2], Duy Pham[1], Daniel A. Skelly[1], Matthew Pankratz[3], Devin K. Porter[3], Greg Keele[1], Ted Choi[3], Steven Gygi[2], Laura G. Reinholdt[1], Christopher L. Baker[1], Gary A. Churchill[1], Steven C. Munger[1] 1) The Jackson Laboratory; 2) Harvard Medical School; 3) Predictive Biology

The phenotypic variability observed across pluripotent stem cell (PSC) lines currently limits their use in personalized medicine. Genetic background is a major driver of this variability, and studies addressing it have relied on transcript abundance as the primary measure for gene expression. However, little is known about how proteins, the functional units in the cell, vary across PSCs and how this relates to variation in other measures. Here we present the first comprehensive genetic study characterizing the pluripotent proteome using 190 unique mouse embryonic stem cell lines derived from highly heterogeneous Diversity Outbred mice. Genome-wide comparisons of protein abundance to chromatin accessibility and transcript abundance showed high levels of co-variation. This co-variation was evident in the quantitative trait loci (QTL) results, as 39% of the total 1,676 significant protein abundance QTL (pQTL) co-mapped with chromatin accessibility and expression QTL (eQTL). Most of these shared QTL mapped proximal to the gene itself and likely reflect cis-regulatory polymorphisms. In contrast, 34% of pQTL were unique to protein abundance, and most of these loci were distal. To distinguish shared and unique drivers of variability across molecular layers, we integrated the genomic data sets using multi-omics factor analysis. Integration resulted in 22 latent factors that cumulatively explain 28%, 39%, and 35% of the variation in chromatin accessibility, transcript abundance, and protein abundance, respectively. Functional characterization showed that these factors capture variation relevant to pluripotency maintenance. For example, Factor 3 was highly correlated to the genotype of cell lines at the *Lifr* locus on Chr 15, which we previously identified as a major eQTL hotspot influencing pluripotency maintenance. As expected, we showed that Factor 3 maps with a significant QTL to the same locus. Remarkably, we found many proteins and transcripts that, although they strongly contribute to Factor 3 and exhibit similar allele-level effects on abundance at the *Lifr* locus, each lack a significant Chr 15 distal QTL. These findings highlight the power of multi-omics data integration in revealing the distal impacts of genetic variation. While QTL mapping with individual traits may be limited due to noise introduced by measurement error, data integration can act to consolidate the influence of genetic signals shared across molecular traits and increase detection power.

339T Diverse environmental perturbations reveal the evolution and context-dependency of genetic effects on gene expression levels *julien Ayroles*[1], Amanda Lea[1,2] 1) Princeton University; 2) Vanderbilt University

There is increasing appreciation that complex traits are determined by poorly understood interactions between our genomes and daily environments. These “genotype x environment” (GxE) interactions remain difficult to map at the organismal level but can be uncovered using molecular phenotypes. We address this broad question by asking how prevalent are context-specific versus ubiquitous eQTL? What are the evolutionary forces (e.g., genetic drift, purifying selection, positive selection) that maintain context-specific versus ubiquitous eQTL, and do these forces differ depending on the evolutionary history of the environmental exposure? To address these questions, we embarked on a large-scale study in which we profiled genome-wide gene expression levels across 12 different cellular exposures using 544 lymphoblastoid cell lines (B cell) derived from the 1000 Genomes Project. We used these data to map the genetic basis of gene expression across all 12 conditions and revealed a context-dependent genetic architecture. Our results highlight the extent to

which gene expression is fine-tuned by the environment. We found that 23% of eQTLs were context-dependent, that is almost 1 in 4 eQTLs are "response eQTL" or SNPs that do not affect variation in gene expression under the control condition but for which genetic effects are revealed by experimental treatments. We also found that the mean per-gene heritability estimates were significantly higher in almost all treatment conditions relative to their respective controls, emphasizing the important role non-linear genetic effects are likely to play in explaining the missing heritability. These experimental treatments included stimuli familiar to B cells such as immune signaling molecules and hormones, but also man-made chemicals that have not co-evolved with B cells through evolutionary time. Evolutionary analyses revealed that positive selection has shaped GxE loci involved in responding to immune challenges and hormones, but not man-made chemicals, suggesting there is reduced opportunity for selection to act on responses to molecules recently introduced into human environments. Together, this work highlights the importance of considering an exposure's evolutionary history when studying and interpreting GxE interactions and provides new insight into the evolutionary mechanisms that maintain GxE loci in natural populations.

340W Exploiting inherent interdependencies among traits for genetic association analysis *Haoran Cai*, David Des Marais Massachusetts Institute of Technology, Cambridge MA

Quantitative trait locus (QTL) mapping and genomic prediction are pivotal for quantitative genetics and crop improvement. Published studies are almost always univariate, considering each trait separately. Problems remain in how to properly exploit the interdependency among traits in QTL mappings. The aim underlying such joint modeling is to increase the statistical power. However, simply using phenotypic or genetic correlation to improve power for QTL analysis may incur problems. Because the correlation between traits does not imply information in a single SNP level. It is the combination of QTLs that shape the interdependency among traits: a strong correlation between traits need not result in a strong pleiotropy of individual QTLs. Here, by leveraging the environmental correlation, we propose a multivariate framework to detect causal QTLs and systematically assess the pleiotropic structure of genotype-phenotype map. The key assumption underlying our approach is that the overlapped covariance structure of environmental and genetic variation provides the inherent interdependencies and modular nature of trait combination. We will demonstrate how our approach provides interpretable results and causal understandings of genetic architecture.

341W Evolutionary change in age at first reproduction in a preindustrial human population is faster at times of high infant mortality *Walid Mawass*[1], Emmanuel Milot[2] 1) University of Arizona, Tucson, AZ, USA; 2) Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

Measuring fitness-associated genetic change in a natural population is a persistent goal for evolutionary biologists. Quantitative genetics (QG) provides theory and methods to extract information about the raw material on which natural selection can work, i.e., the additive genetic variance. So far, Milot et al. (2011) remain the only study in humans to robustly show contemporary evolution in response to selection, using the preindustrial île aux Coudres (IAC) population. It was determined by QG analysis combined with pedigree information that the age at first reproduction (AFR) of married women, which evolved from 26 y to 22 y within 5 to 8 generations, was partially due to a genetic change in response to selection. We hypothesized that fluctuations in environmental conditions might induce changes in the genetic architecture of AFR and hence its rate of evolutionary change, mainly through genotype-by-environment interactions or GxE. We tested this hypothesis using the IAC population by incorporating the infant mortality rate as a proxy of the harshness of conditions in early life. Our results detected the presence of GxE underlying the variation in AFR and relative fitness in this population. Based on our predictions using the Roberston-Price covariance, GxE interactions led to an increase in the expected per-generation genetic change in AFR under harsh early-life conditions compared to more benign conditions. Deep-rooted genealogical information is available for the historical French-Canadian population, and it is vital to determine the reliability of this dataset in terms of QG parameter estimation. The performed power and precision analysis on the reconstructed pedigrees revealed that most of the datasets were powerful enough to detect a simulated genetic effect. Precision and accuracy suffered greatly when the model did not include all sources of similarity between relatives (e.g., similarity due to familial environment) regardless of the sample size and depth of the pedigree. Finally, we tested if a response to selection in AFR also occurred in another larger preindustrial French-Canadian population, the Charlevoix population. Predictions point to a weak expected genetic response to natural selection in AFR in this second population. The fact that the detected temporal change in the average breeding values of AFR is equally as likely under a scenario of drift alone constitutes a robust demonstration of evolutionary stasis.

342T Individual Loci Radically Alter the Genetic Architecture of Complex Traits *Gareth Cromie*[1], Russell Lo[1], Lauren Ames[1], Trey Morgan[1], Katherine Owens[2], Anne Clark[3], Martin Timour[1], Julee Ashmead[1], Michelle Tang[1], Nathan Kutz[2], Joshua Akey[4,5], Aimee Dudley[1] 1) Pacific Northwest Research Institute; 2) Department of Applied Mathematics, University

of Washington, Seattle, Washington, USA ; 3) Department of Genome Sciences, University of Washington, Seattle, Washington, USA ; 4) Lewis Sigler Institute, Princeton University, Princeton, New Jersey, USA ; 5) Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, USA

While advances in DNA sequencing technology have greatly increased our ability to map phenotype to genotype, many biomedical, agricultural, and evolutionary phenotypes of interest are complex and remain difficult to genetically dissect. Here, we present a powerful resource for delineating the genetic architecture of complex and quantitative traits in the budding yeast, *Saccharomyces cerevisiae.* Using an eight-parent funnel cross design that captures a significant proportion of the genetic diversity of the global yeast population, we have generated an unstructured mapping population of ~11,500 genetically diverse strains and determined their genome sequences. Using these data, and accurate individual phenotying of each strain, we investigated the genetic architecture underlying ten quantitative traits. Confirming the remarkable power of the resource and leveraging only standing genetic variation, we demonstrate that we can identify the regulatory and mechanistic components of a complete biological pathway. We also detect genetic heterogeneity underlying strong quantitative trait loci (QTL), and dissect one example down to the level of two closely-linked quantitative trait nucleotides (QTN). Finally, we demonstrate strong effects of genetic stratification, identifying individual alleles defining subpopulations with radically different genetic architectures for the same trait.

343T Network analysis of complex trait evolution *Elli Cryan*[1, 2], Daniel Kliebenstein[1], Jeffrey Ross-Ibarra[2] 1) UC Davis Plant Sciences, Davis, CA; 2) UC Davis Evolution and Ecology, Davis, CA

Traits rarely evolve through changes in a single gene. Often a network of genes must evolve in concert to give rise to a new complex trait. During this process a change in one gene will often influence changes in other genes. A unique model to study complex trait evolution is C4 photosynthesis. This pathway has evolved independently more than sixty times and confers an important adaptive benefit in certain environmental conditions, notably including hot dry climates. The evolution of C4 photosynthesis involves changes in many genes that each have a function in the ancestral C3 photosynthetic pathway, so the trait is not driven by evolution of any one novel gene. We compare protein and gene interactions in C4 plants and their non-C4 relatives in a network model to study the process of complex trait evolution while accounting for the trait architecture. Analysis of the network model allows us to study whether genes that have been exapted, or co-opted from ancestral networks for use in the C4 pathway, tend to have specific placements within the network architecture. We also use the model to compare rates of molecular evolution with gene connectivity in the network, which represents a non-sequence based measurement of gene copy redundancy. By comparing the gene networks of pairs of species on either side of independent C4 evolution events in the grasses, we hope to better understand the process of complex trait evolution.

344W Complex genetics cause and constrain fungal persistence in different parts of the mammalian body Martin Mullis[1], Caleb Ghione[1], Michael Lough-Stevens[1], Ilan Goldstein[1], Takeshi Matsui[2,3,4], Sasha Levy[2,3,4], Matthew Dean[1], *Ian Ehrenreich*[1] 1) Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA; 2) Joint Initiative for Metrology in Biology, Stanford, CA; 3) SLAC National Accelerator Laboratory, Menlo Park, CA; 4) Department of Genetics, Stanford University, Stanford, CA

Determining how genetic polymorphisms enable certain fungi to persist in mammalian hosts can improve understanding of opportunistic fungal pathogenesis, a source of substantial human morbidity and mortality. We examined the genetic basis of fungal persistence in mice using a cross between a clinical isolate and the lab reference strain of the budding yeast *Saccharomyces cerevisiae*. Employing chromosomally-encoded barcodes, we tracked the relative abundances of 822 genotyped, haploid segregants in multiple organs over time and performed linkage mapping of their persistence in hosts. Detected loci showed a mix of general and antagonistically pleiotropic effects across organs. General loci showed similar effects across all organs, while antagonistically pleiotropic loci showed contrasting effects in the brain and the kidneys, liver, and spleen. Persistence in an organ required both generally beneficial alleles and organ-appropriate pleiotropic alleles. This genetic architecture resulted in many segregants persisting in the brain or in non-brain organs, but few segregants persisting in all organs. These results show complex combinations of genetic polymorphisms collectively cause and constrain fungal persistence in different parts of the mammalian body.

345W Exploiting the natural diversity of *Caenorhabditis elegans* to discover chemical actuators of the nervous system *Emily Fryer*[1], Hodan Farah[1, 2], Sujay Guha[2], Lucero Rogel[2], Tessa Logan-Garbisch[2,3], Ehsan Rezaei[2], Iris Mollhoff[1,2,4], Adam Nekimken[2], Angela Xu[1], Sylvia Fechner[2], Alakananda Das[2], Jason Casar[2], Shaul Druckmann[5], Lauen O'Connel[4], Thomas R. Clandinin[5], Seung Y. Rhee[1], Miriam B. Goodman[2] 1) Plant Biology, Carnegie Institution for Science, Stanford, CA; 2) Molecular and Cellular Physiology, Stanford University, Stanford, CA; 3) Neuroscience Graduate Program, Stanford University, Stanford, CA; 4) Biology, Stanford University, Stanford, CA; 5) Neurobiology, Stanford University, Stanford, CA

Plants have evolved complex chemical strategies to communicate with neighboring plants, prevent infection, repel herbivores and attract pollinators. Many of these compounds (called specialized metabolites (SMs) have medicinal properties, which humans have exploited before the advent of modern pharmacology. Identifying new drug candidates among the myriad of plant SMs continues to be a strategy in drug development. Identifying bioactive SMs requires efficient screening methods and new target identification strategies. Given their coevolutionary history with plants, we surmise that nematodes have the ability to detect many plant SMs and we hypothesize that laboratory studies of nematode behavior can identify plant SMs with therapeutic potential. With a fast generation time, richly annotated genome and well documented chemosensory behaviors, C. elegans is an ideal model system for achieving this goal. With ~1453 GPCRs encoded in its genome, C. elegans is a rich source that we can explore using high-content experimental methods, complemented with molecular and quantitative genetic methods to identify targets of plant SMs.

The Neuroplant Project developed a chemotaxis screening platform that efficiently screens 100s of plant SMs against multiple C. elegans strains. Using this platform, we identified 19 compounds that induce attraction/repulsion in the animal, two of which are known neuromodulators. To link these compounds to their receptors, we are mapping the chemosensory neurons that are needed for chemical attraction/repulsion in the laboratory strain, N2 (Bristol); null alleles of the candidate receptors expressed in those neurons are then screened for loss of behavioral responses. To complement these approaches we are leveraging the natural diversity of C. elegans and the genomic resources made available by the CeENDr project (Cook et al. 2016). As proof of concept we are using our platform to conduct chemotaxis assays with the CB4856 strain (Hawaii) against a panel of compounds known to elicit attraction or repulsion in the N2 lab strain. We plan to extend this work to include 10 additional divergent strains curated by CeNDR to represent the genomic variation among wild C. elegans strains. We will present results for N2 against a panel of 96 compounds and pilot studies comparing the sensitivity of N2 and CB4856 to a set of established chemosensory attractants or repellants.

Funding
Wu Tsai Neuroscience Big Ideas

346T Relationships between germline mutation rates and reproductive success in the collaborative cross mice *Alexis Garretson*[1,2], Beth Dumont[1,2] 1) The Jackson Laboratory for Mammalian Genetics, Bar Harbor, ME; 2) Graduate School of Biomedical Sciences, Tufts University, Boston, MA

The mutation rate is a critical genomic parameter, as germline mutations are the cause of heritable genetic disorders and the ultimate source of evolutionary novelty. The germline mutation rate is a complex trait, but we know little about how it is modified by heritable genetic features. In humans, elevated mutational burdens have been tentatively linked to reproductive capacity, with higher mutation rates associated with lower reproductive success and earlier age at last birth. However, the phenotypic effects of natural variation in mutation rates are incompletely understood. We expect elevated mutation rates in mice will be associated with lower breeding success and more rapid decline in reproductive capacity during aging, and we hypothesize shared genetic mechanisms related to DNA damage repair. Here we harness the breeding funnel design of the Collaborative Cross (CC) mouse population, analogous to a mutation accumulation experiment, to study the accumulation of mutations over ~30 generations of organized inbreeding and evaluate associations between mutation rate and phenotypic diversity. Using a powerful pan-genome, graph-based variant caller applied to publicly available whole-genome sequences for 69 completed CC strains, we comprehensively identify *de novo* mutations specific to each CC line and absent from the eight founder strains. Overall, we identify ~11,000 DNMs, including many with projected functional effects. The per generation rate of mutation accumulation varies 11-fold among strains, above simulated expectations. These findings suggest the presence of mutation rate modifiers segregating among the 8 CC founder strains. To explore the broader phenotypic consequences of this diversity, we assessed correlations between strain-level mutation rates and trait values, focusing on phenotypes related to reproduction and fertility. We find a negative relationship between reproductive capacity metrics, including the average litter size, the interbirth interval, and the age at the first litter. We also find a negative relationship between mutation rates and reproductive capacity, with higher mutation rates associated with a more rapid increase in the interbirth interval and a more rapid decrease in the litter size as the mice age. Overall, we demonstrate that germline mutation rates in healthy mice are related to reproductive aging and may suggest shared genetic mechanisms underpin variation in mutation rate and fertility.

347T The geography of GWAS: Genome-wide association mapping within a local *Arabidopsis thaliana* population more fully reveals the genetic architecture for defensive metabolite diversity *Andrew Gloss*[1,2], Amélie Vergnol[2], Timothy Morton[2], Peter Laurin[2], Fabrice Roux[3], Joy Bergelson[1] 1) New York University, New York, NY, USA; 2) University of Chicago, Chicago, IL, USA; 3) LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France

A paradoxical finding from genome-wide association studies (GWAS) in plants is that variation in metabolite profiles typically maps to a small number of loci, despite the complexity of underlying biosynthetic pathways. This discrepancy may partially arise from limitations presented by geographically diverse mapping panels. Widespread characteristics of metabolic pathways that impede GWAS by diluting the additive effect of a causal variant, such as allelic and genetic heterogeneity and epistasis, often increase in severity with the geographic range of the mapping panel. We hypothesized that a population from a single locality would reveal an expanded set of associated loci. We tested this in a French Arabidopsis thaliana population (< 1 km transect) by profiling and conducting GWAS for glucosinolates, a suite of defensive metabolites that have been studied in depth through functional and genetic mapping approaches. For two distinct classes of glucosinolates, we discovered more associations at biosynthetic loci than previous GWAS with continental-scale mapping panels. Candidate genes underlying novel associations were supported by concordance between their observed effects in the TOU-A population and previous functional genetic and biochemical characterization. Our findings highlight how local populations complement geographically diverse mapping panels to reveal a more complete genetic architecture for complex traits.

348W Understanding the local and global structure of pleiotropy using a yeast cross *Shreyas Gopalakrishnan*, Artur Rego-Costa, Eliot Fenton, Michael Desai Harvard University, Cambridge, MA

Pleiotropy, the phenomenon where a single locus affects multiple traits, is an important feature of the genetic architecture of traits. Current descriptions of pleiotropy have been limited by the poor spatial resolution of causal loci and the difficulty in measuring a large number of traits. In this study, we overcame these limitations by using DNA barcode sequencing based fitness assays to measure the fitness of 100,000 F1 segregant genotypes from a yeast cross in 111 different environments. These 111 environments consisted of a set of perturbation gradients (e.g. a temperature gradient) imposed on four very different base environments and thus contained both similar ("local" pleiotropy) and dissimilar environments ("global" pleiotropy). The two major aims of this study are to determine if the effect of a perturbation is the same in the different base environments and if the fitness variation in many environments can be explained by variation along a small number of dimensions or "core phenotypes". In a pilot study using a smaller number of F1 segregant genotypes, we found that perturbation gradients showed two major patterns of correlation: clustering by base environment and clustering by gradient concentration at high concentrations. We identified causal loci responsible for fitness variation by quantitative trait locus (QTL) mapping; these causal loci were jointly mapped across environments to improve spatial resolution. Preliminary joint QTL mapping identified several genes with known fitness effects and causal loci whose effect sizes were correlated with the gradient concentration. This work provides a detailed description of pleiotropy across a variety of laboratory environments and infers a lower-dimensional space of "core phenotypes" that has good predictive power and provides biological insight into the genetic architecture of fitness.

349W Identification of drought-adaptive QTL underlying variation in root system architecture in *Zea mays* *Kirsten Hein*, Patrick Woods, Jack Mullen, John McKay Colorado State University

Complex phenotypes are influenced by genetic variation, environmental differences, and genotype-by-environment interactions (G×E). These interacting factors and their relative contributions are therefore of critical importance in understanding complex traits, but at the same time they are challenging to study comprehensively. In part, this challenge stems from the necessary scale: experimental studies of the interaction of genotype and environment must be able to manipulate both factors at a large enough scale to provide the power to detect the relatively small effects of most loci underlying complex traits. Plants, including crops are ideal systems in that thousands of full or half siblings can easily be replicated across multiple environments. In addition, in crop species such as maize, the G×E variance for yield is larger than the corresponding genotypic main effect variance. Because it is predicted that the impact of climate change on drought frequency will impose significant economic cost in the U.S., especially in the southwest and Rocky Mountain states, it is imperative to find solutions to maintain or increase agricultural productivity in ways that ameliorate rather than exacerbate climate change. Enhanced root systems with deeper architecture are predicted to improve seasonal water-use efficiency, predominantly under drought conditions. To understand the genetic basis for root system architectural variation as it relates to crop production and adaptation to target environments, it is imperative to perform phenotyping under agronomically applicable field conditions. The goal of this study is to evaluate how genes interact with droughted and well-watered environments to create complex root phenotypes in the maize (*Zea may*s L.) system. Quantitative root measurements were collected from small plot field trials of 380 inbred lines of maize across two levels of the environmental factor soil moisture. Genome-wide association (GWAS) identified 85 SNPS significantly associated with root traits, including 12 SNPs that show significant G×E across soil moisture. We then selected biparental recombinant inbred lines that were segregating variation at the genomic regions and gene models identified in the GWAS analysis. Utilizing a computational-based framework proposed by Wen, Pique-Regi, and Luca (2017), we will integrate significant molecu-

lar QTLs identified through QTL analysis with the root trait-associated genetic variants to evaluate the enrichment and colocalization of both types of association signals. The result from the integrative analysis has the potential to address more biologically relevant hypotheses by reducing the list of genes within the associated gene set to those genes with the greatest contribution to the overall trait variability using available gene annotation and synteny data.

350T Antibiotic treatment affects the effect sizes of spontaneous mutations on bacterial population-growth characters *Wei-Chin Ho*, Jadon Gonzales, Michael Lynch Arizona State University

The nature of spontaneous mutations, including rates and effect sizes, largely determines genotypic and phenotypic evolution dynamics. Therefore, studying how environmental factors impact the mutations is critical for understanding the organismal evolution in different environments. While the mutation rates have been found to be plastically different in various environments, whether the effect sizes of mutations are so remains unknown. To answer this question, for ~80 lines of *Escherichia coli* that had accumulated spontaneous mutations in rich medium with or without the antibiotic norfloxacin, we measured their growth curves in their home environment using the microplate spectrophotometer. By analyzing the phenotypic effects and the numbers of mutations among these mutation-accumulated lines, we quantified mutational effect sizes for two traits: maximum growth rate and carrying capacity. The results show that, while the mutational effects of both traits are positively correlated in each home environment, norfloxacin significantly aggravates the mutational effects of maximum growth rates but not the carrying capacities. To further dissect the source of the observed effect-size changes, we measured and analyzed the same traits in the same mutation-accumulated lines in their non-home environment. Besides the direct environmental impact, the results show that the altered spectra of spontaneous mutations in different environments also influence the observed effect sizes. In addition, the mutational correlation of two traits becomes insignificant in the non-home environment. To sum up, our data demonstrate complex context-dependent behaviors of mutational effects and mutaional correlations. These results will enhance our knowledge of genotype-environment interaction, quantitative trait evolution, and microbial responses to antibiotics.

351T Elucidating the patterns of pleiotropy and its biological relevance in maize *Merritt Khaipho-Burch*[1], Taylor Ferebee[2], Anju Giri[3], Guillaume Ramstein[3,4], Brandon Monier[3], Emily Yi[3], M. Cinta Romay[3], Edward Buckler[1,3,5] 1) Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY; 2) Department of Computational Biology, Cornell University, Ithaca, NY; 3) Institute for Genomic Diversity, Cornell University, Ithaca, NY; 4) Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark; 5) USDA-ARS; Ithaca, NY

Pleiotropy has been shown to have effects on traits such as flowering time, leaf architecture, and inflorescence morphology in maize. However, the genome-wide impact of true biological (or horizontal) pleiotropy across all maize phenotypes is largely unknown. Here we investigated the extent to which true biological pleiotropy impacts phenotypes within maize through GWAS summary statistics reanalyzed from previously published physiological, metabolic, and expression phenotypes across the Nested Association Mapping population (US-NAM) and Goodman Association Panel. Through phenotypic saturation of 120,625 traits in maize, we obtained over 480 million significant quantitative trait nucleotides and estimated how pleiotropic each region was in the genome. We then assessed the relationship between pleiotropy and numerous biological features such as gene expression, the prevalence of open chromatin, sequence conservation, and enrichment for gene ontology terms using random forest models. We find very little relationship between pleiotropy and these variables compared to permuted pleiotropy values. Thus, we hypothesize that biological pleiotropy is not a common phenomenon in maize; however, mediated or vertical pleiotropy may be. We suggest that natural selection on large standing natural variation in maize populations will remove deleterious or large-effect variants, leaving the prevalence of biological pleiotropy relatively low. We recommend that the maize and surrounding plant science community accurately describe the types of pleiotropy under investigation and robustly test hypotheses of pleiotropy before claiming its widespread prevalence and causality.

352W The quantitative genetic basis of tolerance to environmental change during early embryogenesis in *Drosophila melanogaster* Sumaetee Tangwancharoen, *Brent Lockwood* University of Vermont

Despite the paradigm of environmental robustness in developmental biology and decades of research into the genetics of developmental traits, relatively little is known about the genetic basis of tolerance to environmental change during embryonic development. We used introgression and selection to uncover the genomic basis of enhanced embryonic heat tolerance in neotropical lines of *Drosophila melanogaster*. After 16 generations of introgression and selection, we mapped embryonic heat tolerance with pooled whole-genome resequencing of 12 replicate crosses. This allowed us to map the genomic basis of embryonic heat tolerance to an average of 640 kb per quantitative trait locus (QTL). We found embryonic heat tolerance to have a complex genetic basis, with a total of 10 unique QTL that represented mapped regions on every chromosome. Two of these QTL were repeatedly selected in multiple crosses and mapped to adjacent

regions on chromosome 2R. These two regions were non-overlapping and consisted of 54 genes known to be expressed in early embryos, notably genes involved in the electron transport chain and response to oxidative stress. Overall, our data suggest that while embryonic heat tolerance is influenced by loci across the genome, key genes involved in redox balance may be critical for the maintenance of embryogenesis in the face of environmental change. This result corroborates recent reports on the important role of redox biochemistry in embryogenesis. Further, our study extends previous work in developmental genetics in *Drosophila* by characterizing the quantitative genetics of an ecologically relevant developmental trait in natural populations.

353W Why most GWAS hits are not eQTLs *Hakhamanesh Mostafavi*, Jeffrey Spence, Sahin Naqvi, Jonathan Pritchard Stanford University

Most findings in genome-wide association studies (GWAS) of complex traits point to non-coding genetic variants with putative gene regulatory effects. However, currently identified expression quantitative trait loci (eQLTs) explain only a small fraction of the GWAS signals. While this lack of overlap may at first appear to be counterintuitive, we wondered if this might be a natural consequence of population-genetic forces such as natural selection. By analyzing GWAS hits for complex traits in the UK Biobank and cis-eQTLs from the GTEx consortium, we show that these assays systematically discover different types of genes and variants: eQTLs cluster strongly near transcription start sites, while GWAS hits do not. Genes near GWAS hits are enriched in numerous functional annotations, are under strong selective constraint and have a complex regulatory landscape across different tissue/cell types, while genes near eQTLs are depleted of most functional annotations, show relaxed selection, and have a simpler regulatory landscape. We describe a model to understand these observations. Specifically, modeling selection on complex traits, we demonstrate that selection disproportionally hampers the discovery of functionally relevant eQTLs, particularly eQTLs regulating genes that contribute most to trait heritability. More broadly, our work provides insight into the utility of intermediate phenotypes for explaining genetic effects on complex traits.

354T Long reads facilitate testing of allele specific expression and estimation of cis- and trans-variance in QTL regions Patricka Williams-Simon[1], Adalena Nanni[2,3], Alison Morse[2,3], Hayes Oken[1], Camille Oster[4], Elizabeth King[4], Paul Schmidt[1], *Lauren McIntyre*[2,3] 1) Department of Biology, University of Pennsylvania, Philadelphia, PA; 2) Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL; 3) University of Florida Genetics Institute, University of Florida, Gainesville, FL; 4) Division of Biological Sciences, University of Missouri, Columbia, MO

A combination of quantitative trait locus (QTL) mapping and allele specific expression analysis was developed and used to identify candidate genes for thermo-tolerance, a quantitative trait, in Drosophila melanogaster. QTL analyses for this trait identified intervals containing genes associated with the trait. Four RIL lines (2 high and 2 low) were selected and crossed to a tester line to assay allele specific expression. We used long reads (Oxford Nanopore) to quantify expression in the F1 and identified loci with allele specific expression. Long reads are highly efficient with more than 70% discrimination between alleles in the same species with very little pre-processing needed as errors in the reads do not affect allele assignment in most cases. We compared allele imbalance (AI) between the two high lines and between the two low lines within the QTL region. Inside the QTL the alleles are shared between the two high lines (low lines) and similar AI indicates a shared *cis*-regulatory effect, as expected based on the QTL. If the two lines with the same QTL allele differ in AI this indicates the presence of a trans-regulatory factor that differs in the two lines. In addition to the analyses within the QTL, the remainder of the genome can be explored for *cis*, and *cis by trans* interactions in AI by grouping alleles using marker loci.

355T Dose Response Modeling of *In Vitro* High Content Screening Identifies Genetic Variants Modulating Sensitivity to Monomethylarsonous Acid Exposure *Callan O'Connor*[1,2], Gregory Keele[1], Whitney Martin[1], Daniel Gatti[1], Ron Korstanje[1], Gary Churchill[1], Laura Reinholdt[1] 1) The Jackson Laboratory, Bar Harbor, ME; 2) Tufts University, Graduate School of Biomedical Sciences, Boston, MA

The mechanisms underlying variation in susceptibility to environmental exposures are generally not well understood which limits our ability to accurately predict risk. Our goal is to establish a model for unbiased gene-by-environment exposure studies that can be used to both identify and validate genetic modifiers of response to toxicant exposure. We exposed 215 primary fibroblast cell lines derived from a genetically diverse mouse population to varying concentrations of monomethylarsonous acid (MMA). We assayed the cells using image-based high content cellular screening (HCS) to capture cellular morphometric features with the goal of identifying heritable, dose-response features for genetic mapping. We found that several experimental factors exacerbate non-linear modeling for these types of data including cellular heterogeneity, imaging artifacts, and cell culture conditions. To mitigate these problems, we introduced changes to our experimental design and used a two-stage estimation approach to capture MMA-exposure responses. First, we

fit the cellular features data to a four-parameter log-logistic dose-response model (DRM) for replicates from the same individual (n = 4). Second, we extracted DRM parameters, which we then modeled using a linear mixed effects model to correct for technical sources of variation. This resulted in 1431 corrected model fit parameter summaries representing slope, asymptotes, and critical effect size estimates (i.e. EC 50) from >300 cellular features. We mapped 10 suggestive quantitative trait loci (QTL) (genome-wide $p < 0.05$). The slope parameter of the fibroblast cell area response to MMA had the strongest QTL (LOD score > 9) at chromosome 10 (82.9 Mbp) demonstrating that 1) changes in cell size occur in response to MMA exposure, and 2) that genetic variation has an impact on the rate of this change during exposure. Within the QTL region, we identified *Txnrd1* as a candidate gene which is highly expressed in fibroblasts and has known interactions with arsenic. We identified genetic variants (SNPs) in the 3'-UTR of *Txnrd1*, notably with alleles specific to the NZO/H1LtJ (NZO) founder strain, associated with higher slopes for cell area. This region of *Txnrd1* regulates the incorporation of the amino acid selenocysteine (Sec) during TXNRD1 synthesis which is critical for the reduction potential of this enzyme. Our findings suggest that the NZO haplotype in the 3'-UTR of *Txnrd1* increases susceptibility of fibroblasts to MMA exposure. To test this hypothesis, we plan to perform parallel *in vitro* and *in vivo* assessments of MMA sensitivity in genetically diverse inbred mouse strains (Collaborative Cross) that possess the NZO allele at this locus. In conclusion, our current data demonstrates the utility of using cellular reference panels derived from genetically diverse populations to identify genetic modifiers of environmental toxicant exposure.

356W Life History and Stress Tolerances in Elevation Adapted Populations of *Drosophila melanogaster* *Camille Oster*, Elizabeth King University of Missouri

In the era of global climate change, not all environments are impacted equally and uniformly. Of ecosystems facing acute changes in climate, high elevation populations have unique constraints on range and adaptation. Researchers have shown evidence of range shift in *Drosophila* species to higher elevations, but thermoregulative behaviors may do little to avoid climate stressors in small arthropods. We expect adaptive trade-offs between climate stress tolerances and life history, but this intersection deserves further study for populations with distinct adaptations to elevation; here we investigate those complex traits.

We phenotyped life history and stress-tolerance traits in 8 *Drosophila melanogaster* populations from sub-Sarahan Africa. These lines, collected and inbred by Pool et al, originate from low-elevation Zambia and high-elevation Ethiopia, both with year-round tropical climates. First, we assayed lifespan and weekly measurements of fecundity in cages of ~1000 individuals with a two to one male to female ratio. Second, we measured starvation resistance by placing adults in non-nutritional agar vials. Finally, we phenotyped thermal tolerance using a Pelt-5 temperature controller in a ramping protocol from room temperature to incapacitation. Due to established sex differences in stress tolerances in *Drosophila spp*., we prioritized female flies in our assays. Here, we show the effects of genotype and population on this suite of traits and use a multivariate approach to show how these different traits relate to one another.

357W Many factors contribute to reproductive isolation between self-pollinating and outcrossing morning glory *Kate Ostevik*[1,2], Joanna Rifkin[2,3], Irene Liao[2,4], Mark Rausher[2] 1) University of California Riverside, Riverside, CA; 2) Duke University, Durham, NC; 3) University of Toronto, Toronto, ON; 4) University of California Los Angeles, Los Angeles, CA

Highly selfing plant species have repeatedly diverged from outcrossing relatives. Diverse mechanisms, some specific to the selfer-outcrosser context, can lead to reproductive isolation between selfing and outcrossing species, and multiple mechanisms may act in any speciation event. Here, we dissect two reproductive barriers between the highly selfing morning glory *Ipomoea lacunosa* and its mixed-mating sister species *I. cordatotriloba*. We find that the crossing barrier (failure of hybrid seed set) is complex, with contributions from barrier components both before and after fertilization and extensive parental sex asymmetry. Fertilization failure varies by cross direction and is partially affected by pollen and style size. We find strong evidence that fertilized seeds in interspecific crosses fail to mature, and that cross direction affects seed provisioning. Genetic mapping of the crossing barrier also reveals multiple factors with variable strengths contribute to cross incompatibility between these species, parental sex asymmetries are common, and imprinting in hybrid seed provisioning is likely to be a major component. In contrast, hybrid pollen sterility is genetically simple and overwhelmingly caused by two epistatically interacting loci. This study highlights the importance of carefully dissecting reproductive barriers into their components, which in this case are diverse and could have independent evolutionary histories.

358T Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits *Roshni Patel*[1], Shaila Musharoff[1,2], Jeffrey Spence[1], Harold Pimentel[3], Catherine Tcheandjieu[1,2], Hakhamanesh Mostafavi[1], Nasa Sinnott-Armstrong[1,2], Shoa Clarke[1,2], Courtney Smith[1], Peter Durda[4], Kent Taylor[5], Russell Tracy[4], Yongmei

Liu[6], Craig Johnson[7], Francois Aguet[8], Kristin Ardlie[8], Stacey Gabriel[8], Josh Smith[7], Stephen Rich[9], Jerome Rotter[5], Philip Tsao[1,2], Themistocles Assimes[1,2], Jonathan Pritchard[1], VA Million Veteran Program 1) Stanford University, Stanford, CA; 2) VA Palo Alto Health Care System, Palo Alto, CA; 3) University of California Los Angeles, Los Angeles, CA; 4) University of Vermont, Burlington, VT; 5) Lundquist Institute for Biomedical Innovation, Torrance, CA; 6) Duke University, Durham, NC; 7) University of Washington, Seattle, WA; 8) Broad Institute, Cambridge, MA; 9) University of Virginia, Charlottesville, VA

Despite the growing number of genome-wide association studies (GWAS), it remains unclear to what extent gene-by-gene and gene-by-environment interactions influence complex traits in humans. The magnitude of genetic interactions in complex traits has been difficult to quantify because GWAS are generally underpowered to detect individual interactions of small effect. Thus, despite the widespread use of the additive model in quantitative genetics, its applicability to human traits is yet to be determined. Here, we develop a method to test for genetic interactions that aggregates information across all trait-associated loci. Specifically, we test whether SNPs in regions of European ancestry shared between European and admixed African-American individuals have the same causal effect size. We hypothesize that, in African-Americans, the presence of genetic interactions will drive the causal effect sizes of SNPs in regions of European ancestry to be more similar to those of SNPs in regions of African ancestry. Because we focus on comparing regions of shared European ancestry in two different populations, our analysis is not biased by differences in LD structure between European and African ancestries. We apply our method to two traits: gene expression of 319 African-Americans and 499 Europeans in the Multi-Ethnic Study of Atherosclerosis (MESA) and low-density lipoprotein cholesterol (LDL-C) of 72K African-Americans and 298K Europeans in the Million Veteran Program (MVP). We find significant evidence for genetic interactions in our analysis of gene expression; for LDL-C, we observe a similar point estimate although this is not significant, likely due to lower statistical power. These results underscore the role of genetic interactions in human complex traits and highlight the limitations of the additive model.

360W Quantitative genetic analysis of pathogenic response to SARS-CoV-2 and other coronaviruses in an F2 cross of Collaborative Cross strains *Ellen Risemberg*, Sarah Leist, Alexandra Schaefer, Will Valdar, Martin Ferris, Ralph Baric University of North Carolina at Chapel Hill

Severe acute respiratory syndrome (SARS) is a viral respiratory illness that emerged in China in 2003 and became the first of three severe epidemics caused by zoonotic coronaviruses over the next 20 years. The ongoing COVID-19 pandemic and likelihood of future coronavirus outbreaks motivates greater understanding of host genetic factors contributing to variation in severity of coronavirus disease. Genetically diverse Collaborative Cross (CC) strains provide a powerful system for studying these factors and identifying mechanisms by which these genetic differences affect viral disease and pathogenesis. In this study, CC strains CC044/Unc (hereafter CC044) and CC006/TauUnc (hereafter CC006) were chosen for a genetic mapping experiment because preliminary data shows divergent phenotypic outcomes upon infection with SARS-CoV-1, including differential weight loss and immune cell infiltration into the lungs following infection. We perform a genome-wide quantitative trait loci (QTL) analysis in F2 offspring of a CC044 x CC006 cross to identify loci associated with variation in response to coronavirus infection. In addition to standard QTL mapping, we examine polygenic heritability of phenotypes, and how genetic correlations between phenotypes vary by pathogen (SARS-CoV-1, SARS-CoV-2, HKU3-CoV, saline), and evidence for the existence of substantial genetic effects on the phenotypic variability (vQTL). Here we report ongoing progress, including the identification of a QTL for infection-induced weight loss on chromosome 9, the homologous region of which has been reported to be associated with severe COVID-19 in humans.

359T Unravelling the Genetic Architecture of Rolling Behavior in the Domestic Pigeon (*Columba livia*) *Atoosa Samani*, Emily Maclary, Michael Shapiro The University of Utah

Hereditary rolling or tumbling in the domestic pigeon (*Columba livia*) is a backward somersault behavior that is stimulated by attempting to fly. Rolling has excited the curiosity of scholars for centuries; Darwin describes rolling as "…one of the most remarkable inherited habits or instincts ever recorded". Rolling is progressive: it does not present itself until a few weeks after fledging and becomes more severe with age. Rolling affects locomotion, but no anatomical anomalies are known to be associated with it. Roller pigeons walk, eat, and breed normally, suggesting that rolling is a specific context-dependent behavior and not a generalized physiological disorder. Rolling is recessive and highly heritable, yet the molecular genetic basis remains unknown. Therefore, rolling offers a unique opportunity to discover the molecular basis of a complex yet genetically tractable behavioral phenotype. Using a combination of quantitative trait locus (QTL) mapping in a laboratory intercross and comparative genomics among pigeon breeds, we identified several loci associated with rolling behavior. Although rolling is a polygenic trait, one major QTL explains 61% of the phenotypic variance in the laboratory cross. Comparisons between the resequenced genomes of rollers and non-rollers confirm the polygenic nature of this behavior, and that quantitative genetic and GWAS approaches yield overlapping results. Dissection of the

candidate loci at the gene level will deepen our understanding of the molecular basis of involuntary and task-specific movement disorders and other progressive vertebrate behaviors.

361W The genetics of pathogen and microbiome control in the switchgrass leaf *Acer VanWallendael*[1], Gian M. N. Benucci[1], Pedro B. da Costa[1], Linnea Fraser[1], Avinash Sreedasyam[2], Felix Fritschi[4], Thomas Juenger[3], John Lovell[2], Gregory Bonito[1], David Lowry[1] 1) Michigan State University, East Lansing, MI; 2) HudsonAlpha Institute of Biotechnology, Huntsville, AL; 3) University of Texas, Austin, TX; 4) University of Missouri, Columbia, MO

Leaf fungal microbes can be fundamental drivers of host plant success, as they consist of pathogens that devastate crop plants as well as taxa that enhance nutrient uptake, discourage herbivory, and antagonize pathogens. In a replicated diversity panel of biofuel switchgrass, we quantified genetic and environmental variation in leaf fungal relationships, both for the whole microbiome and for a specific pathogen, leaf rust. While fungal colonization of the leaf varies over space and time, we uncovered genome-wide associations (GWAs) with several informative loci. In particular, three cysteine-rich receptor-like kinase genes (crRLKs) were linked to a genetic locus associated with microbiome structure. Since each of these genes is consistently upregulated in switchgrass genotypes typically more susceptible to fungal disease, we conclude that they may play a central role in the plant's response to pathogens. Response to leaf rust is polygenic and environmentally sensitive, but resistance alleles are associated with higher biomass, indicating that breeding for rust-resistant plants will benefit growth without trade-offs in the absence of rust. Switchgrass response to fungal colonists is complex and variable, but an experimental design that accounts for variation over space and time allows for greater definition on genetic loci underlying fungal interactions.

362T Chimeragenesis: a method for generating, selecting, and phenotyping gene variant libraries in yeast *Cory Weller*, Meru Sadhu National Human Genome Research Institute, NIH, Bethesda, MD

Thorough study of genetic diseases requires understanding how amino acid substitutions influence phenotype. Manually engineering individual substitutions can provide insight into small numbers of variants, but assessing many variants requires high-throughput methods. Here, we describe a method that improves upon deep mutational scanning by allowing for high-throughput selection, sequencing, and phenotyping of variants within a single pool. First, we prepare a plasmid library via assembly of variant-generating repair templates amplified from a cost-effective synthesized oligonucleotide pool. Each plasmid, when transformed into a yeast cell expressing Cas9, directs the generation of a specific predetermined variant. We transform our plasmid library into a yeast strain that possesses two copies of our gene of interest separated by a counter-selectable marker. A given repair template induces a large-scale deletion between the gene copies, removing the counter-selectable marker while recapitulating a normal-length gene of interest. Variants can be generated either by virtue of the two gene copies exhibiting amino acid substitutions (thereby generating a recombinant gene) or by introducing novel codons via the repair template. This deletion-based method greatly simplifies CRISPR-Cas9 editing by requiring a single, shared guide RNA for the entire variant library. It also allows for simple selection of mutants and facilitates pooled tracking of variant abundance (inferred from abundance of repair templates in pooled short-read sequencing). This technique will greatly accelerate the generation and assessment of variant effects, providing insight into protein function and the genetic underpinnings of disease.

363T Dose-response and quantitative genetic analyses reveal a complex genetic basis underlying susceptibility to diverse toxicants in *C. elegans* *Samuel Widmayer*[1], Timothy Crombie[1], Janneke Wit[1], James Collins[1], Sophia Gibson[1], Joy Nyaanga[1], Emily Koury[1], Robyn Tanny[1], Erik Andersen[1,2] 1) Department of Molecular Biosciences, Northwestern University, Evanston, IL; 2) Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Chicago, IL

Toxic exposure is a known risk factor in the onset of many human diseases, but the contributions and abundance of specific genetic variants to xenobiotic-induced disease risk at the population level are unknown. Because of the limited power and scale of toxicological assessments across genetically diverse human subjects, a tractable model system is required to characterize hazard levels of xenobiotic compounds and identify genes linked to susceptibility. Toxicological assessments using *C. elegans* have revealed previously unknown and translational features of xenobiotic metabolism, but investigations of natural variation in these responses are extremely limited. We measured the susceptibility of eight genetically diverse *C. elegans* wild strains to an array of toxicants, including several heavy metals, mitochondrial poisons, organophosphate insecticides, fungicides, herbicides, and one flame retardant using dose-response assessments. We measured phenotypic responses to each compound by adapting a high-throughput fitness assay using the Molecular Devices ImageXpress Nano automated imaging microscope and developed open-source software to extract and analyze animal morphology measurements from images. Wild strains varied significantly in susceptibility to most compounds and exhibited variable lowest observed adverse response levels, motivating us to search for quantitative trait loci (QTL) associated with differential responses. To accomplish this search, we measured phenotypic responses to a single dose

of each compound across a panel of 200 *C. elegans* strains and performed genome-wide association mappings. These analyses revealed dozens of xenobiotic response loci with measurable effects on population-wide toxicant susceptibility and genetically correlated responses to compounds with similar modes of action. We conclude that differential xenobiotic susceptibility among *C. elegans* strains is highly heritable and controlled largely by toxicant-specific genetic architectures. Future work will validate the effects of these QTL in complementary recombinant populations in order to characterize their modes of action and determine any conserved functions in diverse human populations.

364W Root Pulling Force Across Drought in Maize Reveals Genotype by Environment Interactions and *Phosphate Transporter 1-2a* as a Candidate Gene *Patrick Woods*, Jack Mullen, Kirsten Hein, John McKay Colorado State University

High-throughput, field-based characterization of root systems for hundreds of genotypes in thousands of plots is necessary for breeding and identifying loci underlying variation in root traits and their plasticity. We designed a large-scale sampling of root pulling force (RPF), the vertical force required to extract the root system from the soil, in a maize diversity panel under differing irrigation levels for two growing seasons. We then characterized the root system architecture of the extracted root crowns. We found consistent patterns of phenotypic plasticity for RPF for a subset of genotypes under differential irrigation, suggesting that root plasticity is predictable. Using genome-wide association analysis (GWAS), we identified 54 SNPs as statistically significant for six independent RPF measurements across two irrigation levels and four developmental timepoints. One interesting candidate gene identified for RPF under irrigated treatments was *Phosphate Transporter 1-2a* (*PHO1-2a*), a gene involved in phosphate transfer from the root to the shoot via the root stele. To functionally validate the effect of *PHO1-2a* on RPF we are conducting both screens of *PHO1-2a* mutants, and quantitative trait loci (QTL) mapping in a recombinant inbred line (RIL) population created from two closely related iodent maize lines (PH207 and L127) which possess the alternate alleles at the *PHO1-2a* SNP identified in the GWAS. Comparisons of RPF between mature field grown wild-type and one allele of *PHO1-2a* mutant plants during the 2021 field season show significant differences ($p < 0.05$) across numerous root system traits including RPF, root mass, and root area suggesting a true role of *PHO1-2a* in determining root architectural phenotypes. Root phenotypes of additional mutant alleles of *PHO1-2a* and the PH207 x L127 RIL population will be analyzed in the 2022 field season for further characterization of this gene's effect on root system architecture in maize.

365W Genetic and morphological basis of variation in pup vocalization behavior in deer mice *Maya Woolfolk*, Nicholas Jourjine, Sade McFadden, John Emory Sabatini, Hopi Hoekstra Department of Molecular & Cellular Biology, Department of Organismic & Evolutionary Biology, Museum of Comparative Zoology, Howard Hughes Medical Institute, Harvard University, Cambridge, MA

Infant vocalization is a behavior critical for eliciting care from parents across vertebrates; in many mammals, vocalizations also serve to establish social bonds between parents and offspring. Deer mice (genus *Peromyscus*) are a group of closely related but behaviorally diverse rodents that offer an opportunity to study natural variation in pup vocal behavior. We first designed and optimized a protocol to record pup vocalizations following isolation from their parents and analyzed these recordings using recently developed machine-learning tools for computational bioacoustics. We identified significant interspecific differences in temporal and acoustic features of pup cries, several of which evolved repeatedly between species. Next, by cross-fostering pups between species with different vocal behaviors, we found that these vocal differences are not significantly influenced by the postnatal environment, suggesting instead they likely have a strong genetic component. By measuring pup vocalizations in F1 hybrids, we found that some spectral and temporal features of vocalizations exhibit different patterns of dominance. To further investigate the genetic basis of this behavioral variation, we performed quantitative trait locus (QTL) mapping with 576 F2 individuals generated by intercrossing two sister species: *P. maniculatus* and *P. polionotus*. We first found that some acoustic features (e.g., pitch) and vocalization rate are uncoupled, suggesting that they are under separate genetic control. Indeed, we identified several loci significantly associated with species-specific differences in pup vocal behavior. In conjunction with this genetic approach, we also targeted the mechanistic basis of vocal variation by performing histological and morphometric analyses of the primary vocal organ, the larynx, which suggests there are potential contributions of laryngeal morphology to variation in acoustic features such as pitch. In this system, we are uncovering mechanisms underlying natural variation in pup vocal behavior to identify genetic and morphological bases of behavioral evolution.

366T Age and diet interact to shape body weight and lifespan of DO mice *Kevin Wright*[1], Andrew Deighan[2], Andrea Di Francesco[1], Adam Freund[1], Vladimir Jojic[1], Gary Churchill[2], Anil Raj[1] 1) Calico Life Sciences, LLC; 2) The Jackson Laboratory

It is largely unknown the degree to which age shapes additive genetic and genotype-environment interaction effects to shape quantitative traits. We used a linear mixed model to quantify age- and diet-dependent genetic contributions to a classic quantitative trait (body weight) measured throughout the life of Diversity Outbred female mice subject to

five dietary treatments. We observed that heritability of body weight declined with age under all diets, except the most extreme calorie restriction diet, which remained high throughout adulthood. We identified age- and diet-specific body weight loci and found all allelic effects were consistently positive or negative at one age and neutral at other ages. We found no evidence of allelic effect trade-offs with respect to age. We measure the association between these body weight effect alleles and animal lifespan and conclude these results are inconsistent with predictions arising from Williams' (1957) antagonistic pleiotropy theory of aging.

367T Circulating polyunsaturated fatty acids and COVID-19: a prospective cohort study and Mendelian randomization analysis *Kaixiong Ye*[1, 2], Yitang Sun[1], Radhika Chatterjee[1], Akash Ronanki[1] 1) Department of Genetics, Franklin College of Arts and Sciences, University of Georgia, Athens, GA; 2) Institute of Bioinformatics, University of Georgia, Athens, GA

Background: Higher circulating polyunsaturated fatty acids (PUFAs), especially omega-3 ones, have been linked to a better prognosis in patients of coronavirus disease 2019 (COVID-19). However, the effects and causality of pre-infection PUFA levels remain unclear.

Objective: To investigate the observational and causal associations of circulating PUFAs with COVID-19 susceptibility and severity.

Design: We first performed a prospective cohort study in UK Biobank, with 20,626 controls who were tested negative and 4,101 COVID-19 patients, including 970 hospitalized ones. Plasma PUFAs at baseline were measured by nuclear magnetic resonance, including total PUFAs, omega-3 PUFAs, omega-6 PUFAs, docosahexaenoic acid (DHA), linoleic acid (LA), and the omega-6/omega-3 ratio. Moreover, bidirectional two-sample Mendelian randomization (MR) analyses were performed to examine the causal associations of eight individual PUFAs, measured in either plasma or red blood cells, with COVID-19 susceptibility and severity using summary statistics from existing genome-wide association studies.

Results: In the observational association analysis, total PUFAs, omega-3 PUFAs, omega-6 PUFAs, DHA, and LA were associated with a lower risk of severe COVID-19. Omega-3 PUFAs and DHA were also associated with a lower risk of testing positive for COVID-19. The omega-6/omega-3 ratio was positively associated with risks of both susceptibility and severity. The forward MR analysis indicated that arachidonic acid (AA) and docosapentaenoic acid (DPA-n3) might be causally associated with a lower risk of severe COVID-19, with OR (95% CI) per one SD increase in the plasma level as 0.96 (0.94, 0.99) and 0.89 (0.81, 0.99), respectively. The reverse MR analysis did not support any causal effect of COVID-19 on PUFAs.

Conclusions: Our observational analysis supported that higher circulating PUFAs, either omega-3 or omega-6, are protective against severe COVID-19, while omega-3 PUFAs, especially DHA, were also associated with reducing COVID-19 susceptibility. Our MR analysis further supported causal associations of AA and DPA-n3 with a lower risk of severe COVID-19.

368W Natural variation in *C. elegans* genomic defense mechanisms mediated by small RNAs *Gaotian ZHANG*, Erik Andersen Northwestern University

RNA interference is an evolutionarily conserved mechanism for endogenous gene regulation and defense against foreign RNA viruses. Natural variation in exogenous and endogenous RNA interference in *Caenorhabditis elegans* has been suggested to be genetically complex, but the underlying causal variants are largely unknown. Our recent work revealed gene expression variation and possible regulatory mechanisms across 207 genetically distinct wild *C. elegans* strains using genome-wide association mappings. We classified expression quantitative trait loci (eQTL) into local eQTL (located close to the genes that they influence) and distant eQTL (located further away from the genes that they influence). We also identified a diverse collection of genomic hotspots enriched for distant eQTL of multiple genes. Here, we investigated the role of small RNAs in eQTL hotspots. We performed gene set enrichment analysis on genes with eQTL in each hotspot and found genes enriched for targets of small RNAs. Using fine mappings and mediation analysis, we identified candidate variants and genes for each distant eQTL of small RNA targets. Our results showed genetic variants in the gene *eri-6* could underlie the expression variation of 10 genes. We found the genetic variant in the isoform *eri-6[e]* reduced the expression of *eri-6[e]* but elevated the expression of *eri-6[a-d]*, which encodes the endogenous RNA interference factor ERI-6/7. Elevated levels of ERI-6/7 promote the biogenesis of endogenous ERI-6/7–dependent small interference RNAs (siRNAs) and likely reduce expression of targets, including the 10 genes above. The ERI-6/7–dependent siRNAs primarily target recently acquired, duplicated genes, and pseudogenes with likely viral origins. Stronger suppression of these targets could provide extra protection from overexpression of endogenous viral elements and future infections by viruses closely related to endogenous retroviruses. Our results reveal the diversity in RNA interference pathways

among wild *C. elegans* strains and provide evidence of the role of small RNAs in *C. elegans* defense mechanisms.

369W Amplification is the primary mode of gene-by-sex interaction in complex human traits *Carrie Zhu*[1,2], Arbel Harpak[1,2] 1) Department of Integrative Biology, University of Texas at Austin, Austin, TX; 2) Department of Population Health, Dell Medical School, Austin, TX

Sexual dimorphism is observed in many complex traits and diseases and is suspected to be in large part due to widespread gene-by-sex interactions (GxSex). To date, empirical evidence for GxSex in GWAS data—with attempts focused on identifying imperfect genetic correlations between the sexes— has been underwhelming. We hypothesized that GxSex may indeed be pervasive, but largely missed by current approaches, because it acts primarily through sex differences in the magnitude of many genetic effects ("amplification"), regulated by a shared cue such as a sex hormone, rather than differences in the identity of causal variants or the direction of their effect ("correlation"). To test this hypothesis, we inferred the genetic covariance structure between males and females across dozens of traits in the UK Biobank, and decoupled GxSex signals into correlation and amplification effects. Amplification was a much more pervasive mode of GxSex. For example, we estimate that 38% of causal variants have a greater effect on urate levels in females than males. In addition, we investigated whether testosterone levels, a continuous measure, may underlie the observed mediation by binary biological sex. For many traits, notably in body mass related traits, testosterone levels are strongly associated with the magnitude of genetic effects in both males and females—but the association is different in magnitude and sign between the sexes. Finally, we developed a novel test of sexually-antagonistic selection linking GxSex signals from GWAS and allele frequency differences between males and females. Our test can help identify specific complex traits that are under contemporary sexually-antagonistic selection. In summary, our results suggest that the systematic amplification of genetic effects is a primary mode of GxSex that is crucial for understanding sexual dimorphism and may be germane for phenotypic prediction.

370T Genetic basis of variation in high sugar-induced diabetes-associated traits and development delay in *Drosophila* *Xuan Zhuang*[1], Fabio Morgante[2], Abaranjitha Muniyasamy[1], Mohan Acharya[1], Michael Ludwig[3], Soo Young Park [4], Yang Li[4], Matthew Stephens[5], Graeme Bell [4], Martin Kreitman[3] 1) Department of Biological Sciences, University of Arkansas, Fayetteville, AR; 2) Clemson Center for Human Genetics, Clement University, Clemson, SC; 3) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 4) Department of Medicine, University of Chicago, Chicago, IL; 5) Department of Human Genetics, University of Chicago, Chicago, IL

Drosophila is a well-established model for investigating complex traits and many human diseases. It provides powerful tools for dissecting the contributions of both genes and environment on development and metabolism. We developed a *Drosophila* model to study high sugar diet (HSD) induced Type 2 diabetes (T2D) associated traits in large populations with different genetic backgrounds. We examined HSD-induced phenotypes in larvae and in adults across a subset of *Drosophila* Genetic Reference Panel (DGRP). Flies on HSD display an increase in whole-body glucose and glycogen level, a decrease in developmental rate, survivorship, body weight, and longevity, compared with flies under a low sugar diet (LSD). The examined DGRP lines display a continuous and wide range of these phenotypes and large broad-sense heritability, suggesting great potential for quantitative trait loci (QTL) mapping. In the meanwhile, we developed a unique experimental system for genetic mapping named Drosophila Recombinant Populations (DRPs), which are consisted of 16 outbred advanced intercross populations (AIPs), each founded with 8 inbred DGRP lines. The DRPs provide about 70,000 genotypically distinct flies that allow us to apply the HSD-induced T2D model to each individual to investigate the genetic architecture of these complex traits. We used one of the HSD-induced traits, namely developmental delay, to perform a bulk segregant mapping analysis of extreme phenotypes. We developed a computational pipeline with a Hidden Markov model (HMM) to impute the whole genome of each fly based on their low-coverage sequenced genomes and the available high-coverage founder genomes. Genome-wide association studies (GWAS) identified 76, 2009, and 373 polymorphisms at $p<10^{-5}$ for the analysis of HSD, LSD, and G x E (gene by diet) respectively. Gene ontology (GO) analysis indicates predominant enrichment for genes involved in some hormones and ketones cellular metabolic processes among HSD candidates, and reproduction and mediator complex among LSD candidates. RNAi validation studies of the top candidates (Cyp9b2 and CG15088 from GxE analysis, wun and Ten-m from HSD analysis) indicated their moderate but still significant effects in increasing the development rate on HSD. This study provides a broad survey of diabetes-associated traits in HSD-induced T2D flies, and further dissects the genetic architecture of development time on diets of different sugar concentrations, and the interaction of gene and diet.

371T Identifying the genetic factors in natural genome backgrounds that modulate essential phenotypic outcomes using *C. elegans* *Afiya Razia Chida*[1,2], Victoria Rodrigues Alves Barbosa[1,2], Xiao Li[1,2], Tatiana Maroilley[1,2], Francesca Jean[1,2], Tahsin Hassan Rahit[1,2], Andrew Galbraith [1,2], Filip Cotra[1,2], Larisa Oncea[1,2], Maja Tarailo-Graovac[1,2] 1) Departments of

Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; 2) Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Canada

Essential genes are vital for cellular development, metabolism, and reproduction; mutations with detrimental effects in these genes can cause severe defects. Essentiality varies based on different conditions, including genetic background. Our aim is to search for genetic factors that modulate gene essentiality in six diverse *Caenorhabditis elegans* natural isolate backgrounds - CB4856 (Hawaii, USA), JU1400 (Seville, Spain), AB1 (Adelaide, Australia), GXW1 (Wuhan, China) and KR314 (Vancouver, Canada), and N2 (Bristol, UK). The plasticity of essentiality might be influenced by genetic modifiers - variants that affect the causative gene by ameliorating or exacerbating a trait or disease. Our research focusses on identifying modifiers of two essential genes – *mat-1* (Metaphase-to-Anaphase Transition defect), which is involved in cell division, and *cgh-1* (Conserved Germline Helicase), which is important in oocyte development and gametogenesis. Previous studies with loss of CGH-1 and MAT-1 by RNA interference have suggested the presence of genetic modifiers in natural isolates (CB4856 and N2). To further investigate that, we individually knocked in conditional lethal temperature sensitive alleles of *cgh-1(tn691)* and *mat-1(ye121)* using CRISPR/Cas9 in all six isolates. At the permissive temperature (15°C), we observed about 100% hatch-rate in all isolates. However, at the restrictive temperatures (23°C & 25°C) both *cgh-1(tn691)* and *mat-1(ye121)* show evidence of phenotypic variability in different natural isolate backgrounds, ranging from 100% embryonic lethal to viable, indicating the presence of modifiers. Thus, to determine the genetic factors modulating the essentiality, the strains were subjected to whole genome sequencing followed by advanced bioinformatics analysis. We have detected and annotated a spectrum of variants, including single nucleotide variants and more complex structural variants, such as translocations and inversions. To identify potential genetic modifiers among them, we use genetic interaction network analysis, as well as in house developed machine learning approaches. To test the candidate modifiers, we are using CRIPSR/Cas9. Studying genetic modifiers using *C. elegans* natural backgrounds may improve the genomic approaches in the discovery of genetic modifiers in human diseases. This is crucial for proper diagnosis, prognosis, and more precise patient management.

372W Genome-wide detection and quantification of genetic background effects using double-barcoded CRISPRi perturbations *Ilan Goldstein*[1], Joseph Hale[1], Takeshi Matsui[2,3], Kevin Roy[3,4], Lars M. Steinmetz[3,4,5], Ian M. Ehrenreich[1] 1) University of Southern California; 2) SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA; 3) Department of Genetics, Stanford University, Stanford, CA 94305, USA; 4) Stanford Genome Technology Center, Stanford University, Palo Alto, California, USA; 5) Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidlberg, Germany

The phenotypic effects of mutations often depend on the genetic backgrounds in which they occur. Our goal is to move towards a fundamental understanding of the genetic and molecular mechanisms producing these background effects using the budding yeast model system. Here, we will measure the fitness effects of ~20,0000 CRISPR interference knock-downs targeting ~5,400 genes across 14 genetically diverse haploid and diploid S. cerevisiae strains. We leverage a double-barcode sequencing strategy in which one barcode denotes a strain genotype and the second barcode denotes a gRNA. The relative fitness of each genotype-gRNA combination will be measured by sequencing the double barcodes for all strains growing in a single pooled competition. By including ≥20 barcodes per gRNA and ≥3 barcodes per strain, we employ a high degree of internal replication that provides the statistical power to detect gRNAs that have different phenotypic effects across strains. With these data, we will determine the prevalence, extent, and character of genetic background effects across diverse genetic perturbations and genotypic contexts.

373W Analysis of ~10,000 CRISPR interference perturbations in a yeast cross *Joseph Hale*[1], Ilan Goldstein[1], Takeshi Matsui[2,3,4], Martin Mullis[1], Kevin Roy[5,6], Lars Steinmetz[5,6,7], Sasha Levy[2,3,4], Ian Ehrenreich[1] 1) Molecular and Computational Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA; 2) Joint Initiative for Metrology in Biology, Stanford, CA 94305, USA; 3) SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA; 4) Department of Genetics, Stanford University, Stanford, CA 94305, USA; 5) Stanford Genome Technology Center, Stanford University, Palo Alto, California, USA; 6) Department of Genetics, Stanford University School of Medicine, Stanford, California, USA; 7) Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Genetic perturbations can show different phenotypic effects across individuals due to epistasis with segregating loci. To achieve a broader understanding of the extent and genetic basis of these background effects, we developed a yeast cross amenable to high-throughput integration and efficient phenotyping of CRISPR interference libraries. In our initial experiment, we examined ~200 segregants, each containing a library of ~10,000 gRNAs targeting primarily essential genes. The relative fitnesses of all segregant-gRNA combinations were measured in a common pool using a double-barcode sequencing strategy, with one barcode denoting a segregant genotype and a second barcode denoting a gRNA. While

analysis is ongoing, we expect these data will enable a broad, systems-level understanding of how genetic differences among individuals cause background effects.

374T Natural genetic modifiers of sensitivity to dopamine-level perturbations in *Drosophila melanogaster* *Ana Marija Jaksic*[1,2], Andrew Clark[2] 1) EPFL Swiss Federal Institute of Technology Lausanne; 2) Cornell University, Ithaca, NY

Dopamine (DA) plays a major role in many animal behaviors, and yet its level is highly variable. It naturally changes over the lifetime in concert with physiological states and environmental stresses. In order to maintain stable expression of many important downstream behaviors, DA level homeostasis needs to be regulated. The way DA homeostasis is achieved on a cellular level has been under extensive investigation, due to its role in Parkinson's disease. However, genetic variation underlying these traits remains relatively unexplored. The extensively characterized genetic diversity existing in the *Drosophila melanogaster* Genetic Reference Panel, as well as the utility of *Drosophila* neurogenetic toolkit, enables us to pursue this question in a systematic way.
Here, we use pharmacological interventions to perturb DA level in diverse genetic backgrounds of the DGRP.This enabled us to simulate exogenously induced dopamine perturbations while avoiding confounding the effects of dopamine level with the systemic phenotypic response to a specific environment. We administered L-DOPA (dopamine precursor) and 3IY (dopamine-precursor agonist) to the DGRP lines and then measured changes in locomotion across genotypes. This enabled us to explore the interaction in 193 genotypes and three dopamine states (nominal, elevated and depleted). Using genome-wide association study we then identified new genetic modifiers of perturbed dopamine phenotypes. These variants point to novel as well as known pathways that affect DA signaling. Namely, we find that the natural genetic variation of enzymes along the cAMP signaling cascade, in the octopamine synthesis, as well as sulfotransferase pathways may play an important role in maintaining locomotion upon perturbation of DA levels. In this study we explore and discuss their functional relevance.

375T Yeast prions regulate host physiology *Janet Chih-chun Lin*[1], Maushmi D. Chitale[1], Jessica Y. Jiang[1], Elissa J. Cosgrove[1], Alexandria C. Van Elgort[2], Asha M. Jain[1], Julia C. Kelso[1], Danial F. Jarosz[2], Andrew G. Clark[1] 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Department of Chemical and Systems Biology, Stanford University, Stanford, CA

Prions, misfolded proteins best known for causing mad cow disease, can help yeasts survive harsh environmental conditions. The prion-directed beneficial traits can also be passed down to their offspring for >100 generations. In addition, more than one-third of wild yeast isolates are found to harbor various kinds of prion proteins. Forming prion proteins seems to be a common strategy for microbes to endure physiological challenges. However, whether and how these yeast prions play a role in host-microbe interaction is entirely unknown. Fruit flies routinely encounter yeasts in the wild and the lab. Yeasts are an essential member of the fly mycobiome. They not only serve as a protein source, but yeasts also provide significant signaling cues regulating fly physiology, reproduction, and behavior. Here we use *Drosophila melanogaster* and *Saccharomyces cerevisiae* as a model to investigate the mechanism by which mycobiome-derived prions impact host health. After testing a battery of prion proteins, we found yeasts with the prion protein [MRPL10+] significantly promote cold tolerance in flies. We also discovered a wide range of cold tolerance levels across global diversity lines (GDLs), a collection of fly lines with different genetic backgrounds. By harnessing the power of large-scale GDL screening, GWAS, functional genomics, and RNAi screening, we identified a list of candidate fly genes functionally crucial in responding to yeast-derived prion protein [MRPL10+]. Following up on several candidate genes, we uncovered the neuronal cause of the cold tolerance phenotype in flies. Our results reveal the novel link between fungal prion proteins and host physiology. This appears to be the first systematic analysis of the genetic interaction of fungal prions in the host. Our approach will facilitate the discovery of beneficial prions in microbes and provide mechanistic insight into the mycobiome field.

376W Modularization, minimization, and diversification of the yeast transcription factor repertoire *Daniel T. Lusk*, Alessandro L.V. Coradini, Cara B. Hull, Oscar M. Aparicio, Ian M. Ehrenreich University of Southern California

Chromosome synthesis can be used to reorganize genomes in ways that help improve understanding of cellular life and evolution. Here, we are engineering a yeast strain that can be used to better explore global transcriptional control and its relationship to phenotypic diversity. We are relocating the roughly ~200 DNA binding, RNA polymerase II-associated transcription factors (TFs) in *Saccharomyces cerevisiae* into a single functional module on a neochromosome. The TF neochromosome is being synthesized by assembling ~260 gene-sized pieces of synthetic DNA into a single molecule. We are also constructing TF-free native chromosomes through cloning and reassembly of TF-free natural DNA segments. The outcome of this work will be a cell in which nearly all transcriptional regulation is controlled by a synthetic module on a distinct chromosome. This module will provide a platform for probing minimal sets of TFs required for viability. A minimal

TF cohort will then enable explorations of the phenotypic diversity achievable through the reintroduction of accessory TFs.

377W Mapping mitonuclear epistasis in *Saccharomyces cerevisiae* *Tuc Nguyen*[1], Meghan Lenhardt[2], Margaret Geertz[2], Austen Tinz-Burdick[2], Francesca Ramirez[2], Anthony Fiumera[2], Heather Fiumera[2] 1) New York University; 2) Binghamton University

The maintenance and heredity of functional mitochondrial DNA (mtDNA) is critical for mitochondrial function and organismal fitness. Because coordination between the mtDNA and nuclear genome is required, selection for mitonuclear interactions that stabilize mtDNAs should be important in shaping mitonuclear coevolution. Understanding and mapping evolutionary important mitonuclear interactions is a major goal in biology. Here, we use the small colony, *petite*, phenotype of *Saccharomyces* yeasts that is produced in the absence of a functional mitochondrial respiratory chain to explore the role of natural genetic variation in mtDNA maintenance and stability. We found that mtDNA stability is a complex trait influenced by mtDNA-dependent mitonuclear interactions, and that, in at least one population, selection has coadapted mitonuclear interactions that increase mtDNA stability. To map mitonuclear interactions, we created a multiparent introgressed recombinant panel of *S. cerevisiae* yeasts originating from 25 natural isolates and three mtDNAs and built an association model to identify SNPs that were dependent or independent of mitotype. Mitonuclear interacting SNPs associating with mtDNA stability included genes involved in mitotic cell growth. We found that in natural isolates, and laboratory-induced conditions, mtDNA stabilities correlated with growth rates, suggesting a fitness tradeoff between rapid cell division and mitochondrial health. SNPs that associated with rates of mtDNA loss in ways that were independent of mitotype included alleles of *MIP1,* the mitochondrial DNA polymerase as well as other genes with known association with mitochondrial activities. This work presents a new tool for mapping mitonuclear interactions and promotes the idea that evolutionary important mitonuclear interactions can influence intracellular signaling pathways.

378T Transcriptome-based gene interaction models reveal epistatic relationships in the barley-powdery mildew pathosystem *Valeria Velasquez-Zapata*[1], Priyanka Surana[2], Antony Chapman[3], Gregory Fuerst[4], Roger Wise[4] 1) Iowa State University, Ames, IA; 2) Informatics Infrastructure Team, Tree of Life Programme, Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK; 3) Phytoform Labs, Rothamsted Research, Harpenden, UK; 4) Corn Insects and Crop Genetics Research, USDA-Agricultural Research Service, Ames, IA

Plant disease resistance often occurs upon direct or indirect recognition of pathogen effectors by host nucleotide-binding leucine-rich-repeat (NLR) receptors. The Triticeae grain crop barley has evolved a diverse series of NLR receptors, including those encoded by *Mildew resistance locus a (Mla*), which are complementary to effectors secreted by the powdery mildew fungus, *Blumeria graminis* f.sp. *hordei* (*Bgh*). *Mla* is essential to a complex gene interaction network that leads to life or death of the host. Epistasis models help to explain those relationships by calculating gene effects on the phenotype and classifying them as additive or the product of gene interaction(s). Here, we used a dynamic transcriptome collected from the interaction between barley and *Bgh* to propose models to infer gene effects and epistatic relationships among *Mla6* and two additional host genes critical to the interaction, *Blufensin1* (*Bln1*) and *Rar3 (Sgt1). Bln1* is a negative regulator of immune signaling and the resistant *bln1* mutant exhibits enhanced basal defense. *Rar3* (*required for Mla6 resistance3*) is required for MLA6-mediated generation of H_2O_2 and the hypersensitive response. The susceptible *rar3* mutant contains an in-frame Lys-Leu deletion in the SGT1-specific domain, which interacts with NLR proteins. The first model contains data from single and double immune mutants allowing for the calculation of epistatic effects between *Mla6* and *Bln1* and demonstrating their genetic interaction. The second model between *Mla6* and *Sgt1,* proposed from single mutant data, revealed that both genes have dominant and equivalent effects to control barley gene expression. In contrast, most of the *Bgh* transcriptome showed dependence on the disease phenotype while some genes would fit the *Mla6*:*Bln1* epistatic- or *Mla6*:*Sgt1* dominant responses, suggesting that its expression is regulated by host-pathogen intercommunication. Genomic location was proposed as a regulation mechanism of the gene effects by associating chromosome hotspots with different genetic effect patterns. Lastly, two gene families were characterized under the models, NLRs for barley and effectors for *Bgh*, which determine the outcome of the host-pathogen interaction. Results from this analysis point to a large perturbation network of the host and pathogen arsenals under different genetic mechanisms that diversify expression patterns and increase robustness of the response.

379T Pigmentation in Drosophila melanogaster and the Genetic Correlation to Fitness Traits. *Patricka Williams-Simon*[1], Hayes Oken[1], Xinwen Zhang[2], Kelsey Sinclair[2], Carter Johnson[2], Conner Traugot[2], Lauren McIntyre[2], Paul Schmidt[1] 1) University of Pennsylvania; 2) University of Florida

Drosophila pigmentation is a complex trait that is thought to be linked to fitness. In *D. melanogaster* there are well

documented altitudinal and latitudinal clines, implicating pigmentation phenotype in spatially varying selection in natural populations. In addition, pigmentation is a complex, polygenic trait with several genes of major effect that are highly pleiotropic. This suggests that variance in pigmentation phenotype may covary with a number of other fitness-associated traits. Quantitative genetic methods can be used to test the hypotheses of genetic correlations between abdominal melanization and other fitness traits, which are surprisingly not well characterized in this system. Here, we used a North Carolina Design II with parents from populations of D. melanogaster derived from **a)** 150 generations of artificial selection on extreme phenotypes, **b)** the latitudinal range in the eastern USA (Maine and Florida), and **c)** spring and fall seasonal inbreds developed from a Pennsylvania orchard population. Parents were selected based on variation in pigmentation phenotype and F1 crosses were measured for pigmentation and fitness related traits. We estimated heritability of individual traits and genetic correlations between pigmentation and fitness associated traits. Future studies will aim to perform direct tests of the adaptive significance of variation in pigmentation phenotype using field based experiments.

380W Exploring genetic variation in the sex determination signal in *Drosophila melanogaster* *Frederick Xu*, Andrew Clark, Daniel Barbash Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY

Sex determination mechanisms vary extensively across taxa, but the extent of variation *within* a sex determination pathway is not fully known. In *Drosophila*, sex is determined by counting X-Signal Elements (XSEs). The five XSEs—*sisterlessA* (*sisA*), *sisterlessB* (*sisB*), *runt*, *unpaired* (*sisC*), and *diminutive* (*myc*)—activate in females *Sex lethal* (*Sxl*), the primary sex determination switch. Activation occurs by sufficient XSE dosage when two X chromosomes are present to overcome maternally deposited repressors in the embryo. Sxl proteins subsequently control the somatic sex determination and dosage compensation pathways. In 1988, Thomas Cline found wide variation among wildtype *D. melanogaster* strains for sensitivity to perturbations in sex determination: some strains skewed heavily female-biased in their sex ratios when *sisA* and *sisB* were duplicated, while others skewed heavily male-biased when heterozygous for *sisA* and *Sxl* mutations. Interestingly, strains with sex ratios more sensitive to XSE duplications were more resilient to XSE mutations—and vice versa, suggesting that wildtype strains contain hidden variation that places them along a continuum of sensitivity to *Sxl*-dependent sex ratio perturbations. This begs the question: what is the source of this variation in Drosophila sex determination, and how does it shape the evolutionary dynamics of sex determination, sex ratios, and reproduction?
The Drosophila Synthetic Population Resource (DSPR) is a collection of ~1700 Recombinant Inbred Lines (RILs) generated from 15 founder lines that can be used for fine-scale mapping to a resolution of ~1-2cM. We mated *D. melanogaster* lines containing *sisA*, *sisB*, and/or *Sxl* mutations or duplicated *sisA* and *sisB* elements to the DSPR founder lines and scored progeny sex ratios to determine which founders exhibit the strongest F1 sex ratio skew, either female- or male-biased. After determining the optimal set of mapping RILs, we will mate the same mutant lines from the founder crosses with each of the selected RILs to find the genomic region(s) that harbor variation in progeny sex ratio. Since the DSPR has extensive SNP data for the founders and RILs, we can determine the founder identity of the region(s) driving sex ratio bias and compare it to the same region(s) in other founders in order to identify candidate genes underlying variation in sex ratio skew. Functional analyses will then elucidate mechanisms behind sex determination bias.

381W Inferring non-additive multi-locus selection in introgressed populations using hidden markov models *Nicolas Ayala*[1, 2], Russell Corbett-Detig[1, 2] 1) Department of Biomolecular Engineering, University of California, Santa Cruz; Santa Cruz, CA 95064, USA; 2) Genomics Institute, University of California, Santa Cruz; Santa Cruz, CA 95064, USA

Admixture combines genetic material from potentially disparate populations and is thought to be a major source of adaptive novelty. As such, multi-locus and non-additive selection on introgressing mutations is potentially common in natural populations. However, existing tools for inferring adaptive introgression only account for additive selection at a single site, overlooking phenomena such as interference among selected loci and dominance. To meet this important need, we present AHMM-GLS, a hidden markov model based tool for inferring and identifying multiple selected sites on a chromosome. This tool numerically calculates local ancestry landscapes for a given MLS model, and then optimizes the model to fit the data. It uses read pileup data in an introgressed population to identify selected sites and estimate a multi-locus selection model. In applying our method to a suite of admixed populations, we find that the estimated strength of selection can be affected by ignoring the contributions of other sites. This method will enable more accurate and detailed analyses of selection in admixed populations than has been possible previously.

382T Precisely calculating relative fitness advantage (s) for diverse mutants that provide drug resistance to better inform treatment models *Daphne Newell*[1,2], Kara Schmidlin[2], Rachel Eder[1,2], Michael Hinczewski[3], Jacob Scott[3,4,5], Kerry Geiler-Samerotte[1,2] 1) School of Life Sciences, Arizona State University, Tempe, AZ; 2) Center for Mechanisms of Evolution,

Biodesign Institute, Arizona State University, Tempe, AZ; 3) Department of Physics, Case Western Reserve University, Cleveland, OH; 4) Translational Hematology Oncology Research, Cleveland Clinic, Cleveland, OH; 5) Case Western Reserve University School of Medicine, Cleveland, OH

Drug resistance in pathogens is a major global health concern. Thus, there is great interest in modeling the behavior of drug resistant mutations; for example, how quickly they will rise to high frequency within a population, and whether they come with fitness tradeoffs that can form the basis of treatment strategies. These models of how resistant mutations behave often depend on precise measurements of the relative fitness advantage (s) for each resistant mutation, as well as the strength of the fitness tradeoff that each mutation suffers in other contexts. Previous studies often determine the maximum growth rate of each mutant individually to calculate s. More recent experiments use direct competitions in batch culture to study which strains perform better in different environments. In many cases, a direct competition experiment is more clinically relevant, as microbes with different resistant mutations often compete within the same patient.

Precisely quantifying s helps us create better, more accurate models of how mutants act in different treatment strategies. For example, *P. falciparum* acquires antimalarial drug resistance through a series of mutations to a single gene. Prior work in yeast expressing this *P. falciparum* gene demonstrated that mutations come with tradeoffs. Computational work has demonstrated the possibility of a treatment strategy which first enriches for a particular resistant mutation, which then makes the population grow poorly once the drug is removed. This treatment strategy still requires knowledge of s and how it changes when multiple resistant mutants are competing with one another across various drug concentrations.

Here, we precisely quantified s in varying drug concentrations for five resistant mutants, each of which provide varying degrees of drug resistance to antimalarial drugs. This was accomplished using DNA barcodes to label each strain, allowing the mutants to be pooled together for direct competition in different concentrations of drug. This will provide data that can make the models more accurate, potentially facilitating more effective drug treatments in the future.

383T Distangaling genotype-by-environment and maternal effects in breed-specific genomic predictions for growth traits *Sara Nilson*[1], Troy Rowan[1,2,3,4], Robert Schnabel[1,5], Jared Decker[1,5] 1) Division of Animal Sciences, University of Missouri, Columbia, MO; 2) Genetics Area Program, University of Missouri, Columbia, MO; 3) Department of Animal Science, University of Tennessee, Knoxville, TN; 4) College of Veterinary Medicine, Large Animal Clinical Science, University of Tennessee, Knoxville, TN; 5) Institute for Data Science and Informatics, University of Missouri, Columbia, MO

Genotype-by-environment interactions influence the productivity of beef cattle as one of the last agricultural species remaining outside for the entirety of their lives. Yet, maternal effects shape the potential of calves' development before they are directly exposed to their environment and sets the baseline for lifetime growth. Currently, national genomic evaluation models do not include interactions despite their potential being acknowledged. Novel models including a direct or maternal genotype-by-environment interaction in addition to a genetic maternal effect, will disentangle the source and amount of variance that is contributed to production traits: birth weight, weaning weight, and yearling weight. Approximately 45,000 Red Angus and 98,000 Simmental cattle distributed across the United States with ~850,000 imputed SNPs were included in developing breed-specific genomic predictions. To test if prediction accuracy was further improved, SNPs were reduced to the top hits from analyses of selection. The top associations from envGWAS (associations between SNP genotypes and environmental variables indicative of local adaptation) were used to estimate genotype-by-environmental effects and GPSM (associations between SNP genotypes and time indicative of directional selection) for direct effects. Variance components were estimated with GCTA v1.93.2 and breeding values were estimated and predicted with the BLUPF90 suite. Utilizing a validation set of the youngest 10% of individuals, two measures of accuracy were calculated, accLR and accPA, to compare the genotype-by-environment inclusive models and the current national evaluation. The amount of variance explained by the effects is expected to shift among sources between the growth traits reflecting changes in environmental interactions among life stages. These genotype-by-environment models will influence selection decisions in the beef industry allowing for performance to be accurately predicted in a rapidly changing climate.

384W Predicting gene expression responses in A.thaliana using natural cis regulatory variation. *Margarita Takou*[1], Emily Bellis[2], Jesse Lasky[1] 1) Pennsylvania State University; 2) Arkansas State University

The evolution of gene expression is critical for adaptation due to its role in shaping organismal function. However, predicting the response to different environmental conditions is complicated as variation in *cis* regulatory elements, including transcription factors, impact the potential of adaptation to novel environments. Incorporating information

about the variance in the genetic code across multiple genotypes from different populations may improve the prediction of response to environmental fluctuations and give insights into the evolution of expression and plasticity. We use machine learning approaches to predict the impact of *cis* regulatory variation on stress responsive expression across *Arabidopsis thaliana* populations. The predicted variation is then tested for conservation across populations. Thus, we explore regulatory elements that are universally important for response to the stress, as well as novel within specific lineages. The impact of selection on the predicted sites is studied, identifying the relative importance of selection in evolving the regulation of expression under stress.

385W Temporal dissection of meristems differentiation and reprogramming by single-tissue and single-cell transcriptome profiling *Zohar Meir*, Ziva Amsellem, Yuval Eshed, Amos Tanay Weizmann Institute of Science

Multipotent tissues in plants comprised of cells with self-renewal and flexible differentiation capacity, which collectively form different types of meristems. New single cell genomics techniques can provide an unbiased approach for mapping differentiation programs within meristem tissues over time, and for monitoring the impact of genetic perturbations at native cellular resolution. We have recently described the application of sensitive MARS-seq protocol for profiling transcription in developing meristems, which resulted in rich temporal model for tissue-level transcriptional trends. This model can be used for "aligning" mutant data over the wildtype reference, e.g., dissecting florigen dependent and independent signaling in the shoot apical meristem. In another experimental system for plants cellular plasticity and commitment, we will describe a new approach for single cell RNA-seq in callus tissue, demonstrating how several parallel transcriptional programs emerge over time within epigenetically reprogrammed calli. Thus far, we characterized the gain and loss of pluripotency in callus tissue driven by injury of mature tomato hypocotyls by single-cell RNA analysis of >40K callus cells. This revealed a surprisingly rich heterogeneity of cellular programs within calli, despite the lack of distinct morphological features. A precise map of cell states in the callus and its combination with massively parallel single-callus bulk RNA-seq profiling pointed at multiple gene programs that are activated during reprograming, where only few of them are associated with a hierarchical regulatory cascade related to shoot differentiation. For example, we identified a large sub-population of cells expressing the photosynthetic apparatus that show no clear similarity to any known plant cell-type. The fraction of these photosynthetic cells increased over time and was coupled with gradual of loss proliferative capacity and meristematic features in the callus. Finally, a functional screen pointed at the transcription factor *WOX4* as an important gene for callus reprogramming and differentiation and a potential regulator of the callus meristematic niche. Interestingly, *wox4* mutated calli show altered kinetics of exit from multipotent state and failed to form the reprogrammed photosynthetic sub-population. This high-resolution analysis of the injury-driven callus provides a new quantitative framework for defining mechanisms of de-differentiation, fate determination and commitment in plants.

386T Polygenic adaptation under recurrent changes in environment *Jiarun Chen*, Guy Sella Columbia University

Polygenic adaptation is expected to be ubiquitous, yet remains poorly studied. Previous work modeled this process for a quantitative, highly polygenic trait that experiences a sudden shift in the fitness optimum, and found that the long-term adaptive response arises from a slightly greater number of fixations of alleles whose effects are aligned with the shift than that of fixations of alleles with opposing effects. These fixations occur over a time scale of $2N_e$, where N_e is the effective population size. However, the fitness optimum plausibly changes repeatedly over this time scale, suggesting that recurrent shifts in the fitness optimum should probably be more biologically realistic. Here we model a quantitative, highly polygenic trait that experiences recurrent shifts in the fitness optimum, where the waiting time between shifts is exponentially distributed, the direction of each shift is random, i.e., with probability 0.5 increasing and decreasing the fitness optimum, respectively, and the magnitude of each shift follows some distribution. In particular, we ask how the fixation probability and allele frequency distribution depends on the allelic effects on the trait, and on the frequency and magnitude of changes to the environment, using analytical methods and simulations. We anticipate our results to have important implications in the genetic basis of polygenic adaptation in humans and other species, and in enduring questions about the rate of molecular evolution and levels of genetic diversity.

387T Distinguishing multiple-merger from Kingman coalescence using the two-site frequency spectrum. *Eliot Fenton*[1], Daniel Rice[2], John Novembre[3], Michael Desai[1] 1) Harvard University, Cambridge, MA; 2) National Center for Biotechnology Information, Bethesda, MD; 3) University of Chicago, Chicago, IL

Many demographic inference methods in population biology use the site frequency spectrum (SFS) to fit past population sizes to present-day genetic data. These methods typically assume the population ancestry is well represented by a Kingman coalescent in which only two lineages can coalesce at once and all lineages are equally likely to coalesce at any time. However, real populations often violate the assumptions of the Kingman coalescent. For example, natural selection and

highly skewed offspring number distributions can each lead to "multiple-merger" coalescence events in which three or more lineages simultaneously coalesce. In both cases, features of the SFS can often be reproduced by a Kingman coalescent with changing population size, making it difficult to infer the evolutionary and demographic forces shaping a population from the SFS alone. Here, I present results on distinguishing changing-population-size Kingman coalescents from multiple-merger coalescents using the two-site frequency spectrum (2SFS) constructed from pairs of linked sites. Using a combination of forward- and backward-time simulations, we show that our method distinguishes fluctuating population sizes from natural selection and skewed offspring number distributions. I additionally present a pipeline for analyzing real-world genomic data and results from applying this technique to *Drosophila melanogaster* strains.

388W The impact of measurement error in mediation analysis *Madeleine Gastonguay*, Gregory Keele, Gary Churchill The Jackson Laboratory, Bar Harbor, ME

Mediation analysis is a class of statistical techniques used to investigate the relationship among interacting variables in a causal system. In the context of genetics, it can be used to identify causal intermediaries of genetic effects on phenotypes and thus illuminate genetic regulatory systems. Mediation relies on variation due to causal effects that propagates through the system to produce characteristic patterns of covariation in the observed data. In addition to this causal variation, observed data also include measurement error that does not propagate through the causal relationships. Typically, mediation analyses do not attempt to distinguish between these sources of variation. In this work, we evaluate the reliability of mediation analysis for determining the structure of a causal system in the presence of measurement error and identify ways to diagnose cases where the resulting inference may be misleading. We focus on the biological context where a quantitative trait locus (QTL; X) regulates a target phenotype (Y), potentially completely or partially through a candidate mediator (M). For example, M could represent transcript or protein expression levels of a gene that co-localizes with the QTL. We define a measurement error model to relate the true unobserved variables to their measured quantities and simulated from various causal models. We confirm that measurement error in the mediator when the causal relationship is complete mediation results in mediation analysis incorrectly inferring partial mediation. Notably, this issue occurs more frequently as sample size increases. We observed similar results from simulations of a co-local relationship in which the QTL affects M and Y independently and there is error in X, indicating that partial mediation may be inferred even when there is no causal relationship between M and Y. Furthermore, in cases when partial mediation is not inferred, the relative errors of X, M, and Y will determine which model is preferred. Based on these observations, we derived guidelines for cases where mediation inferences will be consistent or inconsistent with the unobserved causal relationship. Using examples in data from genetically diverse mouse populations, we demonstrate how these guidelines can be used to assess the reliability of mediation inferences and highlight common scenarios in which they will be incorrect.

389W Unreasonably fast estimates of similarity among loci and individuals Anthony Greenberg Bayesic Research, Ithaca, NY

Increasing amounts of genotype information put pressure on computational resources. Researchers and other consumers of genotype information who do not have access to powerful computer hardware can thus be at a disadvantage. Devising analysis algorithms that efficiently process large data sets is an important component in the drive to democratize access to information. It can also save time and energy consumption of compute clusters. Estimating similarities among loci (linkage disequilibrium, LD) and individuals (relationship matrices) are ubiquitous steps in numerous analysis pipelines. Time to compute LD among loci using exact algorithms grows linearly with the number of individuals in a data set and quadratically with the number of loci. While optimizing individual operations can yield significant improvements, we need approximate procedures to improve on these undesirable scaling properties. I describe an approach that uses similarity-preserving hashes to summarize genotype data. This allows for sparse LD matrix computation that is almost insensitive to the number of individuals and slows down less than quadratically as the number of loci in the data set increases. Conversely, time to estimate sparse genetic similarity matrices is close to insensitive to the number of loci and grows slowly with the number of individuals. In addition, these algorithms require much less memory and allow for explicit precision-time trade-offs. Software implementing these approaches is freely available on GitHub (https://github.com/tonymugen/vash).

390T A Bayesian filtering method for estimating fitness effects of nascent beneficial mutations from barcode-lineage tracking data Huan-Yu Kuo University of California, San Diego, CA

The distribution of fitness effect of new beneficial mutations (DFE) is a fundamental quantity necessary to understand adaptive evolution. Despite its importance, direct measurements of the DFE remain challenging because they require observing many independent and isolated adaptive mutations. Recently, the DNA barcode lineage tracking (BLT) approach

has been developed for this purpose, whereby ~10^5 neutral DNA barcodes are introduced into the genome and used as markers for tracking the frequencies of clonal lineages during evolution. When a new adaptive mutation arises, it is permanently linked to one barcode (as long as the organism reproduces asexually) and leads to an observable increase in this barcode's frequency. The selection coefficients of many beneficial mutations can then be inferred from the frequency trajectories of rising barcode lineages with a single experiment.

Although the BLT method for estimating the DFE shows great potential, identifying lineages that carry adaptive mutations and inferring the selection coefficients of these mutations remains difficult. In particular, current methods for analyzing BLT data do not fully account for barcode lineage extinction, a situation that can occur when selection or drift is strong. In addition, these methods classify lineages based on their entire frequency trajectories, which becomes computationally prohibitive with denser temporal sampling. Here, we develop a novel Bayesian method for inferring the DFE from the BLT data that overcomes these challenges. The central concept of our method is the probability distribution on the lineage's selection coefficient, which is updated one data point at a time. At each time step, we also estimate the global parameters, population's mean fitness as well as the experimental noise, based on the current knowledge of the fitness of individual lineages. We validate our method using simulations and show that it successfully recovers mean-fitness trajectories and the selection coefficients of individual lineages even under relatively high genetic drift, experimental noise, and strong selection. This work paves the way for rigorously and efficiently estimating distributions of fitness effects of beneficial mutations from BLT experiments.

391T Ped_slim: a family pedigree toolkit to investigate distant relative misidentification in long-range familial searching *Joaquin Magana*[1], Miguel Guardado[1,2], Shalom Gutierrez[1], Sthen Campana[1], Kaela Syas[1], Emily Samperio[1], Cynthia Perez[1], Berenice Chavez[1], Selena Hernandez[1], Rori Rohlfs[1] 1) San Francisco State University; 2) University of California San Francisco

Investigative Genetic Genealogy (IGG) is a forensic technique used to identify a criminal suspect through their long-distance relatives, such as a third cousin. IGG uses shared inheritance of autosomal DNA segments, Identical by Descent (IBD), to help identify genetic relationships. Studies of European-ancestry data have shown the power of this technique, with an estimated sixty percent of individuals in the United States identifiable through a third cousin or closer match. However, less is known about IGG accuracy across other ancestral population groups. One reason for this is the lack of software available for identifying complex family relationships and efficiently simulating genomes onto familial pedigrees. We developed a family pedigree software, ped_slim, that can investigate the misidentification rate of IGG for different populations. We created a python command-line-based tool to interrogate family pedigrees, providing a simple interface to perform complex pedigree genetic simulations using SLiM, a forward evolutionary software. Ped_slim comes with three main features: (i) simulate family pedigrees structures, (ii) simulate genomes on pedigree structures, and (iii) identify the familial relationship between all pairs of individuals. We represent family pedigrees as directed graph data structures for these three features, where nodes represent individuals and edges define parent-child relationships. To perform the genetic simulation, ped_slim utilizes a directed graph to convert the family pedigree into a file SLiM can read, in order to simulate the family's genome. We validate our software by estimating the kinship of pairs of genetic relatives in our simulations, initializing the family founders with individuals from the 1000 genomes consortium. We show that the kinship estimated from our simulations fits the expected known genetic relation. Additionally, we showcase the ped_slim feature of identifying familial relationships by showing that as estimated genetic relationships grow more distant, less expected kinships are seen between the pair of individuals. Finally, we utilize our simulated family pedigree feature by comparing standard familial statistics between simulated pedigrees and commonly used nuclear family pedigrees to show the benefit of simulating non-nuclear families. Overall, ped_slim will not only provide an open-source solution to investigate the accuracy of IGG, but also to genome simulations inside medical, forensic, and evolutionary genetic analysis.

392W A kinship-based approach to learn maximally heritable traits from high-dimensional quantitative assays Callan O'Connor[1,2], Seamus Mawe[1], Greg Keele[1], Daniel Gatti[1], Ron Korstanje[1], Gary Churchill[1], Laura Reinholdt[1], *J. Matthew Mahoney*[1,3] 1) The Jackson Laboratory, Bar Harbor, ME; 2) Graduate School of Biomedical Sciences, Tufts University, Boston, MA; 3) Department of Neurological Sciences, University of Vermont Larner College of Medicine, Burlington, VT

Identifying heritable traits for analysis is one of the essential tasks of quantitative genetics, and it is becoming more challenging as biological assays now routinely produce hundreds or thousands of quantitative variables. For example, computer vision systems quantify thousands of morphological features from images. Often, cleaner, more biologically relevant variables can be summarized as composite traits from the measured variables. Recently, Mitteroecker, *et al.* (*Genetics*, 2016) showed that finding maximally heritable composite traits as a linear combination of measured variables is equivalent to canonical correlation analysis (CCA) between the genotype and phenotype data matrices. This theoretical

result provides a systematic approach to synthesize traits that are maximally aligned with genetic variation, but it has several practical limitations. It is typical to have many more genetic variants and measured variables than individuals ("large-p, small-n" data), leading to costly computations on extremely large matrices in CCA. Furthermore, CCA becomes unstable in high-dimensional settings. To overcome these limitations, we have applied two machine learning strategies to CCA to enable robust maximum heritability analysis in high dimensional data. First, we used *kernel CCA* (kCCA) instead of classical CCA to reduce the dimensionality of all data matrices. Second, we employed *bootstrap aggregation* (bagging) to minimize the variance of trait estimation. The inputs to kCCA are a genetic similarity matrix (i.e., a kinship matrix), which can include variant, pedigree, and non-additive effects, and a trait similarity matrix, which can be a covariance matrix or any non-linear positive definite kernel defined on the traits. To test our approach, we applied *bagging kCCA* to high-content screening imaging data of fibroblasts cultured from 200 Diversity Outbred (DO) mice. The raw image features included morphological quantifications of individual cells (e.g., nuclear and cell roundness) and summary measures of whole culture wells (e.g., cell densities) quantified using the Harmony 4.9 software suite for the Operetta high-content screening system. The maximally heritable trait corresponded to distinct differences in cell morphology (h^2=0.67), suggesting that the genetic variation of DO fibroblasts strongly influenced how the cells organized in culture. We identified a significant quantitative trait locus (LOD = 8.7) on chromosome 12 for the composite trait that included the gene *Pxdn*, which is highly expressed in fibroblasts, has a known expression QTL in DO mice with consistent genetic effects to the trait QTL, and is involved in extracellular matrix organization. These results support bagging kCCA as a robust strategy for identifying maximally heritable traits from high-dimensional quantitative assays that can then be used for downstream genetic analyses.

393W Inferring sparse latent structure from genotype-phenotype maps Samantha Petti, *Gautam Reddy*, Michael Desai Harvard University, Cambridge, MA

Inferring the structure of a genotype-phenotype map is a long-standing problem in quantitative genetics. Covariation across phenotypes and covariation of genetic effects on the phenotypes contain useful statistical information about phenotype-determining pathways that share the same genes. Structure discovery therefore benefits from measuring the effects of many genetic perturbations on a large number of phenotypes. We develop a conceptual framework and an accompanying analysis pipeline for joint QTL mapping and structure discovery. First, we use a penalized regression framework to jointly map causal loci and their effects across phenotypes. Second, we develop statistical methods to test for the presence of sparse, lower-dimensional "core phenotypes" that explain covariation patterns. Finally, we show that a penalized matrix decomposition framework can be used to identify this sparse structure. We apply our methods on a variety of fitness-based datasets, including genotype-fitness measurements of 100,000 budding yeast offspring from an F1 cross, fitness effects in diverse environments of adaptive mutations in yeast, genotoxic fitness screens from human cell lines and a large-scale yeast chemogenomics assay. We find that individual genes often affect only a sparse subset of core phenotypes, which however can influence fitness in diverse contexts. The extent of pleiotropy varies across genes and depends in general on the nature of the genetic perturbation. Covarying patterns of pleiotropy allow for clustering genes into putative pathways, which we compare to existing annotations.

394V Natural Variation in Ubiquitin System Genes Creates Complex, Pathway-Specific Effects on Proteasomal Protein Degradation *Mahlon Collins*[1], Gemechu Mekonnen[1,2], Randi Avery[1], Frank Albert[1] 1) University of Minnesota, Minneapolis, MN; 2) Johns Hopkins University, Baltimore, MD

DNA sequence differences that influence gene expression are a key mechanistic link between individual genetic differences and variation in cellular and organismal traits. Many genetic effects on gene expression specifically affect protein levels and understanding the molecular basis of these effects remains an outstanding challenge. Protein-specific effects on gene expression may arise through variants that alter the activity of the cell's primary pathway for targeted protein degradation, the ubiquitin-proteasome system (UPS).

To explore this possibility, we developed a statistically powerful method for mapping quantitative trait loci (QTLs) for UPS activity using the yeast *Saccharomyces cerevisiae*. We applied this approach to 22 UPS substrates that engage multiple UPS pathways and diverse molecular mechanisms of substrate recognition and processing, including the full set of degradation signals for the UPS N-end Rule. We identified 167 UPS activity QTLs, most of which were specific to individual UPS pathways or substrates, demonstrating a highly complex genetic basis of variation in UPS activity.

Resolving four QTLs to their causal nucleotides using CRISPR-Cas9 genome engineering revealed the molecular basis of genetic effects on UPS activity. Specifically, pathway-specific influences on UPS activity resulted from regulatory and missense variants in ubiquitin system genes whose products process (*NTA1*), recognize (*UBR1*, *DOA10*), and ubiquitinate

(*UBC6*) substrate proteins. Evolutionary and population genetic analysis showed that causal variants that decrease UPS activity tend to be derived and at low (< 5%) population frequency, suggesting that they reduce organismal fitness.

To understand how causal variants for UPS activity influence gene expression, we tested the effect of a derived, *cis*-acting causal variant in the *UBR1* promoter on genome-wide protein and RNA levels. The causal *UBR1* variant altered the abundance of 36 proteins without affecting levels of the corresponding mRNA transcripts, implicating genetic influences on the UPS as a prominent source of protein-specific variant effects on gene expression.

Our results define the complex genetic architecture of UPS activity, demonstrate how variation in ubiquitin system genes influences UPS protein degradation, and establish a framework for understanding how genetic effects on the UPS contribute to variation in cellular and organismal traits.

395V Genetic markers associated with medullary and cortical bone in Rhode Island Red laying hens Mohammed Sallam[1], Heather McCormack[2], Bob Fleming[2], Peter Wilson[2], Björn Andersson[3], Matthias Schmutz [3], Cristina Benavides[4], Nazaret Dominguez-Gasca[4], Estefania Sanchez-Rodriguez[4], Alejandro Rodriguez-Navarro[4], Ian Dunn[2], *Dirk Jan de Koning*[1], Martin Johnsson[1] 1) Swedish University of Agricultural Sciences, 75651 Uppsala, Sweden; 2) Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, Scotland, UK; 3) Lohmann Breeding, 7454 Cuxhaven, Germany; 4) Departamento de Mineralogía Y Petrologia, Universidad de Granada,18002 Granada, Spain

Most laying hens display a high tendency to suffer from bone damage (deviations or fractures) either on legs or keels, which is one of the major welfare challenges in the egg production industry. The problem seems to have increased over time from ~30% of commercial layers with at least one bone fracture (Gregory & Wilkins, 1989) up to ~70% of keel bone fractures (Thøfner et al., 2021). To better understand bone health in these breeds, (Dunn et al., 2021) performed detailed bone phenotyping (~55 traits), including measurements for medulla and cortex separately in addition to the whole bone strengths, egg production, and body weights.
Here, we present preliminary results from a genome-wide association study aiming to find genetic marker associations with 47 bone measurements in a cohort of 924 Rhode Island Red laying hens. Hens were killed at 68 weeks of age, weighted, and bone samples were collected including humerus, tibia, and keel bone for further measurements as described in detail in (Dunn et al., 2021). The measurements included: tibia cortical lipid (determined by infrared spectroscopy at main peak 1710 cm^{-1}), tibia cortical mineral (determined by thermogravimetric analysis, represents salts of Ca and PO4 and some CO3 substitution of the PO4), and medullary bone score (on a scale of 0-3 where 0 represents no medullary bone and 3 a diaphyseal medullary cavity filled with bone). For genome-wide association studies, the hens were genotyped on 57,636 single nucleotide polymorphisms using the Illumina Infinium assay. The genotyping was performed by the SNP&SEQ Technology Platform, Uppsala University, Sweden. For GWAS, we used the linear mixed model implemented in GEMMA version 0.98.5.
Three traits showed associations with p-value < 10^{-5}: tibia cortical lipid (on chromosomes 2 and 3), medullary bone score (on chromosome 9), and tibia cortical mineral% (on chromosome Z). This study demonstrates the importance of developing bone measurements that enable large-scale genetic mapping studies and genomic selection of bone traits.

References
Dunn, I. C., De Koning, D.-J., McCormack, H. A., Fleming, R. H., Wilson, *et al*. (2021) *53*(1): 11. https://doi.org/10.1186/s12711-021-00603-8
Gregory, N. G., & Wilkins, L. J. (1989) *British Poultry Science 30*(3):555–562. https://doi.org/10.1080/00071668908417179
Thøfner, I. C. N., Dahl, J., & Christensen, J. P. (2021) *PLOS ONE*, *16*(8):e0256105. https://doi.org/10.1371/journal.pone.0256105

396V Genotype-by-diet interactions regulate gene expression across multiple tissues in the "Three Bears" mouse models of type II diabetes *Isabela Gerdes Gyuricza*, Candice Baker, Jeffrey Harder, Gary Churchill, The Jackson Laboratory Cube Consortium The Jackson Laboratory, Bar Harbor, ME

Type II Diabetes (T2D) is a systemic inflammatory disease characterized by progressive insulin resistance. The development of T2D is associated with the consumption of food with high contents of fat and sugar, *i.e.*, a western diet. Hundreds of genetic loci have been associated with T2D susceptibility, revealing that the underlying mechanisms of the disease are complex. Distinct isogenic mouse models provide an opportunity to study the development of T2D in different genetic backgrounds with multiple environmental exposures, including diets. In this study, we generated bulk RNA sequencing data from eight tissues (adipose, islet, liver, kidney, heart, skeletal muscle, hypothalamus, and dorsal

vagal complex) from mice from three isogenic strains, C57BL/6J (B6), CAST/EiJ (CAST), and NZO/HlLtJ (NZO). The study included female and male mice raised on either low-fat/low-sugar or high-fat/high-sugar diets. These strains, the "Three Bears", represent a range of T2D susceptibilities. The CAST strain is highly resistant to diet effects and T2D; the NZO strain presents spontaneous obesity and high incidence of T2D, especially in males, which is exacerbated by the western diet; and the B6 strain presents moderate resistance, with T2D and obesity emerging only after long-term exposure to the western diet. By quantifying differences in gene expression in response to diet across these strains, we aim to illuminate molecular mechanisms related to variation in metabolic responses to the western diet and their relationship to T2D. We found that islet, adipose, and hypothalamus present the strongest strain-specific gene expression responses to diet. In the islets, we see activation of immune response and depletion of beta cells, especially in male NZO mice. In contrast, CAST mice show adaptive responses to western diet with increased cellular respiration and fatty acid oxidation across multiple tissues. Differences in metabolic efficiency across these strains are associated with differential levels of circulating erythropoietin and the activation of mitochondrial uncoupling proteins (UCP2 and UCP3) in multiple tissues. These findings suggest that higher rates of energy expenditure and fatty acid oxidation contribute to the robust T2D resistance of the CAST strain while on a western diet.

397V Quantitative and qualitative mapping of loci involved in tocopherol composition and oleic acid content control in Russian sunflower (*Helianthus annuus* L.) lines *Rim Gubaev*[1], Stepan Boldyrev[1], Elena Martynova[1], Alina Chernova[1], Tatyana Kovalenko[2], Tatyana Peretyagina[2], Svetlana Goryunova[1, 4], Denis Goryunov[1, 5], Zhanna Mukhina[3], Cecile Ben[1], Laurent Gentzbittel[1], Philipp Khaitovich[1], Yakov Demurin[2] 1) Skolkovo Institute of Science and Technology, Moscow, Russia; 2) Pustovoit All-Russia Research Institute of Oil Crops, Krasnodar , Russia; 3) All-Russia Rice Research Institute, Krasnodar, Russia; 4) Institute of General Genetics, Moscow, Russia; 5) Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

Tocopherols and oleic acid protect sunflower oil against thermooxidation and significantly affect its quality. The tocopherol complex in plants includes four different forms, namely, α-, β-, γ-, and δ-tocopherols. Importantly, vitamin E activity decreases from α- to δ-tocopherol, while *in vitro* antioxidant activity in contrast increases in the row. Oleic acid is a monounsaturated fatty acid that in combination with γ-tocopherol, and δ-tocopherols significantly increases oil thermostability. Therefore, one of the pressing tasks in sunflower breeding is the creation of plants that would allow producing oil with balanced tocopherol composition and oleic acid content. To facilitate the identification of new loci linked to these traits, we performed association mapping of quantitative trait loci (QTL) based on the high-throughput sequencing data for sunflower plants. For association mapping, two F2 populations of 144 plants each were obtained from two independent crosses of parents contrast in tocopherol composition and oleic acid content from a collection of All-Russia Research Institute of Oil Crops (VNIIMK). Tocopherol composition was measured using thin-layer chromatography while the oleic acid content was measured by means of gas chromatography followed by mass spectrometry. We applied genotyping-by-sequencing with subsequent SNP calling in Tassel-GBS to construct two genetic maps using R/qtl package. For quantitative mapping, the proportions of each of the four tocopherols and oleic acid were used. For qualitative mapping of tocopherol composition, plants' phenotypes were classified into four tocopherol classes based on the tocopherol composition. For qualitative mapping of oleic acid phenotypes were classified into high and low oleic classes. Different approaches were applied, including interval mapping and composite interval mapping. Loci associated with α-, γ-, and δ-tocopherol were located on chromosome 8 while loci associated with β-tocopherol content were found on chromosomes 8 and 1. For oleic acid, a locus located on chromosome 14 was shown to be significantly associated. SNPs associated with the studied traits were verified on the independent plant sample. The QTLs and corresponding SNPs identified will facilitate the marker-assisted selection of sunflower as well as bring new knowledge on the genetic control of tocopherol composition and oleic acid content of the sunflower oil.
The reported study was funded by RFBR projects number 20-316-90051.

398V Expression quantitative trait loci in human milk reveal effects of milk composition on infant and maternal health *Kelsey Johnson*[1], Timothy Heisel[1], Dan Knights[1], Katherine Jacobs[1], Michael Rudolph[2], David Fields[2], Cheryl Gale[1], Frank Albert[1], Ellen Demerath[1], Ran Blekhman[1] 1) University of Minnesota, Minneapolis, MN; 2) University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma

Breast milk provides complete nutrition and confers immune protection to the neonate, and plays a critical role in infant gut microbiome development. The dependence of human--and more broadly, mammal-- reproduction on milk across evolutionary history suggests strong selective pressure on milk production and composition. The role of maternal genetics in the production of human milk and in the expression of most milk components is almost entirely unknown. Understanding the genetic and genomic basis of variation in lactation is critical for identifying the mechanisms linking milk composition and production to infant and maternal health.

Here, we report the first expression quantitative trait locus (eQTL) analysis of human milk, using one-month postpartum milk samples from 171 exclusively breastfeeding women. RNA in milk is primarily derived from the milk-producing mammary gland cells. We identified cis-acting eQTLs for 2,686 genes, of which eQTLs at 668 genes were specific to milk (i.e. not observed in other tissues). Milk-specific eQTLs affected genes encoding major milk proteins (e.g. kappa-casein, *CSN3*) and ubiquitously expressed genes (e.g. the circadian gene *CLOCK*). Milk eQTLs present in other tissues were most often shared with other exocrine gland tissues, like the salivary gland. Colocalization analyses highlighted 7 milk eQTLs that may underlie risk loci for breast cancer. These included a milk eQTL affecting the transcription factor gene *LMX1B* at a breast cancer GWAS locus without a previous candidate gene.

To probe connections between the lactating mammary gland and infant health, we tested for associations between maternal genotype at milk eQTLs and the gut microbiome of exclusively breastfed infants. The strongest association was between the lead SNP of the *LCT* (lactase) eQTL in milk and *Collinsella*, a beneficial microbe in the infant gut. This eQTL is on the haplotype that confers lactase persistence in European adults. The lactase expression-increasing allele was correlated with decreased infant gut *Collinsella* at both 1 and 6 months postpartum. We propose that this relationship may be due to genetically-determined differences in milk composition.

Our results uncover new biology linking the lactating breast with maternal and infant health. An improved understanding of milk genetics will reveal new insights into the evolution of lactation and mammary biology.

399V Genome-wide association study of female resistance to male-induced harm in the *Drosophila* Genetics Reference Panel *Sarah Kettelkamp*, Kimberly Hughes, Joseph Travis Florida State University, Tallahassee, FL

Interlocus sexual conflict (IRSC) arises when the evolutionary interests of different sexes are in opposition. Most studies of IRSC rely on phenotypic data, alone, leaving gaps in our knowledge of how IRSC functions and evolves. Furthermore, much of the genetic evidence on IRSC that does exist focuses on male rather than female traits. To close some of these knowledge gaps, we used a model organism that exhibits IRSC, *Drosophila melanogaster*, to identify potential genes involved in female resistance to male-induced harm. Male *D. melanogaster* have evolved proteins in the seminal fluid that increase their own fitness but decrease the lifespan and fecundity of their female partners. The population we used was the *Drosophila* Genetics Reference Panel (DGRP), a population of over 200 fully sequenced, inbred lines. First, we documented variation among 61 of these inbred lines in how males affected female lifespan and fecundity. Second, in 139 lines, we measured levels of female resistance to fecundity-based harm by crossing them to high-harm and low-harm male lines. Finally, we used the data from the female resistance assays to perform a genome-wide association study (GWAS) using the DGRP's GWAS pipeline. We identifed over 98,000 significant variants associated with female resistance to male-induced harm, of which 17 variants in ten genes were highly significant and annotated. The 17 top variants included 14 single-nucleotide polymorphisms and three indels. The protein coding genes *Shaker*, which encodes the structural alpha subunit of a voltage-gated potassium channel, and *CG43078*, whose function is unknown but is expressed in the adult head, adult heart, and embryonic/larval muscle system, were each associated with three of the top variants. Current work is using the GAL4/UAS system and RNA interference to assess the effects of ten of the top variants on female resistance.

400V Imputation of 3D genome structure by genetic-epigenetic interaction modeling in mice *Lauren Kuffler*, Daniel Skelly, Anne Czechanski, Steven Munger, Christopher Baker, Laura Reinholdt, Gregory Carter Jackson Laboratory, Bar Harbor, ME

Gene expression is known to be affected by interactions between local genetic variation and DNA accessibility, with the latter organized into three-dimensional chromatin structures. Analyses of these interactions has previously been limited, obscuring their regulatory context, and the extent to which they occur throughout the genome. Here we undertake a genome-scale analysis of these interactions in a genetically diverse population to systematically identify global genetic-epigenetic interaction, and reveal constraints imposed by chromatin structure. We establish the extent and structure of genotype-by-epigenotype interaction using embryonic stem cells derived from Diversity Outbred mice. This mouse population segregates millions of variants from eight inbred founders, enabling precision genetic mapping with extensive genotypic and phenotypic diversity. With 176 samples profiled for genotype, gene expression, and open chromatin, we used regression modeling to infer genetic-epigenetic interactions on a genome-wide scale. Our results demonstrate that statistical interactions between genetic variants and chromatin openness are common throughout the genome. We found that these interactions occur within the local area of the affected gene, and that this locality corresponds to topologically associated domains (TADs). The likelihood of interaction was most strongly defined by the 3D domain structure

rather than linear DNA sequence. We show that stable 3D genome structure is an effective tool to guide searches for regulatory elements and, conversely, that regulatory elements in genetically diverse populations provide a means to infer 3D genome structure. In stem cells, open chromatin participating in the most significant regression models demonstrated an enrichment for developmental genes and the TAD-forming CTCF binding complex, providing an opportunity for statistical inference of TAD boundaries operating during early development. These findings provide evidence that genetic and epigenetic factors operate within the context of three-dimensional chromatin structure.

401V Interaction of genetic variation and diet on stress resistance in *Caenorhabditis tropicalis* isolates *Tzitziki Lemus Vergara*[1], Leonid Kruglyak[2] 1) University of California Los Angeles; 2) Howard Hughes Medical Institute

The gut microbiome influences many of its host traits. In humans, disruption of the microbiota balance has been associated with various diseases including obesity, metabolic syndrome, and autoimmune disorders. However, it is challenging to study the mechanisms by which bacteria influence their human hosts due to the complexity of bacterial communities and the genetic diversity of humans. The nematode *Caenorhabditis elegans* has recently been used as a model organism to study the influence of the microbiome and diet on several phenotypes. Studies have shown that the worm microbiome/diet affects important traits such as development, life span, metabolism, and resistance to chemotherapy drugs, but the underlying mechanisms are not well understood.
I recently discovered that resistance to cold stress in a related nematode, *Caenorhabditis tropicalis*, is affected by the worm diet. Interestingly, different *C. tropicalis* isolates are differently affected depending on their diet. Isolate JU1639 was highly susceptible to cold stress when grown on *E. coli* HT115 but survived when grown on OP50. Genetic analysis suggests that cold stress resistance is a dominant trait, and initial mapping revealed a potential QTL on chromosome III. Currently, I am working on refining and validating the QTL associated with cold stress resistance, and in dissecting the bacterial elements and pathways involved in the cold stress survival difference. The genetic variants uncovered by this study will further our understanding of the mechanisms by which diet and microbiome modulate an organism's phenotype, and how this modulation depends on the host genetic variation.

402V Epistasis within and across chromosomes exposes expression of marginal effects of QTL in the Virginia body weight chicken lines *Tilman Rönneburg*[1], Yanjun Zan[1], Christa F Honaker[2], Paul B Siegel[2], Örjan Carlborg[1] 1) Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; 2) Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg VA, USA

Genetic analyses of quantitative inherited traits are complex, and studies have often been unable to confidently explain more than a fraction of the heritability of the trait in question. The missing heritability has been attributed to epistasis, or more often to a large number of small, near-zero-effect-size variants. Assuming this, most of the variance will be due to causative variants with minor effects that are difficult to either detect or verify due to power-requirements. In addition, most of the QTL that are large enough to be detected are not single large effect variants, but multiple, smaller QTL in close enough proximity to create a haplotype with an effect large enough to reach the detection threshold. Fine mapping these QTL can be difficult, either due to the small individual effect sizes, haplotype effects, or dependency on the genetic background, thus increasing the requirements for resolution and power. In practice, this leads to a dearth of information on how these broad associations to a phenotype map to the underlying loci. Here, we use low-coverage, whole-genome sequencing data of a large (n>3300), 18-generation advanced intercross line formed from generation 41 of the Virginia body weight lines. The Virginia lines originated from an outbred common stock of White Plymouth Rock chickens and are bi-directionally selected for 8-week body weight. This model system was used to dissect *Growth7*, a major QTL-region for 8-week body weight and growth-related traits on Chromosome 4. previously, one QTL was mapped and implicated in across-chromosome epistatic interactions using a F_2 cross. Here we mapped four distinct loci in the same region using the advanced intercross line. One locus was an epistatic capacitor, where the two alleles either released or suppressed the effects of the other three loci. In turn, this capacitor was found to be under the epistatic control of another locus on Chromosome 7. This locus was itself previously thought to be part of a nearby large, additive QTL, *Growth9*. Using the deep intercross population to dissect complex QTL into their components, we not only obtained better estimates and explained variation for the trait, but also gained more insights into the genetic architecture of quantitative trait loci and their phenotypic traits.

403V Bayesian modeling of skewed X inactivation in genetically diverse mice reveals a novel Xce allele and hidden properties of embryonic composition Kathie Sun, Oreper Daniel, Schoenrock Sarah, McMullan Rachel, Giusti-Rodriguez Paola, Zhabotynsky Vasyl, Miller Darla, Tarantino Lisa, Pardo-Manuel de Villena Fernando, *William Valdar* University of North Carolina at Chapel Hill

Female mammals are functional mosaics of their parental X-linked gene expression due to X chromosome inactivation

(XCI). This process inactivates one copy of the X chromosome in each cell during embryogenesis and that state is maintained clonally through mitosis. In mice, the choice of which parental X chromosome remains active is determined by the X chromosome controlling element (Xce), which has been mapped to a 176-kb candidate interval. A series of functional Xce alleles has been characterized or inferred for classical inbred strains based on biased, or skewed, inactivation of the parental X chromosomes in crosses between strains. To further explore the function structure basis and location of the Xce, we measured allele-specific expression of X-linked genes in a large population of F1 females generated from Collaborative Cross (CC) strains. Using published sequence data and applying a Bayesian "Pólya urn" model of XCI skew, we report two major findings: 1) inter-individual variability in XCI suggests mouse epiblasts contain on average 20-30 cells contributing to brain. 2) CC founder strain NOD/ShiLtJ has a novel and unique functional allele that is the weakest in the Xce allelic series, likely the product of copy number variation in the region. In this talk, we will focus on finding (1), illustrating how X-inactivation patterns in adult mice can reveal retrospective properties of the day 5 embryo.

404V Variation in chromatin determines genotype-by-environment interaction in *Drosophila melanogaster* diapause *Abigail DiVito Evans*[1,2], Regina Fairbanks[3], Paul Schmidt[1], Mia Levine[1,2] 1) Department of Biology, University of Pennsylvania, Philadelphia, PA; 2) Epigenetics Institute, University of Pennsylvania, Philadelphia, PA; 3) Department of Evolution and Ecology, University of California, Davis, Davis, CA

Spatially- and temporally-varying selection can shape variation of heritable phenotypes across geographic space and seasonal time. Many such phenotypes exhibit genotype-dependent plastic responses to the environment (genotype-by-environment interaction, "GxE"). Building evidence suggests genotype-specific gene expression programs mediate observed patterns of GxE. However, the mechanisms that determine this gene expression variation, and ultimately phenotypic variation, remain poorly described. Genomic DNA complexed with specialized packaging proteins (i.e. chromatin) is a classic, environment-sensitive determinant of gene expression. Surprisingly, few studies have addressed the role of chromatin in mediating GxE. To fill this gap, we exploit an environment-sensitive phenotype that varies across both geographic space and seasonal time in a tractable model system. *Drosophila melanogaster,* in response to the cold temperatures and short days of oncoming winter, enters reproductive arrest, called diapause. Diapause is a complex trait characterized by extensive physiological changes that result in suspended egg production. In inbred *D. melanogaster* strains that lack genetic variation, only a portion of individuals in the same environment enter diapause. This phenotypic variation in the absence of genetic variation is a classic signature of chromatin-based epigenetic regulation, and allows us to compare diapausing ovaries and reproductive ovaries under the same environmental conditions. To identify how chromatin mediates genotype-dependent diapause incidence, we exploit two inbred lines that span the spectrum of diapause incidence in response to simulated winter conditions: a high diapause (HD) line with 90% diapause incidence and a low diapause (LD) line with 5% diapause incidence. We discovered that silent histone marks H3K27me3 and H3K9me3, and active mark H3K36me1, are invariant across HD and LD ovaries. In contrast, the active histone mark H3K4me3 is depleted in HD diapausing ovaries while active marks H3K27ac and H3K9ac are depleted in LD reproductive ovaries. This genotype-dependent histone mark abundance implicates genotype-specific gene regulation. Consistent with this hypothesis, gene expression in diapausing ovaries only partially overlaps across genotypes. To determine if genotype-specific histone marks are causally linked to diapause incidence, we used ovary-specific RNAi to deplete histone mark "writer" and "eraser" enzymes. Depleting H3K4me3 increases diapause incidence in both lines and enriching H3K4me3 decreases diapause incidence in the HD line only. Moreover, decreasing H3K9me3 and H3K36me3 in the LD line increases diapause incidence, but has no effect on the HD line. This study is the first to reveal a causal role of chromatin-based, epigenetic regulation in mediating an adaptive, genotype-dependent response to the environment.

405V Tango of Two Genomes: Cytonuclear Interactions Underlying Clock and Growth Robustness in Barley Schewach Bodenheimer[1,2], Eyal Bdolach[1,3], Eiji Yamamoto[4], Lalit Tiwari[1], Dan Koening[5], *Eyal Fridman*[1] 1) ARO; 2) HUJI; 3) BGU; 4) MEIJI; 5) UCR

Although there is evidence that phenotypic plasticity can facilitate the evolution of fixed traits in general, it remains an open question whether it is stability or variability of the different characteristics that have adaptive value and therefore be selected? It remains to explore how much of left-behind variation under crop domestication is adaptive under current climate change scenarios and to what extent plasmotype (chloroplast and mitochondria) condition effects of nuclear quantitative trait loci. To start answering these questions, we have modified the classic NAM design to include plasmotype with nucleotype segregation between cultivated barley (*Hordeum vulgare* cv Noga) and ten diverse *H. vulgare* ssp. *spontaneum* wild barley ecotypes. Preliminary analysis of this first Cytonuclear Multi-Parent population (CMPP) shows apparent effects of the plasmotype on growth for the majority of wild donors. Comparative whole-genome and chloroplast sequencing guide follow-up zooming on candidate loci by genome editing, including directed recombination for finer-mapping. We will discuss the caveats, bottlenecks, and solutions to follow-up experiments. Current and future

examination of this CMPP population in the SensyPAM clock platform and across different environments in Israel and the US will allow a better understanding of GxGxE interactions, including questioning the possible role of pleiotropic effects of clock plasticity and robustness on agricultural yield.

406V Sex-Dependency of Epistatic Interactions in the Hybrid Mouse Diversity Panel *Anna Miller*, Scott Williams, David Buchner Case Western Reserve University, Cleveland, OH

Complex traits are influenced by many genetic variants that together determine the presentation and prevalence of phenotypes. How variants work together, additively or non-linear via epistasis, is unclear. If epistasis contributes to most complex diseases, an improved understanding of the complex architecture of complex traits will be critical to guide precision medicine-based decisions. The Hybrid Mouse Diversity Panel (HMDP) is a collection of over 100 inbred strains that can be used to detect and analyze genetic and environmental factors underlying complex traits. Previously reported HMDP phenotypic and genotypic data were analyzed in sex-stratified analyses for eight complex metabolic traits, to identify epistatic interactions. FaST-LMM, a linear mixed model method that accounts for population structure, measures single locus and two loci interaction associations, was used. In females, no marginal effects and 8 epistatic interactions in fat mass were detected using stringent Bonferroni multiple-testing correction. In males, significant marginal effects and interactions were detected for adiposity (58 marginal, 56,112 interactions), body weight (22 marginal, 66,296 interactions) and fat mass (60 marginal, 111,749 interactions). Interactions were also detected for cholesterol (57) and triglycerides (123). The relative number of detected main effects and interactions were comparable within each trait, as the number of pairwise interactions tested were more than 8,000 times greater than the number of main effects tested. Effect sizes and directions indicate that while the interactions were not significantly replicated between sexes at a Bonferroni adjusted level of significance, 80% or more of each trait's interactions show the same direction of effect in the two sexes, significantly greater than the 50% expected by chance. Beyond the statistically significant interactions, there was also a significant enrichment in the number of marginal effects and interactions with a nominal p-value < 0.05 for all traits in both sexes, except for marginal effects on HDL in females. These results collectively demonstrate that the HMDP is a useful tool for detecting genetic architecture, especially interactions, even for traits that lack significant marginal effects. In conclusion, analysis of the HMDP enabled the discovery of marginal and interaction effects that may differ between sexes.

407V The Candidate Chromosomal Regions Responsible for Milk Yield of Cow: A GWAS Meta-Analysis Lida Taherkhani[3], *Mohammad Hossein Banabazi*[1,2], Naser Emamjomeh-Kashan[3], Alireza Noshary[4], Ikhide G. Imumorin[5] 1) Department of Biotechnology, Animal Science Research Institute of Iran (ASRI), Agricultural Research, Education & Extension Organization (AREEO), Karaj 3146618361, Iran; 2) Department of animal breeding and genetics (HGEN), Centre for Veterinary Medicine and Animal Science (VHC), Swedish University of Agricultural Sciences (SLU), Uppsala 75007, Sweden. ; 3) Department of Animal Science, Science and Research Branch, Islamic Azad University, Tehran 1477893855, Iran; 4) Department of Animal Science, Karaj Branch, Islamic Azad University, Karaj, Iran; 5) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

Milk yield (MY) is highly heritable and an economically important trait in dairy livestock species. To detect the candidate genomic regions for this trait, we did a meta-analysis on genome-wide association studies (GWAS). In the present study, nineteen studies were resulted from web searching on PubMed and journals. Only studies were reviewed and included in the meta-analysis that passed the intended filters. Then was extracted and recorded chromosome number, detected markers and their position, number of animals, and *P-value* was extracted from these studies and recorded. Finally, this study was done with 16 GWAS studies, a total 353,698 cows, and 3,950 markers. The pooled data were analyzed METAL software. Our findings revealed 1,712 significant (*P-value* < 2.5×10-6) genomic loci related to milk yield trait. In addition, gene ontology (GO) was used to explore the biological functions of the genes underlying the significant SNP markers associated with milk yield. Our findings showed the associated markers with the studied trait are spreading on all of the bovine autosomal and sexual chromosomes. But, they are higher dense and significant on a few specific regions of chromosome 14. A comparative transcriptome coverage between two bovine subspecies (*bos indicus vs. bos taurus*) confirmed the main contribution of these regions in milk production. In addition, the significant SNPs located in these regions may be addressed as candidate markers to design an array specifically for genomic evaluations of the milk yield trait.

408V Evaluate Breeding Efficiency of Targeted Recombination for Ordinal Traits Yuan-Chieh Yang, *Yung-Fen HUANG* National Taiwan University

Targeted recombination refers to the occurrence of recombination at a specific genomic position. Such concept has been recently applied to crop breeding and has revealed to be promising for the improvement of quantitative traits. Mean-

while, in a breeding program, traits of interest are frequently scored in ordinal scales, such as disease resistance. However, no studies have investigated whether targeted recombination can be applied to the improvement of ordinal traits. Therefore, the objectives of this study were to use simulated and empirical data: (1) to determine whether the strategy of targeted recombination can be applied to the improvement of ordinal traits, (2) to identify factors influencing on the breeding efficiency of targeted recombination in ordinal traits, (3) to compare the relative efficiency of genotypes carrying targeted recombination estimated based on ordinal data and its compatible continuous data. Our results showed that targeted recombination could effectively improve ordinal traits under various genetic architectures. A bi-parental cross of ca. 150 progenies scored with a five-category scale suffices for marker effect estimation and target site determination. The relative efficiency between targeted recombined genotypes estimated based on ordinal traits and their quantitative counterpart were comparable. Hence, our study suggested that targeted recombination be applied to the improvement of ordinal traits with a similar efficiency of continuous trait.

409V Variation in epigenetic state correlates with gene expression across nine inbred strains of mice *Anna Tyler*, Catrina Spruce, Robyn Ball, Wendy Pitman, Vivek Philip, J. Matthew Mahoney, Mary Ann Handel, Gary Churchill, Jennifer Trowbridge, Michael Stitzel, Kenneth Paigen, Petko Petkov, Gregory Carter The Jackson Laboratory

It is well established that epigenetic features, such as histone modifications and DNA methylation, are associated gene expression across cell types. However, it is not well known how variation in genotype affects epigenetic state, or to what extent such variation contributes to variation in gene expression across genetically distinct individuals. Here we investigated the relationship between heritable epigenetic variation and gene expression in hepatocytes across nine inbred mouse strains. Eight of the inbred strains were founders of the diversity outbred (DO) mice, and the ninth was DBA/2J, which, along with C57Bl6/J, is one of the founders of the BxD recombinant inbred panel of mice. We surveyed four histone modifications, H3K4me1, H3K4me3, H3K27me3 and H3K27ac, as well as DNA methylation. We used ChromHMM to identify 14 chromatin states representing distinct combinations of the four measured histone modifications. We found that variation in chromatin state mirrored genetic variation across the inbred strains. Furthermore, epigenetic variation was correlated with gene expression across strains. The correspondence between epigenetic state and gene expression was replicated in an independent population of DO mice in which we imputed local epigenetic state. In contrast, we found that DNA methylation did not vary across inbred strains and was not correlated with variation in expression in DO mice. This work suggests that chromatin state is highly influenced by local genotype and may be a primary mode through which expression quantitative trait loci (eQTLs) are mediated. We further demonstrate that the mid-range resolution of chromatin states, between that of SNPs and haplotypes paired with gene expression, is useful for annotation of functional regions of the mouse genome. Finally, we provide, to our knowledge, the first data resource to document variation in chromatin state across genetically distinct individuals.

410V Genetic analysis of multi-omics data identifies drivers of protein phosphorylation *Gary Churchill*[2], Tian Zhang[1], Gregory Keele[2], Steven Munger[2], Fernando Pardo-Manuel de Villena[3], Martin Ferris[3], Joao Paulo[1], Steven Gygi[1] 1) Harvard University, Cambridge, MA; 2) The Jackson Laboratory, Bar Harbor, ME; 3) University of North Carolina at Chapel Hill, Chapel HIll, NC

Phosphorylation is a chemical process in which a phosphate group is added to an organic compound. Site-specific phosphorylation of proteins regulates many cellular processes including activation of enzymes and signaling cascades. The abundance of phosphorylated peptides can be measured by mass spectrometry and is determined by the abundance of the target protein/peptide and the proportion of target sites that are phosphorylated. We quantified the abundance of phosphorylated peptides, as well as proteins and transcripts in heart, liver, and kidney tissue samples from 116 genetically diverse female and male mice of the Collaborative Cross strain panel. We identified ~700 phosphorylation quantitative trait loci (phQTL) in the three tissues and applied genetic mediation analysis to identify causal drivers of phosphorylation. Candidate mediators included kinases, phosphatases, cytokines, and other factors with established roles as well as identifying novel interactions between target proteins and genes that regulate protein phosphorylation. Our analysis highlights multiple targets of pyruvate dehydrogenase kinase 1 (PDK1), a regulator of mitochondrial function that shows reduced activity in the NZO/HILtJ mouse, a polygenic model of obesity and type 2 diabetes.

411V Gene expression noise in a pathway is condition-specific *David Laloum*, Raquel Assis Florida Atlantic University

Gene expression noise refers to variation in gene expression among genetically identical cells. It can arise from diffusion, growth, or a particular combination of transcription factor, promoter, enhancer, or 3d position in the genome. Although originally thought to be detrimental, recent studies have highlighted that gene expression noise may confer a selective advantage in stressful or changing environments by producing heterogeneous phenotypes. Noise levels of genes should be determined by the structure of the gene regulatory network, and we expect promoters to play important roles in the

propagation of expression noise of their target genes. However, only noise plasticity increases with the number of regulatory inputs of the promoter, without any change in mean expression. Here we study features of noisy genes within the *E. coli* regulatory network. We show that, whereas some noisy genes are generalists, most can be grouped into biological modules in a condition-specific manner. Finally, we compare the evolutionary and expression features of generalist and condition-specific noisy genes, providing insight into the forces driving gene expression noise in *E. coli*.

412V A Bayesian model selection approach to mediation analysis Wesley Crouse[1], *Gregory Keele*[2], Madeleine Gastonguay[2], Gary Churchill[2], William Valdar[1,3] 1) Genetics Department, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) The Jackson Laboratory, Bar Harbor, ME; 3) Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC

Genetic studies often seek to establish a causal chain of events originating from genetic variation through to molecular and clinical phenotypes. When multiple phenotypes share a common genetic association, one phenotype may act as an intermediate for the genetic effects on the other. Alternatively, the phenotypes may be causally unrelated but share genetic loci. Mediation analysis represents a class of causal inference approaches used to determine which of these scenarios is most plausible. We have developed a general approach to mediation analysis based on Bayesian model selection and have implemented it in an R package, bmediatR. Bayesian model selection provides a flexible framework that can be tailored to different analyses. Our approach can incorporate prior information about the likelihood of models and the strength of causal effects. It can also accommodate multiple genetic variants or multi-state haplotypes. Our approach reports posterior probabilities that can be useful in interpreting uncertainty among competing models. We compared bmediatR with other popular methods, including the Sobel test, Mendelian randomization, and Bayesian network analysis using simulated data. We found that bmediatR performed as well or better than these alternatives in most scenarios. We applied bmediatR to transcriptome and proteome data from Diversity Outbred (DO) mice, a multi-parent population, and demonstrate the power of mediation with multi-state haplotypes. We also applied bmediatR to data from human cell lines to identify transcripts that are mediated through or are expressed independently from local chromatin accessibility. We demonstrate that Bayesian model selection provides a powerful and versatile approach to identify causal relationships in genetic studies using model organism or human data.

413V Increasing power in inbred strain association mapping by recognizing variance heterogeneity Marissa Ashner, Robert Corty, *William Valdar* University of North Carolina at Chapel Hill

Modern quantitative trait locus (QTL) mapping in panels of inbred strains uses a linear mixed model (LMM) to test for SNP-phenotype association while accounting for a random effect of population structure. A decade of mathematical tricks have mitigated the computational expense of repeatedly fitting this complex model for genome-wide applications. Existing procedures, however, make the assumption that the phenotype of each strain (or individual) is known with equal precision. In reality, this assumption does not always hold. We propose a method, weighted Inbred Strain Association Mapping (wISAM), which accounts for heteroscedastic residual variance in the study population using a weighted regression technique and makes use of variance shrinkage methods to stably estimate these weights. Simulation studies comparing wISAM to existing methods demonstrate that it can provide additional statistical power for GWAS. The method is then illustrated using data from studies incorporating the Hybrid Mouse Diversity Panel (HMDP).

416W StRoNG Net: Advancing undergraduate opportunities in non-model genome research *Nicholas Miller*[1], Molly McDonough[2], Cindy Voisine[3], Aaron Schirmer[3] 1) Illinois Institute of Technology; 2) Chicago State University; 3) Northeastern Illinois University

Thanks to rapid technological advances in DNA sequencing, obtaining the genome sequences of non-model organisms is now feasible for individual labs, even in cases of species with relatively large, complex genomes. This is opening up a wealth of new opportunities in the fields of population and evolutionary genomics.

Genome sequence data is accumulating so quickly that the main bottleneck to understanding it is a large enough scientific workforce to analyze the data. This is especially true for non-model organisms, which are often studied by small to moderately sized research communities. Undergraduates are a sometimes-overlooked pool of trainee scientists who can assist in the annotation and analysis of new genome sequences.

Equally, participating in research is known to benefit undergraduates' educational, professional and personal development. Unfortunately, access to opportunities in laboratory and field research is often limited and sometimes unequally distributed. Much of the analysis of genome data can be done using software accessed through a web interface. Because of this, students only need a web browser to do research, making opportunities accessible to a diverse population.

StRoNG Net (Stem Research on Non-model Genomes Network) is a network of scientists and educators dedicated to developing opportunities for undergraduates to engage in scientific research by contributing to the analysis of non-model genomes.

417V Integrating genetic incompatibility research and research ethics training in a course-based undergraduate research experience (CURE) Joseph Ross California State University, Fresno

Improving and assuring public trust in science is globally valuable, yet training in the responsible and ethical conduct of research (RECR) is often first made available to graduate students. Course-based undergraduate research experiences (CUREs) are classes in which students participate in the entire process of scientific inquiry, from project conception through data analysis and dissemination. Thus, CUREs expose students to circumstances in which they will need to employ ethical decision-making, for example in data collection and dissemination (data management practices), data analysis and display (awareness of what constitutes research misconduct), and authorship on research group presentations and publications. CUREs are becoming more prevalent in undergraduate curricula, partly because of their ability to democratize access to research experiences. So, CUREs are ideally poised to facilitate RECR training for developing scientists. I previously redesigned a Cell Biology and Genetics Lab course (BIOL 104) as a CURE, in which students conduct background literature reading and develop and experimentally test a hypothesis about genetic incompatibilities between populations of the microscopic nematode *Caenorhabditis briggsae*, a relative of *C. elegans*. I recently added direct instruction on six "core" RECR topics defined by the US Department of Health and Human Services. To assess the perceived value of RECR training to undergraduates, and to measure the efficacy of that training, pre- and post-semester instruments were used to collect attitude and RECR knowledge data. Student artifacts, such as written responses to prompts, were also collected. Despite participants being predominantly third and fourth year undergraduates, sixteen percent did not initially accurately identify which sentences in a peer-reviewed research manuscript should contain citations. Two-thirds of students were not familiar with policies related to image manipulation. Following instruction, students made significant gains in understanding the ethical norms for the six RECR topics. Perhaps most critically, the participants did not feel that undergraduate scientists have a role in monitoring and alerting others about research ethics violations. These results both underscore the need to provide more direct training of RECR and also highlight the role that CUREs can play in meeting this need. The presentation will also describe the RECR instructional materials and activities, as well as a quantitative evaluation of RCR student learning outcomes.