2018 Population, Evolutionary, and Quantitative Genetics Conference

Full Abstracts

Lightning Talks 1 Uncovering the genotype-phenotype-fitness map of microbes adapting to novel environments. *Grant Kinsler*, Kerry Geiler-Samerotte, Dmitri Petrov Dept. of Biology, Stanford University, Stanford, CA.

Experimental evolutions using DNA barcodes to track millions of independent evolving lineages have recently quantified the spectrum of unique single mutations that can each help microbes adapt to glucose limitation. But how many unique physiological processes are represented by this large pool of beneficial mutations? How many unique ways are there to improve microbial fitness in glucose-limited conditions, or in other novel environments? Using recent developments in DNA barcoding, we precisely measure the fitness of evolved lineages in the environment they evolved in as well as many environments that only slightly differ from this original condition. These data allow us to understand the pattern of correlations amongst adaptive mutants across subtly differing conditions, estimate the number of unique fitness-relevant phenotypes represented by these mutants, and uncover the genotype-phenotype-fitness map for these adaptive lineages. We find evidence that single genetic mutations affect many phenotypes (pleiotropy). Despite this observation, we show that only a small number traits matter for adaptation to the original evolution condition. In particular, our finding sheds light on how adaptation can proceed despite widespread pleiotropy, the key being that not all phenotypes affected by mutation have fitness effects in the current environment. This has wide-ranging implications on how adaptation proceeds in complex phenotype space, specifically regarding the extent to which adaptation is limited by tradeoffs and how environmental dependencies influence the relationship between phenotype and fitness.

Lightning Talks 1 The potential of regularized regression to provide more accurate multivariate selection estimates. *J. Sztepanacz*[1,2], D. Houle[1], T.F. Hansen[2] 1) Florida State University, Tallahassee, FL; 2) Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway.

The breeder's equation $\Delta\mathbf{z}=\mathbf{G}\boldsymbol{\beta}$, allows us to understand how genetics (the genetic covariance matrix **G**) and the vector of linear selection gradients selection ($\boldsymbol{\beta}$) interact to generate evolutionary trajectories. Estimation of $\boldsymbol{\beta}$, using multiple regression of trait values on relative fitness, revolutionized the way we study selection in laboratory and wild populations. Multicollinearity, or correlation of predictors, is a major challenge for any multiple regression approach, that can lead to very high variances and covariances between elements of $\boldsymbol{\beta}$. The usual approach to multicollinear predictors is to discard some of them, thereby losing any information that might be gained from those traits. Using simulations, we show how, on the one hand, multicollinearity can result in inaccurate estimates of selection, and on the other how the removal of correlated phenotypes from the analyses can provide a misguided view of the targets of selection. We show that regularized regression, which places a priori constraints on features of $\boldsymbol{\beta}$, generates more accurate estimates of selection in the presence of multicollinearity. We compare standard and regularized regression estimates of selection on sexual pheromone signals in the Australian fruit-fly *Drosophila serrata*. This analysis, and our simulations suggest that regularized regression should be adopted as a valuable tool in selection analyses.

Lightning Talks 1 The genomic and molecular basis of rapid and polygenic response to selection for long leg length in mice. J.P.L. Castro[1], M.N. Yancoskie[1], M. Marchini[2], S. Belohlavy[3], M. Kucka[1], W.H. Beluch[1], R. Naumann[4], I. Skuplik[2], J. Cobb[2], N.H. Barton[3], C. Rolian[2], *Y.F. Chan*[1] 1) Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany; 2) University of Calgary, Calgary, AB, Canada; 3) IST Austria, Klosterneuburg, Austria; 4) Max Planck Institute for Cell Biology and Genetics, Dresden, Germany.

A major goal in evolutionary genetics is to understand how genomes evolve in response to selection. Here we present a genomic dissection of the Longshanks selection experiment, in which mice were selectively bred for longer tibiae relative to body mass over 20 generations, resulting in 13% increase. We combined whole genome sequencing, modeling and population genetics to show that >100 loci contributing to polygenic selection response (See abstract by N. Barton for modelling the Longshanks experiment). Here we will focus on the evolutionary processes underlying loci responding independently and in parallel between the two Longshanks replicate lines. Out of 329 putatively selected outlier loci, more than half of the loci (56%) were specific to one of two Longshanks lines, vs. 27% showing parallel response and 17% other patterns. However, we found that as the selection responses increases, parallelism became increasingly likely. The selected loci showed significant enrichment for limb developmental genes and enhancers, with the impact of *cis*-acting but not coding changes positively correlating with the strength of selection signature in both Longshanks selection replicate lines. Through functional testing of enhancers at *Gli3* and *Nkx3-2*, we show that both gain- and loss-of-function enhancer variants contributed to selection response. Using the Longshanks experiment we were able to obtain a detailed picture of the polygenic selection response and demonstrate the critical role of regulatory changes for specific loci in rapid intraspecific evolution of a major morphological trait.

Lightning Talks 1 Gene, environment and cellular interactions underlying behavioral variance and their relation to fitness in a long-term *Caenorhabditis elegans* evolution experiment. *L.M. Noble*[1], T. Guzella[2], F.M. Mallard[2], M.V. Rockman[1], H. Teotónio[2] 1) Center for Genomics and Systems Biology, New York University, New York, NY; 2) Institut de Biologie, École Normale Supérieure.

By evolving *C. elegans* from multiparental standing genetic variation we have generated a sequenced panel of recombinant inbred lines with which to study the properties and evolution of genetic architectures. We measured fitness, precisely defined under the experimental regime, and morphological and behavioral traits that vary in their alignment with fitness, in more than 400 lines, under familiar and novel conditions. Additive architectures for fitness and closely related traits are extremely polygenic, as expected, but strong sign epistasis with weak marginal effects also accounts for a large fraction of trait variance.

The converse is true for behavioral traits, which show weaker (phenotypic and additive genetic) correlations with fitness. And, consistent with results from other systems, additive effects are relatively consistent across environments while epistatic interactions are much less so. Using whole animal single cell gene expression data we see that the inferred cellular basis of behavioral variance is strongly dependent on environment, evolutionary history, and fitness alignment. The expression of interacting loci underlying variance in behavior and fitness-related traits also differs markedly, consistent with variable pleiotropy.

Lightning Talks 1 Gene expression drives the evolution of dominance. C. Huber[1], A. Durvasula[2], A. Hancock[3], *K. Lohmueller*[1,2] 1) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA; 3) Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany.

Dominance is a fundamental concept in molecular genetics and has implications for understanding patterns of genetic variation, evolution, and complex traits. However, despite its importance, the degree of dominance has yet to be quantified in natural populations. Here we develop a novel composite likelihood approach that leverages genetic variation data from outcrossing and selfing species to co-estimate the distribution of selection coefficients (*s)* and the dominance coefficient (*h)*. Since selection acts immediately on recessive homozygotes in self-fertilizing organisms, the genetic variation data from a selfing species allows us to discriminate between different values of *h.* Application of our approach to amino acid changing mutations in *Arabidopsis* suggests that most mutations are recessive and that more deleterious mutations tend to be more recessive than less deleterious mutations. We next use our data to test the existing models for the mechanism of dominance. For example, Fisher's model suggests that dominance arose via modifier mutations at other loci and that these loci are subject to selection. Wright, and later Kacser and Burns, proposed a metabolic theory model where mutations in enzymes are predicted to be recessive because the overall flux through a metabolic network is robust to decreasing the amount of one of the enzymes of the pathway by one-half. We find that neither of these models for the evolution of dominance can explain how the inferred relationship between *h* and *s* varies with gene expression level and connectedness of genes. Thus we develop a new model for the evolution of dominance. Our new model predicts that dominance arose as a consequence of the functional importance of genes and their optimal expression levels. Our model matches many of the salient features of the data.

Lightning Talks 1 The unreasonable effectiveness of population genetic inference via image recognition. *D.R. Schrider* University of North Carolina, Chapel Hill, NC.

The availability of population-scale genomic datasets has given researchers a new avenue toward answering questions about populations' recent evolutionary histories. To what extent is a population's genomic variation shaped by the interplay between natural selection and recombination? Has the population experienced substantial changes in size? Has the population experienced gene flow from closely related populations/species, and if so, which portions of the genome were affected? In recent decades a host of theoretical and methodological advances have addressed these problems. Typically these methods summarize patterns of genetic variation with a statistic designed to be sensitive to the phenomenon of interest. More recently, approximate Bayesian computation and machine learning approaches have proved successful in simultaneously examining many of these statistics in order to make far more accurate inference. The rationale for these approaches is that any single statistic will capture only a subset of the discriminatory information present in the original data, and thus a set of complementary statistics will perform better. A more fruitful approach would thus be to perform inference directly on the input sequence data rather than digesting it into a set of numbers. Here we attempt to accomplish this by representing a population genetic alignment as an image and using modern deep learning techniques for image processing. We apply this approach to the problems listed above, and find that in each case it matches or exceeds the accuracy of current state-of-the-art methods. Thus, when applied to images of alignments, modern image recognition algorithms outperform expert-derived statistics and even collections thereof. In light of this result, we argue that the rate of progress in evolutionary genetic inference might be improved by devoting greater effort to exploring the myriad possible image representations of sequence alignments and deep learning architectures for processing them, rather than attempting to devise more powerful summary statistics.

Lightning Talks 1 Methods for detecting selection in admixed populations. *Erin Calfee*, Daniel Gates, Jeffrey Ross-Ibarra, Graham Coop University of California, Davis, CA.

New genetic variation from admixture has the potential to drive rapid adaptive change, at the time scale of ecological processes such as range shifts, competition and invasions. Recent progress has been made to accurately infer the mosaic of local ancestry across admixed genomes and identify outlier loci under selection. However, current methods to test for selection are not robust to drift, which limits our ability to analyze admixed populations with complex or unknown demographic histories. We show how shared evolutionary history post-admixture can be inferred from genome-wide patterns of ancestry variance and covariance. Analogous to a phylogenetic independent contrast, we correct for this background non-independence between sampled populations, mitigating false-positives. Our proposed statistical framework can be used to identify loci with an excess or deficit of a specific ancestry, or an association between ancestry and an environmental variable, beyond what can be explained by shared drift among sampled populations. We apply these methods to identify signatures of parallel adaptation in populations of admixed *Zea mays* across Mexico. This work adds fine-scale genomic resolution to previous findings that adaptive introgression from a highland-endemic wild relative, *mexicana*, facilitated maize's range expansion to colder and more UV-intense high altitude environments. The methods presented here can be readily applied to other systems to understand more generally how admixture shapes range shifts or invasions and how consistently specific loci adaptively introgress or maintain barriers to gene flow across different populations or environments.

Lightning Talks 1 Population complexity trumps model complexity in understanding trait variation. *M.G. Sterken*, L.B. Snoek, R.P.J. Bevers, R.J.M. Volkers, J.A.G. Riksen, J.E. Kammenga Laboratory of Nematology, Wageningen University, Wageningen, NL.

The study of expression quantitative trait loci (eQTL) through the use of recombinant inbred lines has yielded detailed information about the transcriptional regulation of complex traits. However, it has proven difficult to apply more advanced genetic models explaining genetic variation underlying gene expression differences. Here, we make use of the difference in genetic complexity of two types of inbred population in the nematode *Caenorhabditis elegans* to estimate the number of loci affecting gene expression.

We measured gene-expression in a recombinant inbred line (RIL) and an introgression line (IL) population constructed from crossing the strains N2 and CB4856. Both populations received a heat-shock treatment and gene-expression profiles were obtained before (48h at 20ºC), directly after heat-shock (2h at 35ºC), and after a recovery period (2h at 20ºC). Making use of the difference in genetic make-up between the populations - few loci from one parent in the IL versus many in the RILs - allowed for the identification of transcripts regulated by multiple loci. By measuring the transcript variance within each population, for over 1,000 genes across the three conditions we found strong evidence for multiple eQTL underlying gene expression variation. Importantly, most of these multi-loci eQTL are environment-specific. Furthermore, we observed over 200 genes where the phenotypic variation in the IL panel significantly exceeded that in the RIL panel, suggesting evidence for complex genetic buffering.

In conclusion, by using two types of inbred populations the complexity of trait architectures can be investigated without reliance on models of higher complexity. The genetic complexity of a trait is directly observed, rather than estimated *post-hoc*. Therefore, relying on *population complexity* rather than *model complexity* can provide valuable insight in the architecture of quantitative traits.

Lightning Talks 1 Selection minimizes introression around incmopatabilitie and regions under strong linked selection in *Capsella*. *Y.J. Brandvain*[1], Tyler Kent[2], Stephen Wright[2], Krzsztof Stankiewicz[3] 1) Plant and Microbial Biology, University of Minnesota, St Paul, MN; 2) Department of Ecology & Evolutionary Biology. University of Toronto. Toronto, ON; 3) Mathematics and BioSciences Group. University of Vienna . Vienna, Austria.

Upon secondary contact some genomic regions rapidly introgress while others remain distinct. We wish to know what determines these outcomes. We examine the density of ancestry from Capsella rubella across 180 genomes ot its sister species, Capsella rubella. We find that a parameter that summarizes the extent of linked selection strongly predicts genome-wide heterogeneity in the extent of introgression introgression. This result suggests that, rather than selecting against a small number of incompatibilities or differential adapted alleles, selection acts to remove introgression across most functional regions of the genome. In the face of this strong, genome wide trend, we also find a dearth of introgression on a chromosome arm in which C. rubella carries a derived incompatibility.

Lightning Talks 1 To TE, or not to TE, that is the question: transposable element dynamics in hybrid and naïve genomes. *C. Smukowski*, M. Dunham Genome Sciences, University of Washington, Seattle, WA.

Transposable elements (TEs) are repetitive, mobile DNA elements that can have deleterious effects on their hosts by triggering changes in gene expression, chromosomal rearrangements, and genome size expansion. Due to these consequences, TEs and their hosts exhibit an evolutionary arms race scenario in which hosts evolve mechanisms to silence transposition and TEs mutate to maintain transposition, successfully colonizing nearly every organism across the tree of life.

The yeast *Saccharomyces uvarum*, a relative of *Saccharomyces cerevisiae*, is therefore quite unusual, as it has expelled its' TEs, leaving a genome with no active TEs and only small fragments of former TEs. One of only a handful of organisms ever observed with this pattern, *S. uvarum* thus represents a unique model for understanding the genetic and environmental factors that regulate transposition and how TEs colonize naïve genomes. Utilizing laboratory experimental evolution, we have found novel TE insertions in evolved *S. cerevisiae* haploids and diploids, but no new insertions after hundreds of generations in *S. cerevisiae* x *S. uvarum* evolved hybrids, suggesting active suppression of transposition. To test this hypothesis, we have directly assayed transposition rate in these and other interspecific hybrids, in addition to *S. uvarum* with an artificially introduced TE. Finally, we propose and test a genetic mechanism that may be responsible for the observed inhibition of transposition. In summary, we shed new light on one of the most important interactions driving genome evolution across every domain of life.

Lightning Talks 2 A phylogenetic analysis of the *Drosophila* metabolome. *D. Promislow*[1], J. Hoffman[2] 1) Department of Pathology and Department of Biology, University of Washington, Seattle, WA; 2) Department of Biology, University of Alabama at Birmingham, Birmingham, AB.

In recent years, researchers have begun to incorporate measures of 'omic' domains (metabolome, transcriptome, etc.) in attempts to build more accurate genotype-phenotype maps, capturing at least some of the missing heritability typical of GWA studies. To bridge the gap between genotype and phenotype, we have focused in particular on the metabolome. The metabolome measures the hundreds of small molecules that make up the structural and functional building blocks of all organisms. Surprisingly little is known about how this complex high-dimensional 'trait' evolves over time. We carried out a comparative analysis of genetic and phylogenetic diversity in the metabolome across 50 million years of evolution in the genus *Drosophila*. We measured both targeted and global metabolomic profiles using mass spectrometry. Comparative analysis of such data is challenging, given the very large number of variables. We present several approaches that attempt to address the statistical challenges of combining phylogenetic comparative analysis and high-dimensional data. We find that the metabolome has a strong phylogenetic signature, and identify some of the first evidence for metabolome gain/loss over evolutionary time. We also identify metabolites with evolutionary conserved patterns of change between between sexes and across ages. In addition to presenting our findings, we will discuss the challenges, opportunities, and some possible solutions in bringing a phylogenetic perspective to high-dimensional physiologically relevant data.

Lightning Talks 2 *Crossover positions influence the total amount of genetic shuffling*. *Carl Veller*[1,2] 1) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Program for Evolutionary Dynamics, Harvard University, Cambridge, MA.

The total amount of genetic shuffling that occurs in gamete production—the genome-wide recombination rate—is a critical quantity for comparative studies in evolutionary genetics, but it has never been measured directly. Traditional measures simply count the average number of crossovers per meiosis, but a crossover in the middle of a chromosome causes more genetic shuffling than a crossover at the tip, and independent assortment of homologs also causes genetic shuffling. We propose a measure of the genome-wide recombination rate that takes into account these features: $\bar{r}$, the probability that a random pair of loci recombine in the production of a gamete. $\bar{r}$ is easily measured using modern cytological or sequencing data. We provide the first direct measurement of the total amount of genetic shuffling by estimating $\bar{r}$ in male and female humans, using both cytological and single-gamete sequencing data. $\bar{r}$ can be decomposed into separate components deriving from crossing over and independent assortment of chromosomes. Performing this decomposition for humans, we find that about 30 times more genetic shuffling derives from independent assortment than from crossovers. We show that $\bar{r}$ is larger when crossovers are more evenly spaced, with the intriguing implication that crossover interference will tend to increase genetic shuffling. $\bar{r}$ should be used in comparative studies when the total amount of genetic shuffling is a variable of interest.

Lightning Talks 2 Comprehensive identification of *cis*-regulatory variants in yeast promoters. *F.W. Albert*[1,6], R. Cheung[2,6], L. Day[3,4,5], S. Kosuri[2], L. Kruglyak[3,4,5] 1) Department of Genetics, Cell Biology, & Development, University of Minnesota, Minneapolis, MN; 2) Department of Chemistry & Biochemistry, University of California, Los Angeles, CA; 3) Department of Human Genetics, University of California, Los Angeles, CA; 4) Department of Biological Chemistry, University of California, Los Angeles, CA; 5) Howard Hughes Medical Institute, University of California, Los Angeles, CA; 6) joint first authors.

Regulatory genetic variation in a species influences gene expression levels and is a key source of variation in complex traits. Considerable progress has been made in identifying genomic regions that harbor regulatory DNA variants (expression quantitative trait loci; "eQTL"). However, resolving these regions to the underlying causal variants remains challenging due to linkage between neighboring variants. Our limited ability to identify causal variants at scale precludes a systematic understanding of how natural regulatory variation shapes gene expression and complex traits.

To tackle this challenge, we have developed a massively parallel reporter assay to comprehensively identify *cis*-regulatory variants in yeast (*Saccharomyces cerevisiae*) promoters. Our assay probes DNA variants in promoters in two ecologically and genetically different strains: a laboratory strain and a vineyard isolate. Each variant is represented by pairs of synthetic promoter fragments that differ only in the given variant allele. We synthesized a library of 27,000 synthetic DNA

oligonucleotides to assay 9,605 natural variants in 3,848 promoters. The library covers two thirds of all putative *cis*-acting variants in these strains, including all variants within 200 bp immediately upstream of each gene. We also synthesized pairwise allele combinations for closely linked variants, allowing us to measure non-additive interactions. We cloned our libraries into reporter plasmids *en masse*, such that each promoter fragment is coupled to hundreds of unique expressed barcodes. We quantify gene expression driven by each promoter fragment via high-throughput sequencing of these barcodes. We determine the effect of each variant by comparing the expression driven by alleles that differ only at the given variant. Our design assays *cis*-acting variants comprehensively and with single variant resolution.
Using a subset of our library, we have so far identified 315 *cis*-acting alleles. These variants include several that destroy or create transcription factor binding sites between the two strains. We are currently finalizing data collection for the entire library and anticipate discovery of hundreds of additional causal variants. By relating these causal variants to the excellent annotations of regulatory elements in yeast, we will ask to what extent the effects of natural promoter variants on gene expression can be predicted from genome sequence alone.

Lightning Talks 2 A Bayesian approach to quantitative genetics for high-dimensional traits. *D. Runcie*[1], J. Ta[1], L. Crawford[2], S. Mukherjee[3] 1) Plant Sciences, University of California Davis, Davis, CA; 2) Department of Biostatistics, Brown University, Providence, RI; 3) Statistical Science, Duke University, Durham, NC.

Statistical models for Genome-Wide Association Studies, QTL analysis, and Genomic Prediction, are the foundation of modern quantitative genetics and crop improvement. Driven by the explosion of whole-genome genotype data, recent improvements to these models allow for analyses of millions of markers at a time. However, similar advances for modeling large phenotype datasets is lacking. New phenotyping technologies collect thousands of observations on each individual plant or line – changes in morphology through time, molecular phenotypes such as gene expression or metabolite levels, or performance measures across multiple environments. Jointly modeling these high-dimensional traits can provide insight into developmental and physiological mechanisms that link genotype and phenotype. We propose a robust and efficient method for modeling the genotype-phenotype relationship of high-dimensional traits. The key idea underlying our model is that groups of traits will be highly correlated due to genetic and developmental pleiotropy. We leverage these correlated modules to prioritize the most important signals in big data. We will demonstrate how our method provides powerful and interpretable estimates of genetic architecture using two high-dimensional datasets: a time-series analysis of growth curves, and a dataset of genome-wide gene expression.

Lightning Talks 2 Genome-wide characterization of differences in mutation fitness effects between populations. Alyssa Fortier[1], Alec Coffman[1], Travis Struck[1], Jose Burguete[2], Aaron Ragsdale[3], PingHsun Hsieh[4], *Ryan Gutenkunst*[1] 1) Molecular and Cellular Biology, University of Arizona; 2) Center for Genomic Sciences, National Autonomous University of Mexico; 3) Applied Mathematics, University of Arizona; 4) Ecology and Evolutionary Biology, University of Arizona.

The fitness effect of a mutation may differ between populations, depending on environmental and genetic context. To quantify genomic patterns of such differences, we extended the concept of a distribution of fitness effects (DFE) to a joint DFE between populations. To infer the joint DFE, we fit parametric models that included demographic history to genomic data summarized in the joint allele frequency spectrum. We applied this framework to African and European populations of both Drosophila and humans, finding that mutation fitness effects are much more similar between populations of humans than Drosophila. Among gene sets, genes involved in immunity typically showed low similarity of fitness effects, whereas genes involved in reproduction showed high similarity. Our results represent the first genome-scale quantification of mutation fitness effect differences between populations and point toward gene functions that are more likely to experience divergent selection.

Lightning Talks 2 A map of highly constrained coding regions in the human genome. *A.R. Quinlan*, J.M. Havrilla, B.S. Pedersen, R.M. Layer Human Genetics, University of Utah, Salt Lake City, UT.

Deep catalogs of genetic variation collected from many thousands of humans enable the detection of intraspecies constraint by revealing coding regions with a scarcity of variation. While existing metrics such as RVIS and pLI summarize constraint for entire genes, single metrics cannot capture the fine-scale variability in constraint within each protein-coding gene. To provide greater resolution, we have created a detailed map of constrained coding regions (CCRs) in the human genome by leveraging coding variation observed among 123,136 humans from the Genome Aggregation Database (gnomAD). The most constrained coding regions in our map are enriched for both pathogenic variants in ClinVar and de novo mutations underlying developmental disorders. While observed regions of constraint are generally correlated with interspecies conservation, many CCRs are highly constrained in the human lineage but not strongly conserved across species. CCRs also reveal protein domain families under extreme constraint, suggest unannotated or incomplete protein domains, and facilitate the prioritization of previously unseen variation in studies of disease. We explore the tissue specificity of and impact of CCRs on protein-protein interactions. Finally, we demonstrate that a subset of the most constrained CCRs likely exist within genes that cause yet unobserved human phenotypes owing to strong purifying selection.

Lightning Talks 2 Convergent regressive evolution of the eye and the identification of eye-specific regulatory elements. R. Partha[1], B.K. Chauhan[2], Z. Ferreira[1], J.D. Robinson[3], K. Lathrop[2], K.K. Nischal[2], M. Chikina[1], *N.L. Clark*[1] 1) Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA; 2) Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) Molecular and Cell Biology, University of California, Berkeley, CA.

Several lineages of subterranean mammals have independently adopted an exclusively underground lifestyle. Extreme examples include the naked mole-rat, blind mole-rat, star-nosed mole, and cape golden mole. In response to their dark environments, these species evolved greatly reduced eyesight in a process known as regressive evolution. Using our new comparative genomic methods we found that hundreds of genes show parallel increases in evolutionary rate specifically in these subterranean species. Most genes were accelerated due to loss of functional constraint, chiefly those involved in eye physiology, such as lens crystallins and photoreceptors, while other genes were accelerated in an apparent adaptation for tunneling and excavating through hard substrate. Regressive evolution proceeded very differently between eye tissues; while lens and retinal genes are highly accelerated, corneal genes remain under constraint, perhaps because they continue to provide a protective outer barrier for the vestigial eye. Moreover, genes important for embryological eye development remain highly conserved, likely because they are important for the development of other tissues. For example, the coding portion of *PAX6*, a key transcription factor in eye, forebrain and pancreas, showed no signs of regression. In contrast, we found that *PAX6's* eye-specific enhancers were evolving at a much faster rate in subterranean species, likely due to relaxed constraint. We then performed a genome-wide screen and identified hundreds of new candidate eye-specific cis-regulatory sequences, which preferentially clustered near confirmed eye development genes. Thus, the results of convergent, regressive evolution provide a powerful means to assign functions to uncharacterized elements in the genome. Upon further examination, mole-accelerated regulatory sequences preferentially lost sequence motifs recognized by transcription factors active in late development and adult tissues, while motifs involved in early development remain relatively conserved. This pattern supports the hypothesis that initial stages of eye development are important cues for proper development of neighboring tissues, and hence early eye stages remain conserved, despite devastating regression at later stages in subterranean mammals. We conclude that eye-related genes and regulatory elements show convergent patterns of loss and retention during repeated instances of regressive evolution, and that regression occurs differentially across tissues, physiological processes and developmental stages. Broadly, this strategy of studying phenotypic convergence in a comparative genomic context is emerging as a powerful approach to characterize functional elements underlying evolutionarily important phenotypes.

Lightning Talks 2 Pleiotropic genetic effects percolate through underlying networks of traits. *K.A. Geiler-Samerotte*[1,2], Annalise Paaby[3], Austin Taylor[2], Shuang Li[2], Charalampos Lazaris[2], Chelsea Ramjeawan[2], Naomi Ziv[2], Mark Siegal[2] 1) Stanford University, Palo Alto, CA; 2) New York University, New York, NY; 3) Georgia Institute of Technology, Atlanta, Georgia.

Understanding the mapping from genotype to phenotype is a major goal of biology. When one gene contributes to multiple traits – a phenomenon called pleiotropy – this mapping becomes more complex. We recently identified ~50 genes contributing to single cell morphology in a model eukaryote (budding yeast) and found that the majority of these genes each contribute to more than one morphological trait; some contribute to as many as 70 single cell features! To understand the mechanism by which pleiotropic genes influence so many traits, we study thousands of clonal cells from within each yeast strain. We find that the relationships between morphological features are multifaceted. For example, morphological features are related through geometric constraints (*e.g.* nuclear density decreases as nucleus area increases) as well as through cell division (*e.g.* nuclear area and density both increase during mitosis). These multifaceted relationships can obscure one and other, making it seem like pleiotropic genes influence independent traits when in fact the traits in question are related in many ways. We show that pleiotropy in this system does not often result from each gene having multiple independent functions. Instead, pleiotropic genetic effects percolate through underlying networks of traits. This conclusion is promising in terms of mapping genotype to phenotype. It suggests that a more comprehensive understanding of cell and organismal biology can lead to predictions about when a genetic change will have pleiotropic effects.

Lightning Talks 2 Variance QTLs as a means to identify epistatic interactions in genome-wide association studies. *A.R. Marderstein*[1,2], E.R. Davenport[3], A.G. Clark[2,3] 1) Tri-Institutional Program in Computational Biology & Medicine, Weill Cornell Medicine, New York, NY; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.

Many human genome-wide association studies have focused on identifying the particular genetic loci that impact the phenotype. Although model organism studies have often been successful in identifying epistatic (GxG) and genotype-by-environment (GxE) interactions, few studies in human genetics have had adequate power. To address this gap, we screened for genetic associations with the variance of a phenotype (vQTLs), which can not only reveal direct genetic control over the variance of a trait, but also a potential mean-based gene x gene (GxG) interaction or genotype x environment (GxE) interaction that underlies the variance association. Since exhaustive all-pairwise interaction testing is a computationally intensive task and decreases statistical power via multiple hypothesis testing correction, screening for vQTLs can provide a powerful inroad to discovering genetic interactions by generating a subset of loci that serve as promising candidates for an interaction, whether GxG or GxE. Using simulations, we found that two variants affecting the mean of a phenotype through

an interaction are also highly likely to be associated with the phenotypic variance. We then adopted a sequential approach to identify epistatic interactions impacting gut microbiome composition in UK Twins, a cohort of over two thousand twins. First, genome-wide association studies with the variance of bacterial relative abundances were performed across 935 taxa. 1.3 million SNPs were tested using a two-step squared-residual linear mixed model approach. Second, to discover novel interactions with the phenotypic means, we tested significant vQTL loci in mean-based interaction models with all other variants. In this dataset, we discover 885 genome-wide significant interactions ($P < 5 \times 10^{-8}$) across the 935 microbial phenotypes. We identify a host genetic interaction between *SLC2A13,* which has been significantly associated with Crohn's disease and Parkinson's, and long intergenic non-coding RNA *LINC00877* impacting *P. copri* abundance, a taxon previously correlated with autoimmune disease and autism. In summary, vQTLs provide a practical approach to perform genome-wide, population-based analysis of epistatic interactions using human genome-wide association studies.

Lightning Talks 2 Combining population genomics and fitness QTL to identify the genetics of local adaptation in *Arabidopsis thaliana*. *N. Price*[1], B. Moyers[1], J. Lasky[2], J. Monroe[1], J. Mullen[1], L. Lopez[2], G. Oakley[3], J. Lin[1], J. Ågren[4], D. Schrider[5], A. Kern[5], J. McKay[1] 1) Bioagricultural Sciences & Pest Management, Colorado State University, Fort Collins, CO; 2) Department of Biology, Pennsylvania State University, University Park, PA 16802, USA; 3) Department of Botany and Plant Pathology and Center for Plant Biology, Purdue University, West Lafayette, IN 47907, USA; 4) Department of Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden; 5) Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA.

Evidence for adaptation to different climates in the model species *Arabidopsis thaliana* is seen in reciprocal transplant experiments, but the genetic basis of this adaptation remains poorly understood. A major open question is whether local adaptation results from 1) genetic tradeoffs (GT), where alleles that maximize fitness in the home environment are deleterious in alternative environments or 2) conditionally neutral (CN) alleles that are advantageous in the home environment but neutral in alternative environments. Addressing the above, will help enhance our understanding on how temporally or spatially varying selection maintains genetic variation, which population genetic signatures can be used to identify local adaptation in the genome, and finally the biological process underlying local adaptation.
Field-based quantitative trait loci (QTL) studies provide direct but low-resolution evidence for the genetic basis of local adaptation. Using high-resolution population genomic approaches that included the identification of: (1) recent sweeps; (2) SNPs showing significant allele frequency divergence (*Fst*); and (3) SNPs showing significant correlations to climate, we examine local adaptation along previously identified genetic tradeoff (GT) and conditionally neutral (CN) QTL for fitness between locally adapted Italian and Swedish *A. thaliana* populations (Ågren et al. *PNAS*, 2013). Using permutation tests, we examine whether GT or CN QTL peaks are found significantly close to genomic regions enriched in the aforementioned population genomic signals of local adaptation. We find that genomic regions enriched in high *Fst* SNPs are found significantly close to GT QTL peaks while peaks of CN QTL are close (not significantly) to regions enriched with SNPs showing correlations to climate in Eurasia. Among the high *Fst* regions identified, some show significant correlations to climate in Eurasia and evidence of recent sweeps in Sweden. Examining unfolded site frequency spectra across genes containing high F_{ST} SNPs suggests genetic tradeoffs may be due to more recent adaptation in Sweden than Italy. Finally, we collapse a list of thousands of genes spanning GT QTL to 42 genes that likely underlie the observed genetic tradeoffs and explore potential biological processes driving these tradeoffs.

Keynote Session 4: Session Chairs The Aging Proteome: Is Aging Programmed in our Genes? *Gary Churchill*, Yuka Takemon, Ron Korstanje Jackson Lab, Bar Harbor, ME.

Cellular protein composition changes with age. These changes could reflect a transcriptional program that is driving the aging process. Alternatively, the effects of age on the proteome may due to post-transcriptional mechanisms of protein homeostasis. Here we examine the role of transcriptional regulation in shaping the aging proteome through transcriptional and proteomic profiling of kidney in a genetically diverse mouse model. We apply mediation analysis to determine the extent to which age-related changes in the proteome are determined by changes in RNA abundance and find that, although variation in transcript levels drives variation in the majority of proteins, the effects of age on protein abundance are largely not mediated by corresponding changes in transcription. In contrast to the transcription-independent effects of age on the proteome, we show that effects of sex on protein abundance are driven primarily by transcriptional regulation. We identify distinct age-related changes in proximal tubules and glomeruli that reflect cell type-specific damage and repair processes. Our findings provide support for the hypothesis that age-related changes in post-mitotic mammalian tissues are not determined by a genetic program.

Keynote Session 4: Session Chairs Genetics and evolution of hybrid lethality between sympatric species of *Mimulus*. M. P. Zuellig, *A. L. Sweigart* Department of Genetics, University of Georgia, Athens, GA.

As a common cause of reproductive isolation in diverse taxa, hybrid incompatibilities are fundamentally important to speciation. Recent work has provided insight into the molecular functions of genes involved in hybrid dysfunction, but we still know little about how such genes initially evolve within species. Here we describe a simple genetic incompatibility that causes

lethality in hybrids between two closely related species of yellow monkeyflower (*Mimulus guttatus* and *M. nasutus*). This hybrid incompatibility, which causes a fraction of F_2 hybrid seedlings to lack chlorophyll and die shortly after germination, occurs between sympatric populations that are connected by ongoing interspecific gene flow. Using genetic mapping and gene expression analyses, we show that lethality occurs in hybrids that lack a functional copy of the critical photosynthetic gene *pTAC14*. In *M. guttatus*, this gene was duplicated, but the ancestral copy is no longer expressed, whereas in *M. nasutus*, the duplication is missing altogether. As a result, hybrids die when they are homozygous for the nonfunctional *M. guttatus* copy and missing the duplicate from *M. nasutus*. In a series of genetic crosses between plants collected from throughout the species' ranges, we have also discovered that both hybrid lethality alleles are widely distributed and rarely fixed within populations. Polymorphism at one of the loci appears to be driven by introgression between the two *Mimulus* species. Although our study suggests that neutral evolutionary processes may play an important role in the evolution of hybrid incompatibilities, it also raises additional questions about how such alleles are maintained in nature.

Keynote Session 4: Session Chairs Heterochromatin and adaptive evolution. *A.G. Clark*, Jullien Flynn, Emily Brown, Sarah Lower, Michael McGurk, Daniel Barbash Molecular Biology & Genetics, Cornell University, Ithaca, NY.

Densely compacted, gene-poor heterochromatic regions of genomes have received relatively scant attention by evolutionary genomicists, in part because of the challenge of assembling these regions of the genome. Using a simple kmer-based approach on raw sequence reads, we have been able to draw inferences about many aspects of evolutionary turnover of heterochromatic sequences. This includes quantitative assessment of mutation rates (both copy number and sequence) based on mutation-accumulation experiments in Daphnia, Chlamydomonas, flies and mice. PacBio and other long-read technologies are also yielding insights about heterochromatin evolution, especially at euchromatin-heterochromatin boundaries. Several studies have shown that the total amount of heterochromatin in a genome impacts chromatin state genome-wide, and we have shown that the composition of the heterochromatin also has target-specific impacts on gene expression. These trans-acting impacts of heterochromatin can potentially have considerable adaptive consequences for the species, imposing relatively strong selection on this component of the genome formerly dismissed as junk. A comprehensive model of the way that mutational processes are balanced by selection acting on functional attributes of heterochromatin still eludes us, but progress in identification of key components of such a model has been accelerating.

Keynote Session 4: Session Chairs The role of parallel genetic changes for plant mating system shifts - on the genetic basis of a recent loss of self-incompatibility in *Capsella orientalis*. J.A. Bachmann, A. Tedder, B. Laenen, M. Fracassetti, A. Désamoré, *T. Slotte* Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, SE.

Flowering plants have many intricate mechanisms to promote outcrossing. A common mechanism is self-incompatibility (SI), which allows plants to recognize and reject self pollen through the action of male and female specificity components, encoded at the *S*-locus. Despite the benefits of outcrossing, SI has been repeatedly lost in many lineages of flowering plants. While theory predicts that loss-of-function mutations that affect male specificity should spread more easily than those that affect female specificity, the role of parallel molecular changes in the recurrent breakdown of SI remains unclear. The crucifer genus *Capsella* offers an excellent opportunity to study multiple transitions from outcrossing to selfing, but so far, little is known about the transition to selfing in the diploid *Capsella orientalis*. Here, we combine long-read sequencing, genetic mapping and analyses of genomic and expression variation to investigate the genetic basis and timing of the loss of SI in *C. orientalis*. We show that loss of SI is due to genetic changes at the canonical Brassicaceae self-incompatibility locus (*S*-locus). We further identify a frameshift deletion in the male specificity gene *SCR*, which is predicted to lead to loss of male specificity, and we confirm the loss of male SI specificity by interspecific crosses. Finally, we leverage full-length *S*-locus sequence information to estimate the timing of loss of SI in *C. orientalis.* Our results mirror recent findings in *Arabidopsis* which suggest that mutations that disrupt male SI specificity are more likely to contribute to the loss of SI in wild species, and our estimates of the timing of loss of SI are important for interpretation of genomic data on the evolutionary consequences of selfing in *Capsella*.

Keynote Session 4: Session Chairs Distal expression QTLs: statistical challenges and opportunities. *B.E. Engelhardt*[1], Brian Jo[1], Ashis Saha[2], Alexis Battle[2], GTEx Consortium 1) Princeton University, Princeton, NJ; 2) Johns Hopkins University, Baltimore, MD.

Understanding the genetics of gene regulation provides information on the cellular mechanisms through which genetic variation influences complex traits. Expression quantitative trait loci, or eQTLs, are enriched for polymorphisms that have been found to be associated with disease risk. While most analyses of human data have focused on regulation of expression by nearby variants (cis-eQTLs), distal or trans-eQTLs may have broader effects on the transcriptome and important phenotypic consequences, necessitating a comprehensive study of the effects of genetic variants on distal gene transcription levels. I will discuss the identification of trans-eQTLs in the Genotype Tissue Expression (GTEx) project data, consisting of 449 individuals with RNA-sequencing data across 44 tissue types. Trans-eQTLs are challenging to identify because of statistical reasons -- there are trillions of statistical test to account for -- and biological reasons -- broad-effect trans-eQTLs require

substantial population-wide variation in the genes that regulate transcription for many other genes. We have developed two methodologies to address these challenges, in the areas of hypothesis test correction and identifying broad effect expression QTLs, and describe the results and implications of these approaches applied to the GTEx data.

Keynote Session 4: Session Chairs Phenotypic Plasticity in Recombination Frequency. *N. Singh* Biology, University of Oregon, Eugene, OR.

Phenotypic plasticity is pervasive in nature. However, the genetic and molecular mechanisms mediating phenotypic plasticity are largely unknown. For instance, it remains controversial whether there are independent 'plasticity' genes or whether plasticity in a trait is governed by the same genes that underlie population-level variation in that trait. Here we couple classical genetics and an experimental evolution framework to explore the genetics and evolution of recombination and plastic recombination in *D. melanogaster*. Our data support plastic recombination associated with both temperature and Wolbachia infection. We find significant divergence in baseline recombination between different temperature-based experimental evolution regimes, but no divergence in plastic recombination across the different experimental evolution regimes. These observations suggest that the genetic basis of plastic recombination is independent from the genetic basis of population-level variation in recombination fraction.

1 PacBio genome sequencing reveals rampant structural mutation in the malaria parasite. *Emily R Ebel*[1], Marina McDew-White[2], Timothy JC Anderson[2], Dmitri A Petrov[1] 1) Stanford University, Stanford, CA; 2) Texas Biomedical Research Institute, San Antonio, Texas.

P. falciparum, the deadliest malaria parasite, has arguably the most unusual genome in the world. Its AT-content surpasses 80%; it is highly repetitive, but lacks transposons; and it adapts (to antimalarial drugs and host immunity) primarily through structural changes. However, because of limitations in Illumina sequencing, the landscape of structural variation in *P. falciparum* remains largely unknown. Here, we have addressed this gap by using long PacBio reads to generate high-quality, *de novo* genome assemblies of samples from a mutation accumulation experiment in *P. falciparum*.
Our complete assemblies reveal an unexpectedly dynamic parasite genome, in which simple sequence repeats underlie the vast majority of mutations. Besides the high frequency of small indels in repeat units, simple repeats also appear to facilitate large-scale, potentially adaptive structural changes. These include the fluctuating copy number of a drug resistance gene, as well as the formation of a novel immune evasion gene from fragments of two others. Although single base pair changes make relatively minor contributions to overall diversity, we still find hundreds more than expected from Illumina estimates, almost all clustered in complex repeats with other indel mutations. Together, these results indicate that the AT-rich, highly repetitive *P. falciparum* genome is uniquely prone to major mutations and mutational hotspots, which constantly provide raw material for adaptation. Furthermore, we show that PacBio technology opens a window onto thousands of structural variants—here comprising nearly all genetic variation in *P.* falciparum—which are completely missed by reference-based Illumina approaches.

2 Deep taxon sampling and epigenetic profiling reveal the evolutionary dynamics of nematode orphan genes. N. Prabh, M. Werner, C. Roedelsperger, *R.J. Sommer* Max-Planck-Institute for Developmental Biology, Tuebingen, Germany.

The widespread identification of genes without detectable homology in related taxa is a hallmark of genome sequencing projects in animals. Such genes have been called novel, young, taxon-restricted, or orphans, but little is known about the mechanisms accounting for their origin, age and mode of evolution. Phylogenomic studies relying on deep and systematic taxon sampling and the employment of the comparative method can provide insight into the evolutionary dynamics acting on novel genes. We used a phylogenomic approach for the nematode model organism *Pristionchus pacificus* and sequenced eight *Pristionchus* and two outgroup species. This resulted in 10 genomes with a ladder-like phylogeny, sequenced in one laboratory using the same platform and analyzed by the same bioinformatic procedures. Our analysis revealed that 68-81% of genes are assignable to orthologous gene families, the majority of which defined nine Age classes with presence/absence patterns that can be explained by single evolutionary events. Contrasting different Age classes, we find novel gene families preferentially arise at chromosome arms, are typically lowly expressed and evolve rapidly. Over time they settle down at chromosome centers, increase expression, and become more constrained. However, younger gene families also show a higher propensity to being lost. Similarly, we performed epigenetic profiling in *P. pacificus* and analyzed orthologous, paralogous and novel genes using CHIP of seven histone modifications, ATAC seq, Iso-seq and RNA-seq. Consistent with previous findings we find that young genes are on average less expressed than older genes. Surprisingly however, the subset of orphan genes that are expressed exhibit distinct chromatin states from similarly expressed conserved genes. Orphan gene transcription is determined by a lack of repressive histone modifications, and transcriptional start sites that resemble enhancers defined by H3K4me1, H3K27ac and ATAC-seq peaks, in contrast to the strict correlation of conserved genes and promoters bearing H3K4me3 and H3K27ac. We also find that the majority of orphan genes are located on chromosome arms with repressive histone marks, yet expressed orphan genes are more randomly distributed. Our results support a model of new gene origination by rare integration into open chromatin near enhancers.

3 Cytonuclear coordination and evolution in allotetraploid wheat. *Joel Sharbrough*[1], Justin Conover[2], Corrine Grover[2], Jonathan Wendel[2], Daniel Sloan[1] 1) Department of Biology, Colorado State University, Fort Collins, CO; 2) Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA.

Whole-genome duplications (WGDs), in which the number of nuclear genome copies is elevated as a result of autopolyploidy or allopolyploidy, underlie many of the major episodes of diversification in eukaryotes. While enormous progress has been made in understanding the myriad genomic and transcriptomic consequences of polyploidy, the genetic and evolutionary forces that WGD imposes upon cytoplasmic genomes are less well understood, despite the central role that cytonuclear interactions play in eukaryotic function and fitness. In particular, cellular respiration and photosynthesis depend upon successful interaction between the 3000+ nuclear-encoded proteins destined for the mitochondria or plastids and the gene products of cytoplasmic genomes in multi-subunit complexes such as Rubisco, the enzymes that comprise the mitochondrial electron transport chain, and organellar ribosomes. Allopolyploids are thus faced with the critically important task of maintaining successful interactions and coordinated gene expression between nuclear and cytoplasmic genes that were inherited from different species. Because maternal homoeologs are expected to be more closely "matched" to cytoplasmic genomes than paternal homoeologs, incompatibilities between the organelle genomes and paternal subgenomes of allopolyploids may lead to relaxed selection on paternal vs. maternal homoeologs of genes targeted to the mitochondria or plastids. To test this hypothesis, we compared rates of molecular evolution in maternal vs. paternal homoeologs of organelle-targeted genes in the allotetraploid *Triticum turgidum* (pasta wheat). Together, our results provide the first assessment of cytonuclear coevolution in diploid vs. polyploid wheat, and suggest that cytonuclear incompatibilities likely played a role in the formation and evolution of wheat.

4 Hostile genomic takeover by transposable elements in the Strawberry poison frog. *R.L. Rogers*[1], L Zhou[3], C Chu[5], R Marquez[9], A Corl[2], T Linderoth[2], L Freeborn[7], M McManes[8], Z Xiong[3], J Zheng[3], X Xun[3], M. Kronforst[9], K Summers[6], Y Wu[10], CL Richards-Zawacki[7], G Zhang[4], R Nielsen[2] 1) Bioinformatics and Genomics, UNC Charlotte, Charlotte, NC; 2) Integrative Biology, University of California Berkeley CA; 3) Beijing Genomics Institute, China National Genbank, Shenzhen, China; 4) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 5) Department of Biomedical Informatics, Harvard University, Cambridge MA; 6) Department of Biology, Eastern Carolina University, Greenville, NC; 7) Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA; 8) College of Life Sciences and Agriculture, University of New Hampshire, Durham, NH; 9) Department of Ecology & Evolution University of Chicago, Chicago, IL; 10) Computer Science and Engineering Department, University of Connecticut, Storrs, CT.

We sequenced the genome of the strawberry poison frog, *Oophaga pumilio,* at a depth of 127.5X using variable insert size libraries. The total genome size is estimated to be 6.76 Gb, of which 4.76 Gb are from low differentiated, high copy number transposable element families encompassing DNA transposons, RNA transposons, and LTR retrotransposons, including at least 0.4 and 1.0 Gb of *Mariner/Tc1* and *Gypsy* elements, respectively. Expression data indicate high levels of *gypsy* and *Mariner/Tc1* expression in ova of *O. pumilio* compared to *Xenopus laevis*. We further observe phylogenetic evidence for horizontal transfer (HT) of *Mariner* elements, possibly between fish and frogs. The elements affected by HT are present in high copy number and are highly expressed, suggesting ongoing proliferation after HT. Our results suggest that the large amphibian genome sizes, at least partially, can be explained by a process of repeated invasion of new transposable elements that are not yet suppressed in the germline. We also find changes in the spliceosome that we hypothesize are related to permissiveness of *O. pumilio* to increases in intron length due to transposon proliferation. Finally, we identify the complement of ion channels in the first genomic sequenced poison frog and discuss its relation to the evolution of auto-resistance to toxins sequestered in the skin.

5 The X chromosome of the German cockroach, *Blattella germanica*, is homologous to a fly X chromosome despite 400 million years divergence. *Richard Meisel*[1], Judith Wexler[2,3] 1) Department of Biology and Biochemistry, University of Houston, TX; 2) Center for Population Biology, Department of Evolution and Ecology, University of California, Davis; 3) Department of Entomology, University of Maryland.

The chromosomes that are sex-linked can differ between closely related species. In contrast, other sex chromosomes have been conserved for >100 million years. Cases of long-term sex chromosome conservation are informative of factors that prevent sex chromosome turnover. We used a comparative analysis of male and female genomes to discover that many of the same genes are found on the German cockroach, *Blattella germanica*, X chromosome and the ancestral X chromosome of higher flies. We also show that three *trans* regulators of transcription and chromatin on the fly X chromosome are conserved in the cockroach genome. We hypothesize that the common ancestor of cockroaches and flies had an X chromosome that resembled the extant cockroach/fly X. Cockroaches and flies diverged ~400 million years ago, and we believe that this is the longest documented conservation of a sex chromosome. Cockroaches and most flies have different mechanisms of sex determination, suggesting that the X chromosome was conserved despite evolution of the sex determination pathway.

6 piRNA-mediated silencing of an invading TE evolves rapidly through abundant beneficial *de novo* mutations. *S. Zhang*, E. Kelleher Department of Biology and Biochemistry, University of Houston, Houston, TX.

Transposable elements (TEs) are selfish genetic entities whose mobilization can reduce host fitness by producing deleterious mutations and inciting genome instability. In many metazoans, TE activity is regulated in the germline by small Piwi-interacting RNAs (piRNAs), which are derived from specialized genomic loci known as piRNA clusters. Two non-mutually exclusive hypotheses are proposed to explain how piRNA silencing evolves after a genome is invaded by a new TE. First, host repression may evolve through *de novo* mutation, with random transpositions into piRNA clusters producing repressor alleles. Alternatively, repression may also evolve through epigenetic mutation, if a non-piRNA producing insertion of the invading TE is converted into a novel piRNA cluster.

Drosophila P-element DNA transposons provide a unique opportunity to disentangle the contributions of *de novo* and epigenetic mutation to the evolution of piRNA-mediated silencing. *P*-elements invaded *D. melanogaster* genomes around 1950, and in response, many natural populations evolved piRNA-mediated repression in less than 50 years. However, numerous strains collected prior to 1950 are retained in laboratories and stock centers, creating a historical record of ancestral piRNA clusters that were active before the *P*-element invasion. We used published small RNA libraries from 8 of these historic collections to annotate a set of ancestral piRNA clusters. We then developed and applied a new method for identifying TE insertions in repeat-rich heterochromatin, where piRNA clusters reside, to 205 recently collected wild-derived genomes comprising the *Drosophila* Genetic Reference Panel (DGRP).

We discovered that ~95% of DGRP genomes contain at least one *P*-element insertion in an ancestral piRNA cluster, supporting *de novo* mutation as the predominant mechanism producing *P*-element repressor alleles. In addition, our analyses uncovered no fewer than 122 unique *P*-element insertions into piRNA clusters, indicating that the ubiquitous repressive phenotype is underpinned by unprecedented genotypic diversity. Finally, we found that *P*-element insertions in ancestral piRNA clusters segregate at a significant higher frequency than those outside of piRNA clusters, suggesting that they confer a selective advantage. Taken together, our results suggest that piRNA mediated silencing evolves through a novel, soft-sweep like process, in which a high *de novo* mutation rate produces an abundance of adaptive alleles at different genomic sites.

7 A mechanistic model of assortative mating in a hybrid population. *Amy Goldberg*[1,2], Noah Rosenberg[2] 1) Integrative Biology, UC Berkeley, Berkeley, CA; 2) Biology, Stanford University, Stanford, CA.

Mating pairs often exhibit levels of similarity in phenotypic or genotypic traits that differ systematically from the level expected under random mating, produced by assortative mating. For example, in admixed human populations, spouses possess correlated ancestry components suggestive of positive assortative mating on the basis of ancestry. Additionally, assortative mating has been proposed as a mechanism for hybrid and sympatric speciation. Using a two-sex mechanistic admixture model, we devise a model of preferential mating based on source population during hybridization. Under the model, we study the distribution of genetic ancestry on the autosomes and X chromosome for positive and negative assortative assortative mating, allowing migration to the hybrid population to vary between sexes and over time. We demonstrate that, whereas the mean admixture under assortative mating is equivalent to that of a randomly mating population, the variance of admixture generally increases with higher levels of positive assortative mating and decreases with negative assortative mating, analogous to classic theory on assortative mating by single locus genotypes or traits. However, perhaps contrary to previous work, we identify cases in which positive assortment can decrease the variance because mate preferences co-occur in multiple populations—the parental and hybrid populations. The effect of assortative mating is smaller on the X chromosomes than the autosomes because inheritance of the X in males depends only on the mother's ancestry, not on the mating pair. As the variance of admixture has been used to infer the timing of hybridization and sex-biased admixture, we consider the implications of assortative mating for inferring population history and speciation. Our model provides a framework to quantitatively study assortative mating under flexible scenarios of mating and hybridization over time.

8 Detecting polygenic adaptation in maize. *E. Josephs*[1,2], J. Berg[3], J. Ross-Ibarra[2,4], G. Coop[1,2] 1) Department of Evolution and Ecology, University of California, Davis, Davis, CA; 2) Center for Population Biology, University of California, Davis, Davis, CA; 3) Department of Biological Sciences, Columbia University, New York, NY 10027, USA; 4) Department of Plant Sciences, University of California, Davis, Davis, CA.

Characterizing the genetic basis of adaptation is not only a longstanding goal of evolutionary biology, but is also an important component of understanding adaptation. Adaptation in quantitative traits likely often occurs through subtle shifts in allele frequencies at many loci, a process called polygenic adaptation. Even though many traits have a polygenic basis, conventional methods lack power for detecting polygenic adaptation. In this talk, I describe strategies for detecting polygenic adaptation at the phenotypic and genotypic level. I show that we can leverage trait-associated loci identified from genome-wide association studies to detect the coordinated shifts in allele frequency expected under polygenic adaptation. Application of my methods to different maize populations shows evidence for polygenic adaptation in a number of traits in both inbred lines from the USDA germplasm pool and European landraces. Ultimately, these methods can be applied to multiple domesticated and wild species to give us a

broader picture of the the specific traits that contribute to adaptation and the overall importance of polygenic adaptation in shaping trait variation.

9 Population genetic models for highly polygenic disease. *Jeremy Berg*, Guy Sella Biological Sciences, Columbia University, New York, NY.

A decade into the era of well powered and reproducible genome wide association studies, one thing is clear: many complex diseases are extremely polygenic, with thousands or perhaps tens of thousands of segregating variants contributing to variation in risk among individuals. However, our understanding of the reasons for variation among diseases in their prevalence, as well as in the number, frequencies, and effect sizes of mutations which contribute to variance in risk, is limited.

We construct and analyze a model of a highly polygenic complex disease at evolutionary equilibrium under mutation-selection-drift balance. In our model, disease arises due to a global epistasis among mutations which act additively on the liability scale (i.e. a liability threshold model). Selection occurs at the level of the disease phenotype, while selection coefficients experienced by individual loci arise as dynamic variables of the system. We show that in fact, so long as the disease is sufficiently polygenic, the selection coefficients of individual loci are insensitive to the fitness cost of the disease, and instead depend on the distribution of effect sizes and the degree of mutational bias toward increased disease liability. This result is robust to the assumption of a strict liability threshold, and also holds in the presence of some forms of pleiotropy. We also show that the results of genome wide association studies appear to be qualitatively inconsistent with mutation-selection-drift equilibrium, assuming modern prevalence and fitness cost of disease, but that pleiotropy and/or very recent environmental change can potentially explain these inconsistencies.

10 Evolutionary dynamics of influenza across spatiotemporal scales. *K.S. Xue*[1,2], T. Stevens-Ayers[3], A.P. Campbell[3], J.A. Englund[4,5], S.A. Pergam[3,6,7], M. Boeckh[3,6,7], J.D. Bloom[1,2] 1) Genome Sciences, University of Washington, Seattle, WA; 2) Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 4) Seattle Children's Research Institute, Seattle, WA; 5) Department of Pediatrics, University of Washington, Seattle, WA; 6) Department of Medicine, University of Washington, Seattle, WA; 7) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

The rapid global evolution of influenza viruses begins with *de novo* mutations that arise in individual infected hosts. Recent advances in high-throughput deep sequencing have made it increasingly possible to measure influenza's within-host genetic diversity and compare this diversity to the virus's global evolution. We demonstrate that influenza evolution within infected humans recapitulates many evolutionary dynamics observed at the global scale. We deep-sequence longitudinal samples from four immunocompromised patients with long-term H3N2 influenza infections. We find parallel evolution across three scales: within individual patients, in different patients in our study, and in the global influenza population. In the viral surface protein hemagglutinin, a small set of mutations arises independently in multiple patients. These same mutations emerge repeatedly within single patients and compete with one another, providing a vivid clinical example of clonal interference. Many of these recurrent within-host mutations also reach a high global frequency in the decade following the patient infections. These results demonstrate that influenza viruses can evolve rapidly in chronic infections in ways that mirror global viral evolution. We close by discussing major open questions about how genetic drift, purifying selection, and positive selection combine to shape influenza's evolution within hosts during more typical, acute infections, and how this within-host diversity contributes to the virus's global evolution. Altogether, our analyses illuminate how evolutionary forces act on viral populations across interlocking scales of space and time.

11 Intra-patient evolutionary dynamics of HIV drug resistance evolution in time and space. *A.F. Feder*[1], Z. Ambrose[2], R.W. Shafer[1], S-Y. Rhee[1], S. Holmes[1], J. Hermisson[3], P.S. Pennings[4], D.A. Petrov[1] 1) Stanford University, Stanford, CA; 2) University of Pittsburgh, Pittsburgh, PA; 3) University of Vienna, Vienna, Austria; 4) San Francisco State University, San Francisco, CA.

At the beginning of the HIV epidemic, drug resistance to treatment evolved quickly and predictably across all patients. Now we treat HIV with combination therapies of three drugs so any single HIV mutation is insufficient for viral replication. The rate of drug resistance evolution has plummeted in response. Despite these advances, a minority of viral populations become resistant nonetheless. Why and how do certain populations overcome efficacious combination therapy? In answer to the first question, we analyze how the mode of drug resistance evolution changed within patients throughout the epidemic using historical HIV sequences. We find evolution has shifted from multiple origins of drug resistance ("soft sweeps") to single origins ("hard sweeps") as treatments have improved. This observation suggests that while drug resistance was once inevitable, now patients that fail combination therapy are merely unlucky, and not predestined to fail due to factors like poor adherence. However, questions remain about **how** hard sweeps of drug resistance can occur at all under combination therapy. Theory suggests that spatial structure of the intra-patient population drive multidrug resistance through creating pockets of spatial monotherapy whereby mutations can be acquired sequentially instead of simultaneously (Moreno-Gamez et al, 2015). However within-body population structure remains unknown. To bridge this gap, I analyze Simian-HIV-infected

macaques sampled spatially and temporally during drug resistance evolution. We observe that populations from different organs (gut, plasma, lymph node, vagina) within the same macaque can be significantly different although the magnitude of the difference varies through time as drug resistance emerges. From these data, we quantify the population genetic parameters of the intra-patient environment to aid modeling efforts such as the spatial-monotherapy work. Notably, we develop a new ABC framework for estimation in adapting populations that can estimate migration rates much larger than those possible to estimate from tracking neutral alleles. This represents the first quantitative description (to our knowledge) of the within-patient spatial structure of HIV that accounts for adaptation.

12 Genetic variation at a conserved non-coding element contributes to microhabitat-associated behavioral differentiation in Malawi African cichlid fishes. *E.C. Moore*, R.B. Roberts WM Keck Center for Behavioral Biology, Biological Sciences, North Carolina State University, Raleigh, NC.

Successful behavioral adaptation to habitat is fundamentally important to species fitness, but linking such behaviors to genes has proven difficult due to plasticity of phenotypes and rarity of within-species genetic variation underlying adaptive behavioral patterns. The East African cichlid fishes of Lake Malawi are ideal for investigating behavioral adaptation to environment, as within genera, fine-scale niche partitioning has resulted in sympatric sister species that live in definable microhabitats with distinct selection pressures. We tested species of Malawi cichlids found in rocky reefs, open sand habitats, or the sand-rock interface for a variety of environment-usage phenotypes in a controlled laboratory setting. Computer-aided analysis of fish response to new environments revealed distinct behavioral patterns among sand, rock and interface species. In the lab, we created a hybrid cross between two genera that differ in microhabitat in the wild and behavior in the lab, and used a ddRAD-seq linkage-mapping strategy to identify quantitative trait loci (QTL) associated with species-specific behaviors. Comparative genomic analysis from 82 wild and wild-derived species grouped by microhabitat-use identified variation corresponding with one of these QTL, further supporting broad association with habitat use across the Malawi cichlid radiation. This locus contains a conserved non-coding element (CNE) upstream of three neuronal cell adhesion molecule (NCAM) genes, where the derived "sand" allele appears to disrupt a neuronal transcription factor binding motif. Additionally, we used interface species that were naturally segregating the "rock" allele and "sand" allele at the NCAM CNE to confirm that genotype at the locus is associated with behavioral variation among full siblings within a species. Finally, allele-specific expression indicates that expression of one of the three NCAM genes has a two-fold reduction in expression when linked to the derived "sand" allele at the CNE. Together, these integrated results suggest that evolution of gene expression at the NCAM locus has accompanied behavioral adaptation to microhabitat, which could ultimately reinforce speciation through spatial isolation.

13 Dissecting complex life-history trait in *Arabidopsis thaliana* using extremely large segregating population. *Wei Yuan*[1,2], Jonathan Flowers[2,3], Dustin Sahraie[2], Ian Ehrenreich[4], Michael Purugganan[2] 1) Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Baden-Wuerttemberg, Germany; 2) Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY, USA; 3) Center for Genomics and Systems Biology, NYU Abu Dhabi Research Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates; 4) Molecular and Computational Biology Section, University of Southern California, Los Angeles, CA, USA.

Understanding complex life history traits, their expression and genetic architecture under various environments are of key biological interest. While theories predict that life history trait variation are driven by numerous small-effect genetic variants; empirical studies often, in contrary, map them to relatively few broad genomic regions (and in few cases genes) of large effects, while leaving a significant fraction of phenotypic variation unexplained. Whether the observed simplicity is representative of complex trait genetics, or rather biased due to low power of extant mapping methods remain open to question. I hypothesized that an extremely large mapping population can overcome the limiting factor of recombination events, and improve both power and precision to pinpoint causal variants of small effects. For my doctoral thesis, I developed the Extreme-QTL mapping platform in *Arabidopsis thaliana*. The method employed high-throughput selection-based phenotyping and bulk-segregant genotyping to enable mapping of an extremely large pool of segregants. With this platform I mapped seed germination speed, an important yet poorly understood trait, and its response to a novel, saline environment. I demonstrated in 17 *A.thaliana* natural accessions that the range of germination speed variation reaches 30% of the trait value (p^5 F_3 recombinants, I identified 3 peak clusters on chromosome 1, 3, and 4 that account for rapid germination. Individual re-sequencing of 191 fast-germinating F_3s suggested that each peak cluster is highly likely driven by multiple causal polymorphisms. A parallel mapping of germination speed under extreme salinity (250 mM NaCl) revealed an increased phenotypic variation due to increased additive genetic variance. Besides a region on the distal arm of chromosome 4 that is required for rapid germination under both tested conditions, many regions that are neutral under non-salt condition exhibited response to salinity. My result revealed that the genetic architecture of germination speed variation is highly complex, determined by many linked and conditional-neutral variants. The method achieved precision (~ ±1 cM) that is expected from bulk-segregant mapping studies in other organisms. However, its inability to dissect complex traits to single-gene level, as was previously sucessful in *S. cerevisiae*, highlighted the linkage between causal variants in the trait studied. By conducting a mapping study using 2 order-of-magnitude greater of recombination events accessible to a typical traditional

QTL mapping design, my research demonstrated the promise of high throughput genotyping to mapping complex traits in multicellular eukaryotes, while also pointed to high-throughput phenotyping as the immediate bottleneck to achieving high mapping precision.

14 GWAS reveals antagonistic pleiotropy, polygenic adaptation and fluctuating selection within a natural population. A Troth[2], J Puzey[2], J Willis[2], *John Kelly*[1] 1) ku, Lawrence, KS; 2) duke u, nc, usa.

Evolutionary biology now endeavors to identify nucleotide level variants underpinning phenotypic variation. To this aim, we whole-genome sequence 187 lines from a single natural population of yellow monkeyflower (*Mimulus guttatus*) and assay allelic effects on morphological and life history traits. The collection of significant variants show remarkably consistent features. Alleles that delay development, and increase plant size at reproduction, are nearly always less frequent than their "small/fast" alternatives. We find that "big/slow" alleles are more abundant in other populations of *M. guttatus* with greater average plant size, indicating a polygenic basis to local adaptation. Three years of field study directly identify significant effects of the polymorphisms on survival and reproduction. Environmental fluctuation are implicated as a means to preserve variation, with big/slow alleles conferring a fecundity advantage under favorable conditions. The synthesis of genomic, phenotypic, and fitness data connects evolutionary processes responsible for both geographical variation and the persistence of high genetic variation within local populations of *M. guttatus*.

15 Patterns of genetic variation at the *FLO11* locus suggest a form of kin recognition in the yeast, *Saccharomyces cerevisiae*. *H.A. Murphy*, Z.J. Oppler, M.E. Parrish Biology, William and Mary, Williamsburg, VA.

Cooperative behaviors have long fascinated and puzzled evolutionary biologists. It is now clear that most examples of apparent altruism can be explained by kin selection, with kin recognition common in many systems. In clonally growing microbes, cooperative behaviors involving cells adhering to one another generally rely on "kind" recognition. A single locus or trait, referred to as a greenbeard, is enough to signal cooperation regardless of overall relatedness across the genome. However, in behaviors that require motile microbes to locate one another, rare examples of kin- and self-recognition have been reported; in these cases, membrane-associated proteins with variable extracellular domains confer discrimination. In the non-motile budding yeast, *Saccharomyces cerevisiae*, Flo11 is a membrane-anchored adhesin required for most social phenotypes (i.e., biofilms, mats, pseudohyphal growth), and has been hypothesized to play a role in recognition. We amplified and sequenced the regulatory and coding regions of *FLO11* in 78 environmental isolates that vary in their social phenotypes, and generated *de novo* assemblies of the locus. The analysis resolved the unique sequence in the A and C domains of the gene, as well as the upstream and downstream regions, but not the repetitive B domain. Population genetic analyses suggest that the precise regions implicated in cell-cell adhesion exhibit a signature of positive selection, while the rest of the gene is under purifying selection. Phenotypic assays demonstrate that different natural *FLO11* alleles generate diverse biofilm architectures in an otherwise constant genetic background. Our results suggest a "shades of greenbeard" system: the ability of Flo11 to adhere to like kinds, but combined with a preference for self. Unlike in motile microbes where cheater avoidance is likely driving the evolution of recognition, homophillic binding of Flo11 may be selected during competition among clones. Thus, the interplay between inter-clone competition and intra-clone cooperation in spatially structured microbial communities can potentially lead to recognition systems.

16 Genetic dissection of chromatin accessibility and transcript abundance underlying ground state pluripotency in mouse embryonic stem cells. D.A. Skelly[1], C. Byers[1], A. Czechanski[1], C. Spruce[1], S. Aydin[1], A. Stanton[1], T. Choi[2], G.A. Churchill[1], *S.C. Munger*[1], C.L. Baker[1], L.G. Reinholdt[1], The Cellular Systems Genetics Consortium 1) The Jackson Laboratory, Bar Harbor, ME; 2) Predictive Biology, Inc., Carlsbad, CA.

Many adult-onset diseases are caused by genetic variants that have proximal effects at the earliest stage of development – the pluripotent, ground state. In order to better understand the role of genetics in maintenance of pluripotency, we performed genetic and genomic analysis on a panel of undifferentiated embryonic stem cell lines derived from genetically heterogeneous Diversity Outbred mice (DO mESCs). We profiled chromatin accessibility (ATAC-seq) and transcript abundance (RNA-seq) of each DO mESC line maintained in cell culture conditions that promote the pluripotent state. We mapped thousands of loci with genetic variants that alter chromatin accessibility (caQTL) and transcript abundance (eQTL). Many distant QTL co-localize and appear as prominent trans-bands, suggesting that a common regulator may drive them. One locus on chromosome 15 altered the expression of 208 genes including many with known functions in maintenance of pluripotency. We applied mediation analysis and identified *Lifr* (leukemia inhibitory factor receptor) transcript abundance as the causal intermediate for these eQTL. Interestingly, sex-specific differences in many of these genes suggest a nonlinear response to *Lifr* dosage. Joint mediation analysis of eQTL by chromatin accessibility revealed a variable region of open chromatin upstream of the *Lifr* gene containing a single SNP that predicts the allelic effects on *Lifr* expression and its downstream targets. This suggests a causal chain of molecular events starting from a single SNP that modulates chromatin state in a *Lifr* enhancer that in turn affects *Lifr* transcription, which ultimately regulates transcript abundance of 208 target genes, including both known and novel pluripotency-associated genes. To validate these predictions and demonstrate their

effects on ground state pluripotency and early lineage commitment, we have developed complimentary genetic resources including F1 hybrid mESC lines from Collaborative Cross strains and CRISPR-modified inbred, founder mESC lines. Future studies will integrate additional cellular assays within this renewable systems genetic resource to expand our understanding of genetic influence on differentiated cell types and ultimately to fetal and adult *in vivo* phenotypes.

17 Modeling ancestry-dependent phenotypic variance accurately corrects for population structure and detects variance effects. *S. Musharoff*[1], D. Park[2], J. Galanter[3], X. Liu[4], S. Forsberg[5], S. Huntsman[6], C. Eng[1], J.F. Ayroles[5], E.G. Burchard[1,6], N. Zaitlen[1] 1) Lung Biology, University of California, San Francisco, San Francisco, CA; 2) Encompass Bioscience Inc; 3) Genentech, South San Francisco, CA; 4) Department of Human Genetics, University of Chicago; 5) Ecology and Evolutionary Biology and Lewis-Sigler for Integrative Genomics, Princeton University, Princeton, NJ; 6) Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

Many complex human phenotypes vary dramatically in their distributions between populations. Most studies to date have focused on the relationship of variation in ancestry and differences in phenotypic mean, either explicitly or to prevent confounding. While analyzing data from the UKBioBank, we discovered that variation in ancestry is also associated with differences in phenotypic variance. We observe a highly significant relationship between population structure (defined by PCA) and phenotypic variance for several physiological and morphological traits including BMI, lung function, hemoglobin amount, and hair color. Interestingly, this pattern follows the north-south latitudinal cline previously reported. The prevailing interpretation is that changes in phenotypic variance ultimately result from genotype-by-genotype interaction (i.e. epistasis) or genotype-by-environment interactions leading to de-canalization. Such variance heterogeneity can produce large differences in disease burden.

We developed an analysis method based on the double generalized linear model (ADGLM), which accounts for relationships between ancestry and phenotypic variance in association studies. We show that the standard approaches of linear regression and linear mixed models, which do not account for these relationships, can produce inflated or deflated test statistics. For example, linear regression applied to simulated data with no genetic effect from a population with ancestry-dependent phenotypic variance produces inflated test statistics (λGC of 1.56), whereas ADGLM produces well-calibrated test statistics (λGC of 1.00). Furthermore, ADGLM has better power than the standard approaches, with power increases as large as 66%. We applied ADGLM to the Genetics of Asthma in Latino Americans (GALA) study. In Mexicans and Puerto Ricans, who have different asthma prevalences, we find a significant association of asthma variance with African (p-value 2.6e-7) and European (p-value 5.6e-6) ancestry proportion. In GALA Puerto Ricans, we detect associations of ancestry with methylation variance at 44 CpG sites (p-values < 1.6e-7), which may be due to gene-by-environment interactions. Consistent with simulations, ADGLM identifies 2% more cis-meQTL associations of genotype with methylation than linear regression. In summary, we show in simulated and real data that ancestry proportion can affect phenotypic variance, linear regression and linear mixed model tests are miscalibrated when ancestry is correlated with phenotypic variance, and ADGLM tests are well-calibrated and more powerful. By focusing on the effect of genetic variation on trait means and ignoring its effect on variance, we are missing an important axis contributing to phenotypic variation and disease emergence. Modeling phenotypic variance with ADGLM will enable discoveries along this axis.

18 A mitochondrial-nuclear interaction compromises immune function and reveals life-history tradeoffs in females. *K.L. Montooth*, J.L. Buchanan, C.D. Meiklejohn School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE.

Physiological responses to short-term environmental stress, such as infection, can have long-term consequences for fitness, particularly if the responses are inappropriate or nutrient resources are limited. Genetic variation in energy metabolism may limit cellular energy availability for these responses and influence resource-allocation tradeoffs even when environmental nutrients are plentiful. Given the central role of the mitochondria in the energy budget of eukaryotic organisms, there is reason to expect that genetic variation for mitochondrial function will contribute to the evolution of these life-history tradeoffs. We used well-characterized Drosophila mitochondrial-nuclear genotypes to test whether disrupted energy metabolism interferes with nutrient-sensing pathways, and whether this disruption has consequences for tradeoffs between immunity and fecundity. We found that flies with a mitochondrial-nuclear incompatibility that compromises mitochondrial function were remarkably resistant to the effect of rapamycin – a drug that stimulates nutrient-sensing pathways in a manner analogous to resource limitation. Resource limitation via diet also compromised survival in these energetically-compromised genotypes, suggesting that this genotype may have little excess energetic capacity and senses fewer cellular nutrients, even when environmental nutrients are not limiting. Accordingly, we found that immune function was compromised in this genotype, but only in females, and that these females experienced immunity-fecundity tradeoffs that were not evident in genotypic controls with normal energy metabolism. Thus, genetic variation in energy metabolism may act to limit the resources available for allocation to life-history traits in ways that generate tradeoffs even when environmental resources are abundant. I will discuss these results in a larger framework of energy demand as an important context for the differential expression of genetic variation in metabolism across life stages, tissues, sexes and environments.

19 The evolution of locally adaptive seasonal camouflage in snowshoe hares. *M.R. Jones*[1], L.S. Mills[2,3,4], P.C. Alves[2,5,6], C.M. Callahan[1], J.M. Alves[5,7], D.J.R. Lafferty[4,8], F.M. Jiggins[7], J.D. Jensen[9,10], J. Melo-Ferreira[5,6], J.M. Good[1] 1) Department of Organismal Biology, Ecology, and Evolution, University of Montana, Missoula, MT; 2) Wildlife Biology Program, University of Montana, Missoula, MT; 3) Office of Research and Creative Scholarship, University of Montana, Missoula, MT; 4) Fisheries, Wildlife, and Conservation Biology Program, North Carolina State University, Raleigh, NC; 5) Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Universidade do Porto, Vairão, Portugal; 6) Departamento de Biologia, Universidade do Porto, Porto, Portugal; 7) Department of Genetics, University of Cambridge, Cambridge, UK; 8) Department of Biology, Northern Michigan University, Marquette, MI; 9) School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 10) School of Life Sciences, Arizona State University, Tempe, AZ.

Adaptation to environmental change is central to the origin and persistence of biodiversity. At least 22 mammal and bird species undergo seasonal color changes as part of a coordinated suite of responses to seasonally varying environments. Yet the genetic underpinnings of dynamic seasonal shifts in pelage coloration, and seasonal plasticity in general, remain largely unknown. Snowshoe hares (*Lepus americanus*) maintain seasonal camouflage by molting to a white winter coat. Direct field estimates of hare survival have shown that mismatch between coat color and snow cover increases predation probability, demonstrating direct climate-induced fitness consequences of camouflage mismatch. Consequently, in regions with low snow cover some hares have evolved brown winter coats. We used association mapping in natural populations and functional experiments in captive animals to show that polymorphic winter camouflage reflects *cis*-regulatory evolution modifying the expression of a single gene during the autumn molt. Causative coat color alleles show population genetic signatures of strong positive selection for the local maintenance of seasonal camouflage. In addition to these substantial molecular insights, phylogenetic analyses and coalescent simulations reveal a crucial role of introgression in the origin of this locally adaptive camouflage polymorphism in snowshoe hares. Ongoing work in this system is addressing the history and genomic consequences of introgression and the precise molecular mechanisms underpinning the development of winter white coats. Despite growing evidence for hybridization between animal species, introgression has only rarely been directly linked to ecological adaptation of traits with known fitness consequences. This research represents the first documentation of hybridization shaping local adaptation of a seasonal trait. Temperate and boreal snow cover duration is predicted to dramatically decrease over the next century, which may further intensify directional selection for winter-brown camouflage. Thus, this dynamic color polymorphism is likely to be a critical component of ongoing adaptation to rapidly changing seasonal environments in this iconic ecological model.

20 Speciation genes are more likely to have discordant gene trees. *R. Wang*, M. Hahn Indiana University, Bloomington, IN.

Speciation genes form the underlying genetic basis for reproductive isolation between species. The genealogies of isolating loci are thought to more faithfully represent species trees because of their direct participation in the process of speciation. This property could provide unique evolutionary insights and help determine the true history of species divergence. Here, we formally analyze whether genealogies from loci participating in Dobzhansky-Muller (DM) incompatibilities are more likely to be concordant with the species tree under incomplete lineage sorting (ILS). The genealogies of individual loci differ from the species tree with a predictable frequency due to ILS. We combine these coalescent expectations with the emergence of epistatic incompatibility loci, according to the DM model, to determine the probability of concordance at reproductively isolating loci. Contrary to existing verbal models, we find that reproductively isolating loci that follow the DM model are more likely to have discordant gene trees. These results depend on the pattern of isolation, the time between speciation events, and the time since the last species diverged. However, we found support for a higher probability of discordance regardless of whether incompatibilities were allowed to segregate in the same population and regardless of whether DM pairs were derived-derived or derived-ancestral. Our overall results suggest that DM loci are unlikely to be especially useful for reconstructing species relationships and may in fact be positively misleading.

21 Speciation, sex chromosomes, and the sensitivity of spermatogenesis. *Erica Larson*[1], Jeffrey Good[2] 1) Department of Biological Sciences, University of Denver, Denver, CO; 2) Division of Biological Sciences, University of Montana, Missoula, MT.

Sex chromosomes play a large role in speciation. Hybrid sterility—one of the most common and rapidly evolving causes of speciation—is largely associated with the X chromosome (in XY systems). At the same time, the proper regulation of the sex chromosomes is a critical component of normal male fertility. The sex chromosomes are transcriptionally silenced early in meiosis and then repressed during postmeiotic development. Disruption of sex chromosome inactivation causes abnormal sperm development—similar to the sterility syndrome observed in hybrid males. These parallel observations suggest that regulatory control of the sex chromosomes may be an underlying developmental mechanism for the evolution of hybrid male sterility. We used fluorescence activated cell sorting to developmentally stage spermatogeneic gene expression in two closely related subspecies of house mice, and their sterile and fertile F1 hybrids. We found that sex chromosome expression was disrupted at every major stage of spermatogenesis, and that each likely reflects distinct mechanistic and genetic bases. Disruption of X inactivation during meiosis was associated with increased autosomal asynapsis, consistent with divergence at

the hybrid incompatibility gene, *Prdm9*. Postmeiotic disruption was associated with the over- and under-expression of sex chromosomes in reciprocal hybrids, supporting a hypothesis that genomic conflict between sex-linked genes leads to dosage imbalances in hybrids. Our results indicate that disruption of sex chromosome expression plays a central role in evolution of hybrid male sterility and the large X-effect in mouse speciation.

22 The snowball effect requires simple epistasis. *R. B. R. Azevedo* Dept Biol & Biochem, Univ Houston, Houston, TX.

Genetic incompatibilities can accumulate as a by-product of genetic divergence. Dobzhansky and Muller proposed that an allele that fixes in one population may be incompatible with an allele at a different locus in another population when the two alleles are brought together in a hybrid. Orr proposed that these Dobzhansky–Muller incompatibilities (DMIs) should accumulate faster than linearly—i.e., snowball—as two lineages diverge. Kalirad & Azevedo (2017) showed that, under the Orr model, the number of inviable introgressions at single divergent sites (IISs) is also expected to snowball. They also investigated the snowball in evolutionary simulations on a holey fitness landscape defined by a computational model of RNA folding. They found that IISs snowball, but more slowly than expected under Orr's model. Here, I conduct a similar investigation on holey fitness landscapes defined by the NK model. I find that the IISs snowball approximately as predicted by the Orr model only when the complexity of epistasis, defined by K, is low. As K increases the pattern of accumulation of IISs becomes more and more linear. When epistasis is maximally complex (i.e., K = N – 1), the holey landscape becomes completely random and IISs accumulate linearly with the level of divergence at a rate equal to the probability that a genotype is inviable. Thus, the snowball effect for inviable introgressions at single divergent sites requires simple epistasis.

23 The optimal mating distance resulting from heterosis and genetic incompatibility. *Xinzhu (April) Wei*, Jianzhi Zhang Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI.

The genetic distance between the two parents of an individual, or mating distance, influences the individual's fitness via two competing mechanisms. On the one hand, increasing the mating distance is beneficial because of the phenomenon of heterosis. On the other hand, too large of a mating distance is harmful owing to genetic incompatibility. It is thus believed that the fitness of a genotype is a hump-shaped function of the mating distance, culminating at an intermediate distance referred to as the optimal mating distance (OMD). However, decades of research has generally failed to validate this belief or identify the OMD. Here we address this question using large datasets from the plant *Arabidopsis thaliana*, fungus *Saccharomyces cerevisiae*, and animal *Mus musculus*, including phenotypic measures of multiple fitness-related traits from tens to hundreds of crosses and whole-genome sequence-based mating distance estimates. In each species, we find the hybrid phenotypic value a humped quadratic polynomial function of the mating distance for the vast majority of traits examined, with different traits exhibiting similar OMDs. OMDs are generally slightly greater than nucleotide diversities but smaller than the maximal observed genetic distances within species. Hence, the benefit of heterosis is at least partially offset by the harm of genetic incompatibility even within species. These results have implications for speciation, conservation, agriculture, and human health.

24 Population genetic tests for the direction and relative timing of introgression. *M.S. Hibbins*, M.W. Hahn Indiana University, Bloomington, IN.

Introgression is a pervasive biological process, and many statistical methods have been developed to infer its presence from genomic data. However, many of the consequences and genomic signatures of introgression remain unexplored from a methodological standpoint. Here, we develop a coalescent model of introgression using a parent-tree framework, and from it propose two new test statistics: D_1 and D_2. D_1 provides information on the timing of introgression relative to speciation, with the null hypothesis being that they are effectively simultaneous, and the alternative being that introgression occurs significantly afterwards. This statistic will prove useful for addressing hypotheses related to both the origin of an admixed population and homoploid hybrid speciation, in which speciation and introgression can be considered to occur simultaneously. D_2 provides a test for the direction of introgression, and has the advantages of being generally applicable and requiring data from only four taxa, similar to the ABBA-BABA test. Simulations show that D_1 and D_2 have the power to reject their null hypotheses in reasonable regions of parameter space. We apply the D_1 statistic to genomic data from the wild yeast *Saccharomyces paradoxus*, a proposed case of homoploid hybrid speciation, demonstrating its use as a test of this model. These statistics provide new and powerful ways to address questions relating to the timing and direction of introgression.

25 Tracking short-term evolution in a pedigreed wild population. *N. Chen*[1], I. Juric[2], E. Cosgrove[3], R. Bowman[4], J. Fitzpatrick[5], A. Clark[3,5], G. Coop[1] 1) Department of Evolution & Ecology, Center for Population Biology, University of California, Davis, Davis, CA; 2) Cleveland Clinic, Cleveland, OH; 3) Department of Molecular Biology & Genetics, Cornell University, Ithaca, NY; 4) Avian Ecology, Archbold Biological Station, Venus, FL; 5) Department of Ecology & Evolutionary Biology, Cornell University, Ithaca, NY.

Recent studies have demonstrated phenotypic evolution on ecological timescales in many different organisms, but

investigations of short-term evolutionary dynamics at the genomic level remain challenging and rare. Here, we directly characterize the relative roles of different evolutionary processes in shaping patterns of genetic variation through time using a 25-year genomic, phenotypic, and pedigree dataset in the Florida Scrub-Jay (*Aphelocoma coerulescens*), an iconic species on the U.S. Endangered Species List. 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 the past two decades (3,838 individuals total) at 15,416 genome-wide SNPs. We used gene dropping simulations to estimate individual genetic contributions and model drift and immigration on the known pedigree. We estimated the empirical strength of genetic drift each year by partitioning the proportion of variance in allele frequency change through time. After accounting for drift and gene flow, we identified SNPs whose frequency dynamics were driven by directional or fluctuating selection. Finally, we identified loci under selection acting on specific life-cycle stages, including male gametic selection, sexual selection, and viability selection. By combining pedigree-based models with fine-scale dissection of selection components, this study provides a detailed assessment of the roles of selection, gene flow, and drift in governing allele frequency dynamics in a natural population.

26 Massive variation of short tandem repeats with functional consequences across strains of *Arabidopsis thaliana*. *Maximilian O. Press*[1,2], Rajiv McCoy[1,4], Ashley Hall[1,3], Joshua Akey[1,4], Christine Queitsch[1] 1) University of Washington, Dept. of Genome Sciences, Seattle, WA; 2) current address: Phase Genomics, Seattle, WA; 3) University of Washington, Dept. of Molecular and Cellular Biology, Seattle, WA; 4) current address: Princeton University, Dept. of Ecology and Evolutionary Biology, Princeton, NJ.

Short tandem repeat (STR) mutations may be responsible for more than half of the mutations in eukaryotic coding DNA, yet STR variation is rarely examined as a contributor to complex traits. We assess the scope of this contribution across a collection of 96 strains of Arabidopsis thaliana by massively parallel STR genotyping. We found that 95% of examined STRs are polymorphic, with a median of six alleles per STR in these strains. Modest STR expansions are found in most strains, some of which have evident functional effects. For instance, three of six intronic STR expansions are associated with intron retention. Coding STRs are depleted of variation relative to non-coding STRs, consistent with the action of purifying selection, and some STRs show hypervariable patterns consistent with diversifying selection. Finally, we detect dozens of novel STR-phenotype associations that could not be detected with SNPs alone, and validate two with follow-up experiments. Our results demonstrate that STRs comprise a large, unascertained reservoir of functionally relevant genomic variation. (Preprint describing this work: https://doi.org/10.1101/145128)

27 Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions. Iskander Said[1], Ashley Byrne[2], Victoria Serrano[1], Charis Cardeno[3], Christopher Vollmers[1,4], *Russ Corbett-Detig*[1,4] 1) BME, UC Santa Cruz; 2) MCD, UC Santa Cruz; 3) EEB, UC Davis; 4) Genomics Institute, UC Santa Cruz.

Chromosomal inversions are widely thought to be favored by natural selection because they suppress recombination between alleles that have higher fitness on the same genetic background or in similar environments. Nonetheless, few selected alleles have been characterized at the molecular level. Gene expression profiling provides a powerful way to identify functionally important variation associated with inversions and suggests candidate phenotypes. However, altered genome structure itself might also impact gene expression by influencing expression profiles of the genes proximal to inversion breakpoint regions, or by modifying expression patterns genome-wide due to rearranging large regulatory domains. In natural inversions, genetic differentiation and genome structure are inextricably linked. Here, we characterize differential expression patterns associated with two chromosomal inversions found in natural *Drosophila melanogaster* populations. To isolate the impacts of genome structure, we engineered synthetic chromosomal inversions on controlled genetic backgrounds with breakpoints that closely match each natural inversion. We find that synthetic inversions have negligible effects on gene expression. Nonetheless, natural inversions have broad-reaching regulatory impacts in *cis* and *trans*. Furthermore, we find that differentially expressed genes associated with both natural inversions are enriched for loci associated with immune response to bacterial pathogens. Our results therefore strongly support the idea that inversions in *D. melanogaster* experience natural selection to maintain associations between functionally related alleles to produce complex phenotypic outcomes.

28 How brown rats adapted to life in NYC's concrete jungle. *Arbel Harpak*[1], Nandita Garud[2], Dmitri Petrov[1], Noah Rosenberg[1], Pleuni Pennings[3], Jason Munshi-South[4] 1) Stanford University, San Francisco, CA; 2) University of California San Francisco, San Francisco, California; 3) San Francisco State University, San Francisco, California; 4) Fordham University, Armonk, New York.

Brown rats (*Rattus Norvegicus*) populations have recently grown to enormous sizes in urban environments. From the human perspective, they wreak havoc in their path, causing hundreds of millions of dollars in damage to health and infrastructure. What role did genetic adaptation play in the spread of rats in cities? To answer this question, we collected whole-genome samples from 29 brown rats from New York City (NYC) and scanned for genetic signals of adaptation. We applied multiple genomic tools, each tailored to identify specific modes of adaptation. Intriguingly, we found evidence for

recent selective sweeps in genes associated with rodenticide resistance, olfactory and other sensory perception genes, and genes possibly related to behavioral changes. Finally, we investigated whether these adaptations are specific to urban environments by scanning for similar adaptations in a sample from the presumed ancestral range of brown rats in rural north-east China. Our results suggest a genetic component to the adaptation of rats in response to human activities, and could be applicable to urban rat control around the world.

29 Representing population structure with effective migration surfaces. *J. Novembre*[1], B. Peter[2], H. Al-asadi[1], D. Petkova[1], M. Stephens[1] 1) University of Chicago, Chicago, IL, USA; 2) Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Population structure is a fundamental feature of genetic variation that is crucial to address when carrying out studies in evolutionary biology, conservation genetics, and trait mapping. However, most analytical methods to represent population structure do not incorporate geography directly. Typically, geography must be considered *post hoc* alongside a visual summary of genetic data. Recently, we developed a spatially explicit method that estimates "effective migration surfaces" to visualize how genetic diversity is geographically structured (the EEMS method). Here, we share insights from applications of the EEMS method using examples from multiple species, leveraging both published studies and novel applications. In particular, using a novel analysis of over 8,000 widespread human samples we find surfaces that are "rugged", which indicates the relationship between genetic and geographic distance is heterogeneous and subtly distorted as a rule. We also highlight possible pitfalls of applying the method and share novel lines of development of the EEMS method, including an extension that analyzes long-shared haplotype tracts, which can be used to assess population structure on different time-scales.

30 Detection of shared balancing selection in the absence of trans-species polymorphism. *X. Cheng*[1,2], M. DeGiorgio[2,3,4] 1) The Huck Institutes of the Life Sciences, the Pennsylvania State University, University Park, PA; 2) Department of Biology, The Pennsylvania State University, University Park, PA; 3) Department of Statistics, The Pennsylvania State University, University Park, PA; 4) Institute for CyberScience, the Pennsylvania State University, University Park, PA.

Trans-species polymorphisms have been widely used as the key sign of long-term balancing selection across multiple species. However, such sites are often rare in the genome, and could result from mutational processes or technical artifacts. No methods are yet available to specifically detect footprints of trans-species balancing selection without using trans-species polymorphic sites. In this study, we developed summary- and model-based approaches that are each specifically tailored to uncover regions of long-term balancing selection shared by a set of species by using genomic patterns of intra-specific polymorphism and inter-specific fixed differences. We demonstrate that our trans-species statistics have substantially higher power than single-species approaches to detect footprints of trans-species balancing selection, and are robust to those that do not affect all tested species. We further applied our model-based methods to human and chimpanzee whole genome sequencing data, and have identified the most outstanding candidates to be the MHC locus and the malaria resistance-associated FREM3/GYPE region, consistent with previous findings. A number of regulatory elements also exhibit signals of trans-species balancing selection, and among them we characterized two clusters of elements locating upstream of the innate immunity gene MBL2. Our findings echo the significance of pathogen defense in establishing balanced polymorphisms across human and chimpanzee lineages, and suggest that non-coding regulatory regions may play an important role. Additionally, we have shown that these trans-species statistics can be applied to and work well for more than two species, and have integrated them into open-source software packages for ease of use by the scientific community.

31 Direct estimation of mutation rates in owl monkeys shows that life history is the main determinant of rate variation in primates. *G.W.C Thomas*[1,2], R.J Wang[1], A. Puri[2], R.A. Harris[3,4], M. Raveendran[3,4], J. Rogers[3,4], P. Radivojac[2], M.W. Hahn[1,2] 1) Department of Biology, Indiana University, Bloomington, IN; 2) Department of Computer Science, Indiana University, Bloomington, IN; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Mutation rates vary between species, and many explanations regarding the causes of this variation have been proposed. Many of these hypotheses revolve around changes to the underlying molecular mutational machinery. However, recent genome sequencing of human pedigrees has shown that most mutations arise in males after puberty, highlighting the importance of life history in the determination of a species' mutation rate. It remains unclear whether changes in life history alone are enough to account for variation in mutation rates across species, or if changes to the molecular machinery are necessary as well. Here, we sequence the genomes of 30 owl monkeys (*Aotus nancymaae*) within 5 multi-generation pedigrees. This allows us to identify *de novo* mutations between generations and to identify the parent-of-origin of many mutations. We find that owl monkeys have an average mutation rate of 5.5×10^{-9} per generation per site, roughly 30% lower than that estimated in humans. We show that, as in humans, the number of mutations passed on to an offspring in owl monkeys depends on the age of the father. Finally, we model the effects of life history traits such as age at puberty and length of reproductive period to predict species-specific mutation rates. We show that variation in life history traits alone can explain variation in the per-generation mutation rate among primates.

32 The genomics of conflict: X-chromosome drive in the fly *Drosophila neotestacea*. K.E. Pieper[1], R.L. Unckless[2], *K.A. Dyer*[1] 1) Dept of Genetics, University of Georgia, Athens, GA; 2) Dept of Molecular Biosciences, University of Kansas, Lawrence, KS.

Meiotic drivers manipulate gametogenesis to achieve a transmission advantage, and as a result cause genetic conflict between the driver and the rest of the genome. *Sex-ratio* drive involves selfish elements located on the X-chromosome that inhibit the transmission of Y-bearing sperm in males. As a consequence, the *sex-ratio* X-chromosome is transmitted to all of the male's offspring, which are all daughters. These *sex-ratio* (SR) X-chromosomes have evolved independently many times in flies and are frequently associated with large inversions. Here we study the genetic causes and genomic consequences of *sex-ratio* drive in the fly *Drosophila neotestacea*. *Sex-ratio* drive in this species occurs at high prevalence in natural populations, and there are no segregating suppressors of drive. The SR X-chromosome of *D. neotestacea* is genetically differentiated from the wild-type chromosome due to large chromosomal inversions, but unlike many other drive systems there is substantial genetic variation present on SR and evidence of some limited gene flow between chromosome types. In this study, we use a combination of transcriptomics and molecular evolution to identify widespread expression and sequence differentiation between the standard (ST) and SR X-chromosomes of *D. neotestacea.* We find that the X-chromosome is enriched for transcripts that are differentially expressed between ST and SR. There is widespread sequence divergence between ST and SR, and many transcripts had elevated K_a/K_s values. We identify a set of candidate transcripts, including a testis-specific, X-linked duplicate of the nuclear transport gene *importin-α2* that is overexpressed in SR*.* Population genetic analyses of this gene suggest that positive selection has occurred in the lineage leading to the duplicate but that relaxed purifying selection has been prevalent in ST, consistent with the involvement of this gene in the mechanism of drive in this species. Finally, we infer that this duplicate is found in two closely related species that have SR drive, but not found in other related species that do not have SR drive. In sum, this gene represents a strong candidate for involvement in *sex-ratio* drive. We suggest that nuclear transport may be a common target for genetic conflict, as the mechanism of the autosomal *Segregation Distorter* drive system in *D. melanogaster* involves the same pathway*.*

33 The effect of strong purifying selection on mutational trajectories and genetic diversity. *I. Cvijovic*[1,2], B. H. Good[3], M. M. Desai[1,2,4] 1) FAS Center for Systems Biology, Harvard Univeristy, Cambridge, MA 02138; 2) Department of Organismic and Evolutionary Biology Harvard University, Cambridge, MA 02138; 3) Departments of Physics and Bioengineering, University of California, Berkeley CA 94720; 4) Department of Physics, Harvard University, Cambridge, MA 02138.

Purifying selection reduces genetic diversity, both at sites under direct selection and at linked neutral sites. This process, known as background selection, is thought to play an important role in shaping genomic diversity in natural populations. Yet despite its importance, the effects of background selection are not fully understood. Previous theoretical analyses of this process have taken a backwards-time approach based on the structured coalescent. While they provide some insight, these methods are either limited to very small samples or are computationally prohibitive. I will present a new forward-time analysis of the trajectories of both neutral and deleterious mutations at a nonrecombining locus. Using this approach, we find that strong purifying selection leads to remarkably rich mutational dynamics: neutral mutations can exhibit sweep-like behavior, and deleterious mutations can reach substantial frequencies even when they are guaranteed to eventually go extinct. Our analysis of these dynamics allows us to calculate analytical expressions for the full site frequency spectrum. We find that whenever background selection is strong enough to lead to a reduction in genetic diversity, it also results in substantial distortions to the site frequency spectrum, which can mimic the effects of population expansions or positive selection. Because these distortions are most pronounced in the low and high frequency ends of the spectrum, they become particularly important in larger samples, but may have small effects in smaller samples. This means that extrapolating conclusions from small samples about the effects of background selection can be grossly misleading. We also apply our forward-time framework to calculate other quantities, such as the ultimate fates of polymorphisms or the fitnesses of their ancestral backgrounds.

34 A population genetic interpretation of GWAS findings for human quantitative traits. *Y. Simons*, G. Sella Biological Sciences, Columbia University, New York, NY.

The genetic architecture of a quantitative phenotype (i.e., the number, frequency and effect size of alleles underlying variation in its value) arises from genetic and population genetic processes. Mutations affecting the trait appear at a rate that reflects the target size, and their trajectory through the population is determined by demographic processes and by the selection acting on them. Many phenotypes, including human height and body mass index (BMI), appear to be under stabilizing selection, either because of selection on the trait itself or through the effects of genetic variation on other traits (i.e., via pleiotropy).

With these considerations in mind, we introduce and solve a generative model for the genetic architecture of a continuous trait under direct and pleiotropic stabilizing selection. We derive simple and robust predictions for the distribution of additive genetic variation among loci. We then relate these predictions to observations from GWAS, accounting for how the power to detect a locus depends on its contribution to additive genetic variation.

This new theory allows us to make inferences about the population genetic processes that underlie genetic variation for

height and BMI in Europeans. We find an extremely good fit to GWAS findings (Wood et al. *Nature Genetics* 2014, Locke et al. *Nature* 2015): by fitting a single parameter for each trait, we are able to explain the distribution of additive genetic variation over genome-wide significant associations. Accounting for the demographic history of European populations suggests that the current GWAS is well powered to identify only loci under moderate selection.

35 The polygenic basis of an ancient divergence in yeast thermotolerance. *C. Weiss*[1], J. Roop[1], R. Hackley[1,2], J. Chuong[2], I. Grigoriev[1,3], A. Arkin[1,4], J. Skerker[1,4], R. Brem[1,2] 1) UC Berkeley, Berkeley, CA; 2) Buck Institute for Research on Aging, Novato, CA; 3) US Department of Energy Joint Genome Institute, Walnut Creek, CA; 4) Lawrence Berkeley National Laboratory, Berkeley, CA.

Some of the most unique and compelling survival strategies in the natural world evolved long ago, and are fixed in now-isolated species. Molecular insight into these adaptations has been limited, as classic experimental genetics has focused on the interfertile individuals of a population. Here we dissect a complex thermotolerance difference between yeast species that diverged millions of years ago. Using a new mapping approach that screens mutants in a sterile interspecific hybrid, we identified eight genes that underlie the growth advantage of *Saccharomyces cerevisiae* over its sister species *S. paradoxus* at high temperature. All eight encode housekeeping factors with no known direct function in heat-shock or stress response. Pro-thermotolerance alleles at these mapped loci were required for the adaptive trait in *S. cerevisiae* and sufficient for its partial reconstruction in *S. paradoxus*. Together, our data reveal the genetic mechanism by which *S. cerevisiae* acquired its high-temperature growth advantage in the distant past. And our study lays the groundwork for the mapping of genotype to phenotype in clades of sister species across Eukarya.

36 Detecting signatures of convergent adaptation in population genomic data. *K.M. Lee*[1], J.P. Selby[2], J.H. Willis[2], G.M. Coop[1] 1) Department of Evolution and Ecology, University of California, Davis, Davis, CA; 2) Department of Biology, Duke University, Durham, NC.

Convergent evolution, in which selection for the same trait occurs independently in several lineages, can be leveraged to identify both the ecological and molecular basis of adaptation. When convergent adaptation occurs among closely-related populations, it may be difficult to distinguish adaptive convergence from drift or shared evolutionary history. Here, we develop a method to identify adaptive convergence genome-wide among closely related populations, while accounting for neutral population structure and background selection. We build on our previous results to create a coalescent-model-based, composite-likelihood method to search for regions with lower within-population coalescence times in neutral loci, relative to demographic expectations. We expect to see this pattern at neutral loci near alleles that have increased in frequency due to selection across multiple populations. Once putatively selected alleles are identified, between-population coalescent times are considered to determine whether the alleles were derived independently or have a shared origin via either gene flow or ancestral standing variation. We illustrate our approach with genome-wide polymorphism data from five populations of *Arabidopsis lyrata*, identifying loci involved in the adaptation to serpentine soil in the Eastern United States. We identify novel loci involved in this adaptation, finding alleles shared between sets of serpentine-adapted *lyrata* populations as well as alleles that are likely derived from independent mutational events.

37 Selfing can facilitate transitions between pollination syndromes. *Carolyn Wessinger*, John Kelly Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS.

Pollinator-mediated selection on plants can favor transitions to a new pollinator depending on the relative abundances and efficiencies of pollinators present in the community. A frequently observed example is the transition from bee pollination to hummingbird pollination. We present a population genetic model that examines whether the ability to inbreed can influence evolutionary change in traits that underlie pollinator attraction. We find that a transition to a more efficient, but less abundant pollinator is favored under a broadened set of ecological conditions if plants are capable of delayed selfing rather than obligately outcrossing. Delayed selfing allows plants carrying an allele that attracts the novel pollinator to reproduce even when this pollinator is rare, providing reproductive assurance. In addition, delayed selfing weakens the effects of Haldane's sieve by increasing the fixation probability for recessive alleles that confer adaptation to the new pollinator. Our model provides novel insight into the paradoxical abundance of recessive mutations in adaptation to hummingbird attraction. It further predicts that transitions to efficient but less abundant pollinators (such as hummingbirds in certain communities) should disproportionately occur in self-compatible lineages. Currently available mating system datasets are consistent with this prediction and we suggest future areas of research that will enable a rigorous test of this theory.

38 A temporal signal of linked selection. *V. Buffalo*, G. Coop UC Davis, Davis, CA.

A long-standing problem in evolutionary genetics is to determine the extent to which allele frequency changes are shaped by neutral genetic drift versus by selection. Selection can change allele frequencies either directly, if the genotypes at a locus directly impact fitness, or indirectly, if a polymorphism is affected by selection at linked sites.

Recently progress has been made in quantifying the substantial impact of linked selection on genome-wide levels of genetic diversity. However, we may be missing much of the contribution of linked selection if selection pressures are dynamic over short time periods. We demonstrate how time series data offers a way forward, allowing us to separate allele frequency change into drift and selection components. Our approach relies on a unique signal of polygenic fitness variation in the population: the covariances in allele frequency changes through time. We characterize and model these covariances through new theoretic results and simulations, and show how these allows us to directly observe the degree to which selection shapes allele frequency changes through time. Finally, we show how this theory can be extended to detect adaptation on contemporary timescales.

39 Selection against LTR retrotransposons is balanced by locally adapted transposable element alleles in *Arabidopsis thaliana*. *M.C. Stitzer*, J. Ross-Ibarra University of California, Davis, Davis, CA.

Although transposable elements (TEs) contribute to the genomes of virtually all eukaryotic organisms, their abundances, types, and frequencies differ within and between species. Classical theory posits that TEs are, on average, slightly deleterious to their host genome. But averages disguise evolutionarily relevant variation in TE impacts on the host genome. To address how selection acts on intraspecific variation at individual TE loci, we characterize the positions of LTR retrotransposons in two *Arabidopsis thaliana* reference genomes and use resequencing data from a range-wide sample of 900 individuals to identify over 7,500 polymorphic TEs. Using sequence variation that accumulates within each TE after insertion, we calculate the age of each copy in the genome. We leverage the relationship between the age of a TE allele and its expected frequency under neutrality to characterize the action of natural selection. While the majority of TEs are young and found at low frequency, consistent with widespread negative selection, over 700 ancient TEs remain polymorphic, despite having inserted in the genome millions of years in the past. These ancient non-neutral TE alleles alter expression of adjacent genes, change local epigenetic regulation, and are associated with climatic variation across the landscape. Sixty of these fossil TEs are also polymorphic in sister species *Arabidopsis lyrata*, consistent with balancing selection retaining these alleles in spatially and temporally variable environments. Together, our results provide evidence for complex interactions of TEs with their host genome, blurring the line between their classification as harmful mutagens and adaptive variation.

40 Identifying the molecular basis of convergent adaptation to herbivorous diets in mammals. *W.K. Meyer*[1], K.D. Kohl[2], N.L. Clark[1] 1) Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA.

What genomic changes underlie adaptations to new environments? Herbivory represents an excellent case study of a complex phenotype that has arisen independently numerous times throughout mammalian evolution, necessitating physiological and anatomical adaptations for processing high-fiber, low-nutrient food sources that also frequently contain toxic compounds. Although several studies have investigated adaptation to herbivorous diets in individual species or explored candidate genes involved in multiple cases of herbivory, to date no study has queried the genome-wide molecular basis of convergent adaptation to herbivorous diets across mammals.

We identify genes and regulatory sequences whose patterns of evolution are associated with transitions to herbivory, by finding correlations between changes in phenotype and changes in molecular traits across branches of the mammalian phylogeny. Rather than looking for sequence-level convergence, which is likely rare, we seek patterns of convergence in broader-scale molecular traits associated with transitions to herbivory:
1) Convergent shifts in evolutionary rates of genes and conserved regulatory elements
2) Convergent functional losses of genes
3) Convergence in chemical properties such as isoelectric point of proteins

We here apply these methods to genome-wide data from 64 mammals, adding key species such as the beaver and desert woodrat to publicly available alignments. Genes showing evidence of acceleration within herbivorous lineages include *GALNT10*; this gene plays a role in the glycan biosynthesis pathway, which also involves microbial digestion within the gut. We additionally investigate convergence in chemical properties for genes such as *LYZ*, an enzyme that plays a digestive role specifically in ruminants and leaf-eating monkeys. In these species, *LYZ* is present in the stomach, where it encounters highly acidic environments the protein-cleaving enzyme pepsin; consequently, such co-opted enzymes show patterns of chemical convergence in herbivorous species. We also compare patterns of convergence across sets of species with distinct anatomical and physiological adaptations to plant-based diets, and we investigate the impact of differences in dietary content on patterns of convergence. These analyses improve our understanding of the gene networks underlying adaptation to herbivory across mammals, as well as of the selective constraints experienced by genes that respond to broad or specific shifts in diet.

41 Understanding Adaptation and Fitness Trade-offs in Yeast. *Y. Li*[1], S. Venkataram[1], A. Agarwala[2], D. Fisher[3], D. Petrov[1], G. Sherlock[4] 1) Departments of Biology, Stanford University; 2) Departments of Physics, Stanford University; 3) Departments of Applied Physics, Stanford University; 4) Departments of Genetics, Stanford University.

Few studies have *quantitatively* probed how adaptive mutations result in increased fitness. Even in microbial evolution experiments, with full knowledge of the underlying mutations and specific growth conditions, it is still challenging to determine where within a growth-saturation cycle those fitness gains occur. I have characterized thousands of evolved yeast clones, each carrying a unique DNA barcode and quantified their fitness gains that result from different phases of the growth cycle by measuring their fitness under conditions where lengths of fermentation, respiration and stationary phases were systematically varied. A common implicit assumption is that most benefits derive from an increased exponential growth rate. However, I instead found that while all adaptive lineages gained similar and modest benefits from fermentation, most of the benefits came instead from respiration where there is little cell division. From pairwise fitness competitions for a dozen of these clones, I determined that the benefits *accrued* during respiration are *realized* later as a shorter duration of lag phase in the following growth cycle. These results reveal hidden complexities of the adaptive process even under ostensibly simple evolutionary conditions, and the sensitivity of fitness to subtle quantitative changes of conditions.

This sensitivity of fitness to the growth cycle conditions suggests that adaptation would proceed differently under conditions with a different growth-saturation cycle. To study this, I evolved barcoded yeast population under conditions modified from the previous Original Evolutionary Condition (OEC, which includes lag, fermentation and respiration phases). The Modified Evolutionary Conditions (MECs) have either a shorter growth cycle, without respiration, or a longer cycle with a prolonged stationary phase. I have found, by contrast to adaptation in the OEC, adaptive clones from the MECs have 1) a different genetic basis, e.g. mutations in the HOG pathway instead of the RAS/PKA pathway in OEC, and 2) different adaptive strategies, gain and loss of fitness benefits from different growth phases. My study of adaptation under systematically varied conditions has enabled the dissection of regulatory networks and intrinsic evolutionary constraints underlying yeast growth. This approach provides the opportunity to better understand how organismal evolution can be constrained by a changing environment.

42 Validated SNP for bone strength in laying hens show strong GxE. Fernando Lopes-Pinto[1], Heather McCormack[2], Helena Wall[1], Robert Fleming[2], Ian Dunn[2], Andreas Kindmark[3], *Dirk Jan de Koning*[1] 1) Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, SE; 2) The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9RG, Scotland; 3) Department of Medical Sciences, Uppsala University Hospital, 75185 Uppsala, Sweden.

Bone weakness is a major welfare problem in laying hens which can be improved through genetic selection. QTL studies on divergent crosses and GWAS studies within breeding lines have revealed a number of candidate genes and significant SNP affecting bone strength. For implementation of these markers in marker assisted selection or genomic selection, the effects need to be validated in commercial crossbred layers in a variety of production systems. Here we tested 111 candidate SNP from previous GWAS for their effect on tibial breaking strength on 856 birds from two different companies, kept in two different housing systems.

The experiment included two commercial lines: LSL Classic (Lohmann Tierzucht GmbH, Cuxhaven, Germany) and Bovans Robust (Hendrix-Genetics, Boxmeer, The Netherlands). At our experimental farm Lövsta, 1620 birds from each breed were divided over 18 floor pens (100 birds/pen) and 180 furnished cages (8 birds/cage). Birds were also assigned one of two experimental diets. This allowed us to test any interactions between the breed, diet, and housing systems as well as their respective interactions with the putative genetic effects. Selected birds were evaluated according to a 24-point integument score at 35, 55, and 80 weeks of age as well as at the end of the experiment at 100 weeks of lay. As well as an experiment in its own right, this meant we could test any pleiotropic effects of the candidate SNPs. Other additional studies included evaluation of internal and external egg quality traits at three time periods during lay.

There was a significant effect ($P < 0.001$) whether birds were housed in furnished cages or floor pens. There was a strong interaction between the housing system and the SNP results: all significant SNP ($P < 0.05$) only had an effect in one of the housing systems. The study successfully validated QTL affecting bone strength on chromosomes 1, 3, 8, 13, and 19 with seven significant QTL acting in cages and seven being specific for pen housed birds. SNP associated effects were in the region of 40 Newton with considerable dominance effects. The study shows clear GxE effect for SNP affecting bone strength. This has clear implications for genomic selection for bone strength when selection is focused on purebred great-grandparent lines that are kept in different housing systems than commercial laying hens.

43 The genomic basis of environmental adaptation in house mice. *M.W. Nachman*[1], M. Phifer-Rixey[1,2], K.L. Mack[1], M. Ballinger[1], K. Bi[1], K.G. Ferris[1], M.J. Sheehan[1,3], D. Lin[1], S.M. Keeble[4], J.M. Good[4], T. Suzuki[1] 1) Integrative Biology, UC Berkeley, Berkeley, CA; 2) Department of Biology, Monmouth University, NJ; 3) Department of Neurobiology and Behavior, Cornell University, Ithaca, NY; 4) Division of Biological Sciences, University of Montana, Missoula, MT.

House mice (*Mus musculus domesticus*) are native to Western Europe and have recently expanded into a wide range of novel environments, providing an exceptional opportunity to study the genetic basis of evolutionary change over short time

scales. We sampled mice along a latitudinal transect in North America and sequenced exomes at moderate coverage and whole genomes at low coverage to identify loci contributing to adaptation. We also established a new set of wild-derived lines of mice from the ends of the transect to measure phenotypes and study gene expression in a common laboratory environment. Finally, we studied expression and mapped expression quantitative trait loci (eQTL) in wild mice to identify gene regulatory differences contributing to adaptation. Mice from different latitudes differed in body size, activity level, and metabolic phenotypes when measured in the lab, indicating that these differences are genetically determined. For example, mice from warmer environments were smaller, conforming to Bergmann's Rule. A large majority of mutations contributing to adaptation appear to be regulatory. Mice sampled from the ends of the transect show significant differences in gene expression in liver, hypothalamus and adipose tissue, and we discovered genes with cis-regulatory differences that show strong signatures of selection. The combination of genome-wide analyses of genetic variation, measurement of gene expression, and characterization of phenotypes in the lab identified specific genes underlying environmental adaptation, including two genes that account for a significant amount of phenotypic variance in body size. Surprisingly, we found parallels between the genetic and phenotypic signatures of environmental adaptation in humans and mice, including those related to metabolic disease.

44 Genetic architecture of parallel adaptation across spatially explicit population samples. Andrea Fulgione, Célia Neto, *Angela Hancock* Max Planck Institute for Plant Breeding Research, Cologne, Nordrhein Westfalen, DE.

Reconstructing the details of adaptive processes is challenging due to the complexity inherent in natural populations, but islands can represent useful cases where complexity is reduced and these details can be revealed. The Cape Verde Islands are an archipelago 600 km off the coast of Senegal. Populations living here face a harsh arid climate and subsist on volcanic soil. We sequenced the genomes of 200 *Arabidopsis thaliana* accessions from these islands as well as 70 Moroccan outgroup individuals and reconstructed the colonization history of the island populations. Then, we used a combination of quantitative and population genetic approaches to reconstruct adaptive history. We infer that initial island colonization occurred 7-10 kya by a single seed from Morocco, and subsequently a second island was colonized, again by a single seed ~5000 generations ago. Since the split, the two island populations have been evolving separately with no subsequent gene flow. We identified several variants responsible for ongoing parallel adaptation for fitness-related complex traits between islands. The limited complexity of this natural experiment combined with spatially explicit sampling allows us to uncover precise details of the evolutionary dynamics between these quantitative trait variants over time and space.

45 Periodic variation of mutation rates in bacterial genomes associated with replication timing. *M. M. Dillon*[1,3], V. S. Cooper[2,3] 1) Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, CA; 2) Department of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH.

Spontaneous mutation rates may vary within genome regions across the tree of life, but the causes and consequences of this spatiotemporal variation are uncertain. We examined relationships between local mutation rates and replication timing in three bacterial species whose genomes have multiple circular chromosomes: *Vibrio fischeri*, *Vibrio cholerae*, and *Burkholderia cenocepacia.* Five mutation accumulation experiments analyzed by whole genome sequencing were conducted: three with wild-type strains of each species, and two with mismatch repair-deficient strains of *V. fischeri* and *V. cholerae*. In the absence of mismatch repair, base-substitution mutation rates vary in a mirrored wave-like pattern on opposing replichores of the large chromosome of *V. fischeri* and *V. cholerae*, where concurrently replicated regions experience similar base-substitution mutation rates. The base-substitution mutation rates on the small chromosome are less variable in both species but occur at similar rates as the concurrently replicated regions of the large chromosome. Neither nucleotide composition nor frequency of nucleotide motifs differed among regions experiencing high and low base-substitution rates, which along with the inferred ~800kb wave period suggests that the source of the periodicity is not sequence-specific but rather a systematic process related to the cell cycle. We suggest that variation in dNTP pools governed by ribonucleotide reductase activity may provide a simple explanation for this genome-wide variation and offer a model that recapitulates the observed dynamic. These results support the notion that base-substitution mutation rates are likely to vary systematically across many bacterial genomes, which exposes certain genes to elevated deleterious mutational load.

46 Multi-environment fitness landscapes of a tRNA gene. *C. Li*[1,2], J. Zhang[1] 1) Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI; 2) Biology, Stanford University, San Jose, CA.

The fitness landscape (FL) of a gene maps each mutational variant of the gene to the fitness of the organism carrying the variant. FLs allow explaining as well as predicting evolutionary trajectories, and are therefore of fundamental biological importance. Because fitness depends on the environment, which frequently changes in nature, characterizing FLs in multiple environments offers not only more relevant and accurate evolutionary explanations and predictions but under certain circumstances also insights that are otherwise impossible to gain. Notwithstanding, past measures of FLs typically probe only one environment, because even the characterization of a small fraction of the FL of one gene under one environment had been a formidable challenge until recently. Using a high-throughput method that combines precise gene replacement with

next-generation sequencing, we determine in four environments the in vivo FL of a yeast tRNA gene comprising >23,000 genotypes. We observe pervasive genotype-by-environment interaction (G×E), confirming the value of multi-environment FL mapping. Unexpectedly, however, the G×E pattern observed is so simple that the FL in one environment can be computationally transformed to that in another environment with fitness measures of only a few genotypes in the new environment. Under each environment, we observe prevalent, negatively biased epistasis between mutations (G×G). Epistasis-by-environment interaction (G×G×E) is also abundantly detected, but trends in epistasis changes between environments are predictable. Our study thus reveals simple rules underlying seemingly complex multi-environment FLs, opening the door to understanding and predicting FLs in general.

47 The architecture of an empirical genotype-phenotype map. *Jose Aguilar Rodriguez*[1], Leto Peel[2], Massimo Stella[3], Andreas Wagner[4,5,6], Joshua Payne[7] 1) Department of Biology, Department of Chemical and Systems Biology, Stanford University, CA, USA; 2) Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Universite catholique de Louvain, Belgium; 3) Institute for Complex Systems Simulation, Department of Electronics and Computer Science, University of Southampton, UK; 4) Department of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland; 5) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 6) Santa Fe Institute, NM, USA; 7) Institute of Integrative Biology, ETH Zurich, Switzerland.

Recent advances in high-throughput technologies are bringing the study of empirical genotype-phenotype (GP) maps to the fore. Here, we use data from protein binding microarrays to study an empirical GP map of transcription factor (TF) binding preferences. In this map, genotypes are DNA sequences and phenotypes are the TFs that these sequences bind. We study this GP map using genotype networks, in which nodes represent genotypes with the same phenotype, and edges connect nodes if their genotypes differ by a single small mutation. We describe the structure and arrangement of genotype networks within the space of all possible binding sites for 525 TFs from three eukaryotic species encompassing three kingdoms of life (animal, plant, and fungi). We thus provide a high-resolution depiction of the architecture of an empirical GP map. Among a number of findings, we show that these genotype networks are "small-world" and assortative, and that they ubiquitously overlap and interface with one another. We use polymorphism data from *Arabidopsis thaliana* to show how genotype network structure influences the evolution of TF binding sites in vivo. We discuss our findings in the context of regulatory evolution.

48 Surprising mechanisms underlying evolution in a classic system: Revisiting ADH in *Drosophila*. *Mohammad Siddiq*[1], Joseph Thornton[1,2] 1) Ecology & Evolution, University of Chicago, Chicago, IL; 2) Human Genetics, University of Chicago, Chicago, IL.

A central goal in evolutionary biology is to uncover how genetic variation created by historical and ongoing evolutionary processes lead to phenotypic diversity. Development of statistical tests for selection and deeper genomic sampling of natural diversity has allowed articulation of hypotheses that predict how specific genetic changes affect molecular and organismal functions and fitness in biologically relevant environments. Experimentally testing such hypotheses of molecular adaptation, however, has been a major challenge because doing so requires characterizing the effects of genetic changes on molecules that are now extinct, and because identifying the actual genetic target(s) of selection from co-occurring variation is often not possible. Nevertheless, such tests are necessary to understand how molecular adaptation occurs and to avoid spurious inferences about cases and mechanisms of adaptation. Here, we demonstrate a way of testing such hypotheses by investigating in mechanistic detail predictions about adaptive protein evolution in the alcohol dehydrogenase enzyme (ADH) in *Drosophila*, a classic system in evolutionary genetics. To do this, we reconstruct and characterize ancient ADH enzymes alongside their modern descendants through biochemical assays *in vitro* and through physiological assays of transgenic organisms carrying the ancient alleles *in vivo*. Our experiments show that the most widely-held biological explanations for how selection acted on this gene are not supported on any empirical level and strongly refute nearly every functional prediction from the canonical adaptive ADH hypothesis. Interestingly, our experimental approach reveals other functional changes that occurred in ADH due to sequence changes thought to be relatively inconsequential, suggesting that selection may drive evolution in this gene but through different genetic and functional mechanisms than previously predicted. Together, our work shows a way in which hypotheses of historical molecular adaptation can be more directly tested, why doing so is necessary to avoid accepting intuitively appealing but ultimately spurious accounts of historical molecular adaptation, and that doing so can uncover previously unanticipated but potentially important mechanisms through which molecules and organisms evolve.

49M Severe population crashes decrease genome-wide diversity in an already genetically depauperate insular mammal. *N. Adams*[1], E. Watson[1], X. Wang[2], S. Edmands[1] 1) Biological Sciences, University of Southern California, Los Angeles, CA; 2) Vertebrate Paleontology, Natural History Museum of Los Angeles, CA.

The Channel Island fox, *Urocyon littoralis*, was delisted/down-listed from the federal endangered species list in 2016 after successful conservation efforts brought them back from a 90-99% population decline in the 1990s. While their demographic recovery has been dramatic, much less is known about their genetic recovery. Generally, it is not well understood how/if genomic regions (e.g. X chromosome vs. autosomes, immune function genes vs. non-immune related genes) respond

differently to severe bottlenecks. Additionally, there is no consensus on what metrics should be used to diagnose a recovered species- nucleotide diversity, heterozygosity, effective population size, runs of homozygosity, etc. Therefore we looked at the "natural" experiment of the Channel Island fox population declines, which varied in bottleneck presence, size, cause, and recovery across multiple islands. To assess their recovery, we conducted the first direct genetic comparison at the population level of historical samples (collected prior to the recent bottlenecks) and modern samples (collected after the crashes and subsequent expansions). We sequenced the fox exome using dog baits designed with the recent dog genome (CanFam3) and called variants against the CanFam3 reference assembly. Results show already genetically depauperate populations that were further degraded significantly by the 1990s population crashes and remain low. We also show a decrease in genetic diversity in a population that has not undergone a known population crash and highlight a set of significantly different variants between the historical and modern groups. Thus understanding the population genetics component is important because only the Santa Catalina subspecies is now federally listed as threatened, yet other subspecies such as the San Nicolas population have even lower genetic diversity which may limit their ability to adapt to changing environmental conditions. We can learn from these results that species conservation is more complex than population size and that some Island Fox populations may not be "out of the woods" yet.

50M The effect of spatially varying selection on transposable element insertions in Drosophila. *J. Adrion*[1], D. Begun[2], M. Hahn[1] 1) Indiana University, Bloomington, IN; 2) University of California Davis, Davis, CA.

Natural populations often exist in spatially diverse environments and may experience spatial variation in the strength and targets of natural selection over their ranges. This spatially varying selection can shape both the evolution of genome architecture and adaptation to the environment. Drosophila provides an excellent opportunity to study the effects of spatially varying selection in natural populations, as both *D. melanogaster* and *D. simulans* have recently (within the last 500 years) been introduced in North and South America, having since colonized the bulk of both continents. Previous studies have identified candidate single nucleotide polymorphisms (SNPs) that are potential targets of spatially varying selection, and have described broad patterns of SNP variation along clines in North America and Australia. Here, we investigated how spatially varying selection impacts another important source of genomic variation, transposable elements (TEs). TE insertion dynamics shape genome evolution and, in multiple instances, have been shown to be the causative mechanism underlying adaptation to the environment. Here, we discover clinal TE insertions in whole-genome pooled-population sequence data from six populations of *D. melanogaster* and nine populations of *D. simulans* sampled from the Americas. We characterize parallel variation in the distribution and allele frequency of individual TEs and TE families between species and we explore associations between TEs and features of the host genome. Our results shed light on the role for TE insertions to shape genome evolution and facilitate environmental adaptation in Drosophila.

51M Predicting the selective importance of genes in Drosophila melanogaster by simultaneous modeling of DNA sequence features, expression level, biological function and pathway features. *Hosseinali Asgharian*[1], Asif Zubair[2], Katrina Shebrina[2], Sergey Nuzhdin[2] 1) Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA; 2) Program in Molecular and Computational Biology, University of Southern California, Los Angeles, CA.

Many selectively important genes go through long periods of purifying selection punctuated by short bouts of positive selection, both processes resulting in negative values of Tajimas'D. We improve and previous attempts at predicting genes response to selection in several ways. First, by showing that the correlation of nucleotide diversity and variance in gene expression is strongest at the two extremes of Tajima's D, we provide evidence of concordant purifying selection on DNA sequence and expression levels and rule out different mutation rates as the sole underlying factor. Second, we integrate all the Reactome pathways by their shared nodes; introduce two new network measures, aggregate indegree and aggregate outdegree, to incorporate the effect of indirect interactions; and show that they predict selection response more strongly than simple indegree or outdgree or any of the betweenness metrics. A few previous attempts at associating network features (e.g. upstream or downstream positions) to selection response were unsuccessful due to focusing on single pathways where the small number of genes likely reduced inference power and increased the chance of confounding by factors such as expression level. The third advantage of this study is that by assembling all the Reactome database into a single matrix, including information from average adult expression levels, expression across tissues and developmental stages, GO functional terms, aggregate network connectivity and DNA sequence features (length, no of CDS, etc.) in a single model, we managed to dissect the effect of each parameter while controlling for confounding factors and using a large sample of genes ensuring statistical power.

52M Seasonal Selection across a Fine Temporal Timescale. *Alyssa Bangerter*, Alan Bergland Biology, University of Virginia, Charlottesville, VA.

Many organisms live in temporally heterogeneous environments and experience seasonal changes in selection pressure. There is both genetic and phenotypic evidence that temperate populations of *Drosophila melanogaster* cyclically adapt to seasonal change in selection pressure. To date, our understanding of rapid adaptation to seasonal change in selection pressure is primarily based on pooled allele frequencies from spring and fall samples. To further understand the role of

population structure and the dynamics of seasonal adaptation we sampled *Drosophila melanogaster* bi-weekly from a focal population in central VA and re-sequenced individual F1 male offspring of wild caught flies. By sampling across all four seasons we will analyze allele frequency changes across a full seasonal cycle. Additionally, using individual sequencing, we will be able to identify if there are any signs of population structure. This detailed dataset will also be put into the context of samples collected from Africa, Europe, and across a latitudinal gradient in North America in order to put signals of fine-scale seasonal adaptation into large-scale spatial adaptation.

53M CODIS loci STR's could contain more information than previously assumed. *Mayra Banuelos*[1], Emilia Huerta Sanchez[2], Rori Rohlfs[1] 1) San Francisco State University, CA; 2) University of California, Merced, CA.

The 13 CODIS Loci STRs (the Combined DNA Index System (CODIS) core loci) are short tandem repeats (STRs) adopted by the FBI around 1996 to be used as forensic identification markers. Back in 1996, full genome sequencing was not yet available. DNA sequencing was a very expensive and lengthy process so, when choosing the CODIS loci it was very important that the set of markers was easily sequenced, and that they contained enough genetic information to distinguish one individual from another. The markers selected for this purpose were chosen for being easily replicated via PCR with minimal amplification issues and, in addition, they demonstrated high genetic diversity which made them an ideal genetic source to differentiate individuals. 13 microsatellites were chosen and became the CODIS Loci STRs used today for forensics in the United States.

Up until now, we have assumed these genetic markers have evolved neutrally but, since they were established as forensic tools before whole genome sequencing was available, we haven't studied them in-depth to confirm that assumption. Also, because the primary use of these markers is individual identification, the CODIS Loci STRs are thought not to reveal any other genetic information (particularly medical information) outside from differentiating individuals. We know, however; that the change in length of STRs can alter phenotypes and expression levels in certain genes. Huntington's Disease (HD), for example, is caused by the variation in the length of an STR that changes the expression levels of the HD gene.

In this study, we isolate the 13 CODIS STRs, analyze the correlations between their lengths and individual gene expression levels. The data used for this study is publically available in the Rosenberg Lab at Stanford University. We utilized a genotype data file and a lymphoblast cell line transcriptome data file, both of which are typed on the Human Genome Diversity Project dataset. I was able cross-match 31 individuals with the genotype set.

Our initial study showed no standout patterns of correlation at the genome-wide level. Therefore, we decided to look at the relationships between the STRs and neighboring genes (100kb of distance). The preliminary comparisons suggest that the STRs may not be evolving neutrally after all but further analysis is needed. In the next phase of this study, we plan to expand the analysis to a larger set of data including a more diverse population set.

54M Population genetic structure of a UK *Daphnia pulex* meta-population using whole genome resequencing. *K. Barnard-Kubow*[1], D. Becker[1], A. Edwards[1], E. Voss[1], A. Beckerman[2], A. Bergland[1] 1) Department of Biology, University of Virginia, Charlottesville, VA, USA; 2) Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK.

Daphnia are keystone species in freshwater aquatic ecosystems and are a model for ecological and evolutionary research. *Daphnia*, along with some rotifers and aphids, are cyclic parthenogens, meaning they alternate multiple cycles of asexual reproduction with periodic episodes of sexual reproduction. This mode of reproduction has the potential to significantly influence the neutral population structure of *Daphnia* populations, making the detection of signals of selection or demography using traditional population genetic methods somewhat complicated. Much theory has focused on describing how the reproductive strategy of cyclic parthenogens can influence the population genetic structure. However, extensive characterization of the genetic structure of natural *Daphnia* meta-populations has been restricted to only a few studies, and rarely has the characterization included sequencing of entire genomes. In the current study, we carried out whole genome resequencing of several hundred individual clones, as well as pooled samples, collected across multiple time points from a meta-population of *Daphnia pulex* in the southern UK. For this analysis, we also constructed a new *D. pulex* reference genome using a European *D. pulex* clone. Our results demonstrate variation in species composition, as well as genetic diversity and clonal structure due to differing reproductive strategies, among ponds that are separated by only hundreds of meters. We find evidence for dominant clonal lineages co-existing over multiple seasons. We also find copy number variants among clones of *D. pulex* within a single population as well as between populations, demonstrating the dynamic nature of the *D. pulex* genome. These findings will be used to help establish baseline expectations of patterns of genome wide diversity in this system to assist in future work examining patterns and processes of selection acting in this system.

55M Genome-wide markers reveal a complex evolutionary history involving divergence and asymmetric patterns of introgression in the Abert's squirrel (*Sciurus aberti*) species group. *J.M. Bono*[1], H.K. Pigage[1], J.C. Pigage[1], P.J. Wettstein[2] 1) Dept of Biology, University of Colorado Colorado Springs, Colorado Springs, CO; 2) Dept of Surgery, Mayo Clinic, Rochester

MN.

Genetic introgression between divergent lineages is now considered more common than previously appreciated, with potentially important consequences for adaptation and speciation. Introgression is often asymmetric between populations and patterns can vary for different types of loci (nuclear vs. organellar), complicating phylogeographic reconstruction. The taxonomy of the ecologically specialized Abert's squirrel (*Sciurus aberti*) species complex has been controversial, and previous studies based on mitochondrial data have not fully resolved the evolutionary relationships among populations. Moreover, while these studies identified potential areas of secondary contact between divergent lineages, the possibility for introgression has not been tested. Here, we use RAD-seq to unravel the complex evolutionary history of the Abert's squirrel species group. Although some of our findings reinforce inferences based on mitochondrial data, we also find significant areas of discordance. Discordant signals generally arise from previously undetected introgression between divergent populations that differentially affected variation at mitochondrial and nuclear loci. Most notably, our results support earlier claims (disputed by mitochondrial data) that *S. aberti kaibabensis*, found only on the north rim of the Grand Canyon, is highly divergent from other populations. However, we also detected introgression of *S. aberti kaibabensis* DNA into other *S. aberti* populations, which likely accounts for the previously inferred close genetic relationship between this population and those south of the Grand Canyon. Overall, our results support the emerging view that populations often diversify under scenarios involving both divergence in isolation and gene flow during secondary contact, and highlight the value of genome-wide datasets for resolving such complex evolutionary histories.

56M Methods for detecting selection in admixed populations. *Erin Calfee*, Daniel Gates, Jeffrey Ross-Ibarra, Graham Coop University of California, Davis, CA.

New genetic variation from admixture has the potential to drive rapid adaptive change, at the time scale of ecological processes such as range shifts, competition and invasions. Recent progress has been made to accurately infer the mosaic of local ancestry across admixed genomes and identify outlier loci under selection. However, current methods to test for selection are not robust to drift, which limits our ability to analyze admixed populations with complex or unknown demographic histories. We show how shared evolutionary history post-admixture can be inferred from genome-wide patterns of ancestry variance and covariance. Analogous to a phylogenetic independent contrast, we correct for this background non-independence between sampled populations, mitigating false-positives. Our proposed statistical framework can be used to identify loci with an excess or deficit of a specific ancestry, or an association between ancestry and an environmental variable, beyond what can be explained by shared drift among sampled populations. We apply these methods to identify signatures of parallel adaptation in populations of admixed *Zea mays* across Mexico. This work adds fine-scale genomic resolution to previous findings that adaptive introgression from a highland-endemic wild relative, *mexicana*, facilitated maize's range expansion to colder and more UV-intense high altitude environments. The methods presented here can be readily applied to other systems to understand more generally how admixture shapes range shifts or invasions and how consistently specific loci adaptively introgress or maintain barriers to gene flow across different populations or environments.

57M Epigenetic regulation of aging in *Drosophila melanogaster*. *Q. B. Chen*, T. F. C. Mackay Department of Biological Sciences and Program in Genetics, North Carolina State University, Raleigh NC.

As the average lifespan of the world population continues to increase, deciphering the biological underpinnings of natural variation in aging and lifespan is critical to managing aging-related diseases. Recent studies have strengthened the heterochromatin loss model of aging: organisms exhibit a global loss of heterochromatin over time that results in the aberrant expression of silenced genes and the inability to maintain homeostasis. Here, we use five *Drosophila melanogaster* lines selected for postponed reproductive senescence for over 170 generations (O lines) and five lines from the same base population maintained without selection (B lines) to assess differential chromatin states between long lived and normal lifespan flies. The O lines have twice the lifespan compared the B lines at approximately 70 days and 35 days, respectively. We find that expression of transposable elements, normally silenced in heterochromatic domains, are higher in the B lines than the O lines at both young and old age. Additionally, expression of transposable elements in old O line flies are comparable to young B line flies. These findings suggest that the O and B lines can be utilized to decipher the genetic basis of global heterochromatin state maintenance and aging. We will use ATAC-seq to determine changes in open chromatin from a variety of tissues in both sexes of the O and B lines at one and five weeks and ChIP-seq to target histone modifications. In conjunction with previous genomic, transcriptomic, metabolomic, and phenotypic data, we will derive putative causal relationships between epigenetic modifications and natural variation in lifespan.

58M Inference of Admixture for Cattle with Complex Ancestry. *T. Crum*, R. Schnabel, J. Decker, J. Taylor Division of Animal Sciences, University of Missouri, Columbia, MO.

In many beef production systems, crossbreeding is used to take advantage of breed complementarity and heterosis. Whether the crossbreeding of two purebred individuals occurred recently or many generations ago can be difficult to establish if pedigree data are lacking, yet the additive effects of the foundation breeds contribute to the individual's

performance and to variation among animals for production traits. Estimating ancestry in admixed populations is required in order to control for population structure in genome-wide association studies where spurious associations can occur if breeds differ for a trait and also differ in allele frequencies at random sites throughout their genome. Approaches have been proposed for the estimation of local ancestry (breed of origin of alleles at specific loci) in admixed populations, however, these approaches generally focus on inferences in recently admixed populations. We tested an approach to estimate the global ancestry of individuals using ADMIXTURE and SNPWEIGHTS. ADMIXTURE estimates ancestry using a model-based approach applied to large SNP genotype datasets. The individuals are assumed to be unrelated and a supervised analysis can be performed using animals that have been predetermined to represent specific populations. SNPWEIGHTS infers ancestry using genome-wide SNP weights that are estimated using external reference panels. We constructed an analysis pipeline to determine the ancestry of cattle with potentially complex ancestries using both methods based upon a predetermined reference population dataset. The reference population was established using Breed Association pedigree information and iterations of analysis performed to identify a set of purebred individuals that represented each breed. The pipeline extracts genotypes for a set of 6,900 SNPs that are common to all commercially available high-density cattle SNP genotyping assays for an individual of unknown ancestry which are then processed into the correct file formats for the analysis software and are analyzed against the reference dataset to predict admixture proportions.

59M Population genomic signatures of selection during freshwater invasions by a saline copepod (*Eurytemora affinis*). *T. da Silva Ribeiro*[1], M. Bontrager[2], C. Lee[1,2,3] 1) Department of Integrative Biology, UW-Madison, Madison, WI; 2) Laboratory of Genetics, UW-Madison, Madison, WI; 3) Center of Rapid Evolution (CORE), UW-Madison, Madison, WI.

Invasive populations provide excellent model systems for studying genetic mechanisms underlying rapid adaptation in nature. In the past ~70 years, the copepod *Eurytemora affinis*, originally found in estuarine coastal habitats, has invaded freshwater habitats multiple times. Previous studies have documented the evolution of increased freshwater tolerance and ion transporter activity and expression in freshwater populations of *E. affinis* (under freshwater conditions), relative to their saline ancestors. The goal of this project is to explore population genomic signatures of selection associated with freshwater invasions, in order to detect candidate genes that possibly underlie rapid adaptation to freshwater habitats. Thus, we are investigating the presence of genomic signatures of selection associated with freshwater invasions by *E. affinis* in the Great Lakes, relative to ancestral saline populations in the St. Lawrence estuary. To achieve this goal, we performed whole genome sequencing of 100 pooled individuals per population from 4 populations from the St. Lawrence drainage system. Using windows of 100 SNPs each, we measured the decrease in nucleotide diversity (π) between saline and freshwater populations and Population Branch Excess (PBE), an F_{ST}-based metric comparing a freshwater population to two saline populations. The windows within the 99th percentile for each metric were considered candidate windows under selection in the freshwater habitat. In independent assessments for each lake, we found 150 candidate regions under selection in both lakes (shared candidate regions). Three annotated genes within these regions are related to ion uptake, namely, the Na^+, H^+ antiporter paralogs NHA-1 and NHA-5 and Na^+, K^+-ATPase (NKA) subunit alpha, which had shown evolutionary shifts in gene expression in our previous studies. Gene Ontology enrichment analyses revealed enrichment of many functions related to ion uptake but also unexpected functions such as sensory perception and development. Our results strongly suggest an important role for the evolution of ion transporter function during freshwater invasions and identify other biological functions that might have been crucial for such invasion. The study takes advantage of invasive populations to gain novel insights into rapid adaptation to radical environmental changes in nature.

60M Extensive negative linkage disequilibrium between amino acid changing variants suggests interference among deleterious mutations in the human genome. *J. Garcia*, K.E. Lohmueller University of California, Los Angeles, Los Angeles, CA.

Although the theoretical foundation for the expected value of linkage disequilibrium (LD) between two neutral loci has been established, it is not clear how LD theory extends to large numbers of interdigitated variants affected by natural selection. Forces like negative epistasis, where haplotypes carrying multiple deleterious mutations are less fit than predicted by the marginal fitnesses of both loci, are expected to lead to deleterious mutations being found on distinct haplotypes. This creates an excess of negative LD between derived deleterious alleles. Additionally, Hill-Robertson interference can create an excess of negative LD between pairs of deleterious variants under purifying selection. However, the extent to which these forces shape genome-wide patterns of LD remains elusive. In this study, we aim to assess how various types of selection (direct effects of purifying selection and background selection) on multiple linked loci jointly affect summary statistics of LD in finite diploid populations. First, we used forward in time simulations to examine the extent at which selection might impact these patterns of LD. Our simulations included realistic intron-exon structures, a distribution of fitness effects, recombination, and a mutation rate resembling that of humans. The simulations show that deleterious variants less than 10 kb apart tended to be carried on different haplotypes, generating an excess of negative LD as measured by D and D'. Also, nearby variants that are not under direct selection, but possibly affected by background selection, appear to have a decreased r^2, D, and D' when the

distance between them is less than 10 kb. However, pairs of these neutral variants over 10 kb apart appear to have these LD statistics slightly elevated compared to their counterparts in simulations where no selection is present. We then analyzed human genetic variation data from the 1000 Genomes Project. We find that pairs of derived nonsynonymous variants have, on average, more negative LD between them compared to pairs of derived synonymous variants, even when matching for allele count, physical distance, and genetic distance between pairs of variants (P

61M Array design and SNP ascertainment bias. *J. Geibel*[1], S. Weigend[2], A. Weigend[2], C. Reimer[1], T. Pook[1], H. Simianer[1] 1) University of Goettingen, Department of Animal Sciences, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen, Germany ; 2) Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics, Neustadt-Mariensee, Germany.

SNP ascertainment bias plays an important role in population genetics and genomics. This study aims at elucidating the impact of different steps in the array design on the magnitude of the bias. For this purpose, the design of the Axiom™ Genome-Wide Chicken Array (www.affymetrix.com) was remodeled using a broad set of four individually and 42 pool sequenced chicken populations. Estimates for the allele frequency spectra and the expected heterozygosity from SNPs called after the different steps of the array design were contrasted against the estimates from whole genome sequences and the original array. Array design was also repeated with random assignment of populations to discovery, validation and application groups and varying parameters. Two steps of the array design were found to have major influences on the amount of bias: (i) the discovery of SNPs in a small number of populations and (ii) the reduction of the SNPs to achieve equal spacing of SNPs over the genome by genetic distances. It also could be shown, that there is a more pronounced (~30 %) overestimation of the expected heterozygosity in the populations used for the discovery of the SNPs, compared to other populations. Overestimation was decreased drastically when the number of populations for SNP discovery was increased and focus during the equal spacing step was rather on achieving equal spacing of SNPs over the genome than to aim at a larger proportion of common SNPs as a backbone. Despite overall ascertainment bias, Pearson correlations between whole genome sequence based and array based expected heterozygosities range from 0.92 to 0.97.

62T Genome-wide characterization of differences in mutation fitness effects between populations. Alyssa Fortier[1], Alec Coffman[1], Travis Struck[1], Jose Burguete[2], Aaron Ragsdale[3], PingHsun Hsieh[4], *Ryan Gutenkunst*[1] 1) Molecular and Cellular Biology, University of Arizona; 2) Center for Genomic Sciences, National Autonomous University of Mexico; 3) Applied Mathematics, University of Arizona; 4) Ecology and Evolutionary Biology, University of Arizona.

The fitness effect of a mutation may differ between populations, depending on environmental and genetic context. To quantify genomic patterns of such differences, we extended the concept of a distribution of fitness effects (DFE) to a joint DFE between populations. To infer the joint DFE, we fit parametric models that included demographic history to genomic data summarized in the joint allele frequency spectrum. We applied this framework to African and European populations of both Drosophila and humans, finding that mutation fitness effects are much more similar between populations of humans than Drosophila. Among gene sets, genes involved in immunity typically showed low similarity of fitness effects, whereas genes involved in reproduction showed high similarity. Our results represent the first genome-scale quantification of mutation fitness effect differences between populations and point toward gene functions that are more likely to experience divergent selection.

63M Inferring global population structure and genetic diversity in wild house mice (*Mus musculus*) using a genome-wide SNP array. *Jonathan Hughes*[1], Andrew Morgan[2], John Didion[2], Jeremy Searle[1], Fernando Pardo-Manuel de Villena[2] 1) Ecology and Evolution, Cornell University, Ithaca, NY; 2) Department of Genetics, University of North Carolina, Chapel Hill NC, USA.

The house mouse *Mus musculus* in its domestic form is the mammalian biological model: the laboratory mouse. The wild form of *Mus musculus* and its subspecies are also key models for mammalian evolution, benefitting from research tools developed for laboratory mice. However, the particular origin and history of laboratory mice must be accounted for when applying such tools to wild mice. We discuss how to control for such features when using a 77,000 SNP genotyping array designed for laboratory mice on wild populations, such as using runs of homozygosity in the genome to estimate inbreeding coefficients. We use the genotyping array and available whole genome sequencing data to provide information on nearly four hundred wild house mice from across the world, and examine variation of heterozygosity, inbreeding, and admixture in wild mice, in relation to geography and subspecies, as a baseline for future evolutionary studies. We observe markedly higher nucleotide diversity in *M. m. castaneus* compared to other subspecies. Inbreeding levels show limited difference between inland and coastal populations, while island populations are more inbred than mice on the mainland. Furthermore, we note distinct geographic patterns of heterozygosity in European mice. *M. m. domesticus* and *M. m. musculus* populations in Europe and North America show substantial admixture. Genome-wide SNP arrays are valuable tools for studies of wild mice, when used with care for possible biases.

64M Population Genetic Analysis of Autophagy and Phagocytosis Genes in *Drosophila melanogaster* and *simulans*. *J. Im*[1,2], B. Lazzaro[1,2] 1) Cornell Institute of Host-Microbe Interactions and Disease, Cornell University, Ithaca, NY; 2) Department of Entomology, Cornell University, Ithaca NY.

Dynamic conflict between hosts and pathogens can result in a co-evolutionary adaptation in host genes. Autophagy and phagocytosis are cellular immune mechanisms for internalization and elimination of intracellular and extracellular pathogens. Some pathogens have evolved to inhibit or evade these processes, raising the prospect of adaptive reciprocal co-evolution by host autophagy and phagocytosis genes. To test for molecular evolutionary signatures of adaptation, we performed population genetic analyses on phagocytosis and autophagy genes in *Drosophila melanogaster* and *D. simulans*. We found signatures of recent and recurrent positive selection across several phagocytic receptor genes, as has been previously reported. We additionally discovered that two glutamate transporters involved in phagocytosis (*gb* and *polyph*) show a pattern of elevated sequence divergence between *D. melanogaster* and *D. simulans* while two genes that facilitate autophagy (*Atg14* and *Atg8b*) display signature of recent selection in each species.

65M Identifying recurrent mutations from unphased population-level sequencing data. *K.E. Johnson*[1], B.F. Voight[1,2,3] 1) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA; 2) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Recurrent mutations are a hallmark of Mendelian and complex disease. Studies identifying genes enriched for recurrent mutation have improved our understanding of the genetic basis of disease. Furthermore, as population resequencing continues to expand and we approach mutation saturation, accounting for recurrent mutations will help to improve estimates of the site frequency spectrum from these data. Given the importance of recurrent mutation in health and evolutionary studies, we require an approach to discover recurrent mutations in genetic data that lacks familial relationships. Here, we present a method to infer recurrent mutations in population-level sequencing data without the need for pedigree data or phased haplotypes. The key intuition underlying the method is the observation that the time to the recent common ancestor (TMRCA) differs between recurrent and identical-by-descent (IBD) mutations. As a summary statistic for the local TMRCA around a site, we measure the 'obligate recombination distance': the distance to the nearest opposite homozygote genotype on either side of the target allele for each pair of carriers. We can calculate the likelihood of this collection of pairwise distances under IBD or recurrent mutation scenarios, and identify alleles whose measurements are inconsistent with recent IBD as candidate recurrent mutations. Simulation studies indicate reasonably powered performance of our approach. The performance is dependent on the allele count of the target allele, with improved detection of rarer recurrent variants. We applied our method to SNVs and indels in human whole-genome sequencing data from the UK10K project. The putative recurrent mutations we identify are enriched at CpG sites, consistent with the known elevated rate of mutation at CpGs relative to other sequence contexts. Future applications of this method include incorporating recurrent mutation into tests of rare variant burden in disease, and interrogation of the distribution of fitness effects through improved inference of the site frequency spectrum at rare frequencies.

66M Estimating the contemporary effective population size of *Hypomesus transpacificus* using RAD sequence data. *SEK. Joslin*[1], A. Finger[2] 1) Integrative Genetics and Genomics, University of California, Davis, CA; 2) Department of Animal Science, University of California, Davis, CA.

Hypomesus transpacificus (Delta Smelt) is a small (5-7cm), translucent, panmictic species of Osmeridae fish endemic to the San Francisco Estuary (Delta) in California. A once plentiful fish in the Delta, the species has been listed as federally threatened since 1993. Drought and anthropogenic effects are associated with the rapid collapse of the Delta Smelt population and it is now feared that stochastic processes could push the species to extinction. Management strategies include the maintenance of a conservation hatchery; restoration efforts to improve habitat conditions and vital rates; and monitoring Delta Smelt population size by conducting regular surveys to estimate census size via physical capture. An accurate estimation of effective population size (Ne) will provide a useful component for conservation efforts aimed at maintaining genetic diversity within the wild Delta Smelt population. Previous estimates of Ne used 12-15 microsatellite markers, have wide confidence intervals and the small number of markers leaves entire linkage groups of the genome unexamined. Here, we capitalize on the recent reimplementation of NeEstimator v2.1 to estimate Ne in a non-model organism. We use 17 generations (1993-2014) of restriction site-associated DNA sequencing (RAD sequencing) data cut with Sbf1 to estimate Ne using a two-generation temporal method (Ne_T) and a single-generation bias-corrected linkage disequilibrium method (Ne_{LD}). We believe these estimates to have greater power to more precisely estimate the effective population size of Delta Smelt to better inform the management and conservation practices in the Delta.

67M Limited evidence for selection at linked sites in wild populations of *Mus musculus domesticus*. *Michael Kartje*[1], Peicheng Jing[1], Bettina Harr[2], Diethard Tautz[2], Bret Payseur[1] 1) Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; 2) Max Planck Institute for Evolutionary Biology, Plon, Germany.

Selection acting on deleterious mutations or beneficial mutations can reduce levels of genetic variation at linked neutral sites. Both background selection and genetic hitchhiking predict a positive correlation between neutral diversity and recombination rate across the genome. Although this relationship is observed in many species, the magnitude of the correlation varies, suggesting differences in the extent of linked selection among groups. We compared nucleotide diversity and recombination rate across the genome of the Western European house mouse (*Mus musculus domesticus*). We used whole genome sequences constructed by aligning Illumina reads to the mouse reference genome to estimate neutral diversity in island and mainland populations. We estimated recombination rates (cM/Mb) across the genome independently of the population genomic data using the mouse reference genetic map, which is based on 3,546 meioses and 10,195 informative SNPs. We also considered potential covariates, including the density of functional sites along the genome. In both mainland and island populations, we observed significantly positive correlations between nucleotide diversity and recombination rate ($p << 0.01$), but correlation coefficients were near zero (Spearman's rho < 0.02). We conclude that linked selection does not strongly shape genome-wide patterns of neutral variation in Western European house mice.

68M Two opposite haplotypes that regulate *FADS1* expression underwent recurrent diet-dependent adaptation in worldwide human populations. *A. Keinan*[1,2,3,4], D. Wang[1], K. Ye[1] 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Cornell Center for Comparative and Population Genomics, Cornell University, Ithaca, NY; 3) Center for Vertebrate Genomics, Cornell University, Ithaca, NY; 4) Center for Enervating Neuroimmune Disease, Cornell University, NY.

Biosynthesis of the functional, long-forms of omega-3 and omega-6 relies on fatty acid desaturases (*FADS*). While long-forms are prevalent in animal-based diets, biosynthesis is essential when provided with shorter-forms that are common in plant-based diets. Adaptation has been reported for several different *FADS* variants in different human populations. We recently characterized a long, 85 kb region that extends through the entirety of *FADS1* and most of *FADS2*, with two common haplotypes (M2 and D) on which SNPs have opposite alleles. We established that these haplotypes have been adaptive in pre-Neolithic (haplotype M2) and post-Neolithic (haplotype D) Europeans. SNPs on the haplotypes are eQTLs for *FADS1* such that haplotype D increases (and M2 decreases) the biosynthesis, consistent with the dramatic switch to opposite adaptation with the Neolithic transition from mostly meat-based diets of hunter-gatherers to mostly plant-based diets of farmers. Here, we investigate adaptation of *FADS* genes and the role of these haplotypes in global populations. Specifically, we 1) tested whether previously reported adaptation is consistent with these haplotypes, 2) evaluated a recent claim that positive selection must have operated on different *FADS* variants in different populations, and 3) separated ancestral from recurrent adaptation. We have put together a broad set of extant populations from many sources along with ancient human populations based on data from 25 studies of ancient DNA and have imputed all to a uniform set of variants. We applied tests for selection of three types: various tests based on extant populations that capture positive selection at different time scales; allele frequency shift from ancient to extant populations; and, uniquely, allele frequency time series in ancient samples restricted to a period of interest. We conclude that haplotype D has been adaptive in most populations that traditionally practiced farming, resulting in high frequency in South Asians (75-90%), Europeans (53-71%), Northeast Asians (65-66%), and five African farmer populations (35-48%). Similarly, M2 or an almost identical haplotype (M1) has been adaptive in hunter-gatherer Eskimos from East Siberia (100%), Andamanese (97%), and pre-Neolithic Europeans (40% ~30 kya increased gradually to 67% just before the Neolithic). Our previous results of adaptation in South Asians that focused on an indel are consistent with adaptation of D, which carries the insertion, with the deletion being on M2. Analyzing the indel with inaccurate genotype calls in other studies has contributed to discrepancies in the literature. Combined, the results point to selection on standing variation, unify and generalize all previous studies of *FADS* adaptation, and reveal that adaptation is consistent with the same variant\s being the target of selection in all studied populations.

69M Fine-scale localization of sites under balancing selection. *Evan Koch*, Mohammad Siddiq Ecology and Evolution, University of Chicago, Chicago, IL.

A central goal in evolutionary biology is to understand how genetic variation is maintained in populations. Balancing selection—the name given to any scenario where selection acts to maintain genetic variation—is one mechanism through which genetic variation in populations may be maintained. Balancing selection is theoretically possible under a variety of conditions that may exist naturally and has been a process of interest in evolutionary genetics for as long as the discipline has existed. Despite considerable historical interest, relatively few decisive cases of balancing selection have been identified, much less understood functionally. Recent population genomic studies that have tracked genomic studied natural populations across spatial and seasonal habitats however suggest that balancing selection may be much more widespread than previously appreciated. Nevertheless, two major challenges have hindered the identification and empirical investigation of specific alleles undergoing putative balancing selection. First, signatures of balancing selection are often subtle and remain relatively difficult to identify in genome-wide scans. Second, even when there is a detectable signature of selection on a region that has been affected by balancing selection, the signal is often dispersed across a large number of sites with an impractically high number of candidate mutations for functional validation. Here, we investigate how these challenges may be ameliorated by using a supervised machine learning approach to identify windows and specific sites evolving under balancing

selection. We train our models using simulations and a larger set of summary statistics than could be examined manually. We then test the accuracy and precision of these models to identify the nucleotide site under selection in simulated data, and apply our analysis to genomic regions of *Drosophila* that are thought to harbor balanced alleles. In so doing, we show a way in which existing statistics for evolutionary analysis can be efficiently combined so that historical evolutionary processes leaving only subtle signatures in modern genomes can be better identified and understood. In particular, the procedures we develop here often suggest a small enough set of potentially selected sites that it is feasible to functionally characterize individual mutations.

70M From soil to stein; population genomics of wild and domesticated lineages of the Lager-brewing ancestor; *Saccharomyces eubayanus*. *Q. Langdon*[1], D. Peris[1,2,3], K. Buh[1], R. Moriarty[1,2], K. Sylvester[1,6], J. Eizaguirre[4], C. Lopes[5], D. Libkind[4], C.T. Hittinger[1,2] 1) Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Genome Center of Wisconsin, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI USA; 2) DOE Great Lakes Bioenergy Research Center, University of Wisconsin¬Madison, Madison, WI USA; 3) Department of Food Biotechnology, Institute of Agrochemistry and Food Technology (IATA), CSIC, Valencia Spain; 4) Laboratorio de Microbiología Aplicada, Biotecnología y Bioinformática, Instituto Andino Patagonico de Tecnologías Biológicas y Geoambientales, IPATEC (CONICET-UNComahue), Centro Regional Universitario Bariloche, Bariloche, Río Negro, Argentina; 5) Laboratorio de Microbiología Aplicada, Biotecnología y Bioinformática, Instituto Andino Patagonico de Tecnologías Biológicas y Geoambientales, IPATEC (CONICET-UNComahue), Centro Regional Universitario Bariloche, Bariloche, Río Negro, Argentina; 6) Department of Molecular Genetics and Microbiology, Duke University School of Medicine, NC 27708.

Human society has harnessed the awesome power of yeast genetics for the production of food and beverages for far longer than we have known of either microbes or genetic inheritance. Yet, it is only recently have we begun to explore their genetic diversity, both in industrial and natural settings, and understand their evolution within and interaction between differing environments. Hybrids of *Saccharomyces cerevisiae* and *Saccharomyces eubayanus* have been used to make lager beer for hundreds of years, but it was less than a decade ago that we discovered the wild pure stock of *S. eubayanus*. Through whole genome sequencing of over 100 strains of *S. eubayanus,* we are exploring the extensive natural diversity not found in industrial settings. With this considerable collection we are investigating the population structure and demographic history of this recently discovered species. Within Patagonia, where *S. eubayanus* is frequently isolated, there are two distinct and diverse populations. Isolates outside of Patagonia are rare, but the strains that have been found offer a unique view of the spread of admixed lineages and the origin of brewing hybrids. A handful of strains represent a Holarctic sub-population that has been found in North Carolina, Tibet, and in the *S. eubayanus* sub-genome of lager-brewing yeast. Yet, with this limited pool we have shown that no extant strain is the closest relative of lager-brewing yeast, but rather the standing genetic variation found in this wild Holarctic lineage still persists in the two lineages of lager-brewing yeast. As we further explore this dataset we can understand both variation in the numerous South American strains and in other *Saccharomyces* hybrids found in industrial settings.

71M Impacts of recurrent hitchhiking on divergence and demographic inference in *Drosophila*. *J. Lange*, J. Pool Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI.

In species with large population sizes such as *Drosophila*, natural selection may have substantial effects on genetic diversity and divergence. However, the implications of this widespread non-neutrality for standard population genetic assumptions and practices remain poorly resolved. Here, we assess the consequences of recurrent hithchhiking (RHH) in which selective sweeps occur at a given rate randomly across the genome. We use forward simulations to examine two published RHH models for *D. melanogaster*, reflecting relatively common/weak and rare/strong selection, respectively. We find that unlike the rare/strong RHH model, the common/weak model entails a substantial degree of Hill-Robertson interference, which has implications for the rate of beneficial mutation and for the simulation of RHH models. We also find that the common/weak RHH model is more consistent with our genome-wide estimate of the proportion of substitutions fixed by natural selection between *D. melanogaster* and *D. Simulans* (19%). Finally, we examine how these models of RHH might bias demographic inference. We find that these RHH scenarios have relatively minor effects on the inference of recent between-population parameters. Thus, even for species with important genome-wide impacts of selective sweeps, neutralist demographic inference can have some utility in understanding the histories of recently-diverged populations.

72M Conserved patterns of somatic mutations in human blood cells. *L A Liggett*[1,2], James DeGregori[1,2,4,5,6], Anchal Sharma[3], Subhajyoti De[3] 1) Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO; 2) Linda Crnic Institute for Down Syndrome, University of Colorado School of Medicine, Aurora, CO; 3) Rutgers Cancer Institute, New Brunswick, NJ; 4) Integrated Department of Immunology, University of Colorado School of Medicine, Aurora, CO; 5) Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO; 6) Department of Medicine, Section of Hematology, University of Colorado School of Medicine, Aurora, CO.

It is well known that mutation rates vary by chromosomal location, nucleotide identity, and sequence context, but as somatic mutations are typically rare, measuring them within healthy tissues has been challenging. As a result, most studies

have been limited to studying somatic mutation within clonal expansions, resulting in a deficient understanding of mutation frequencies and patterns within benign tissues. Using a novel method capable of accurately detecting mutations with single base pair resolution at allele frequencies as rare as 10-4, we find a surprisingly high somatic mutation burden in peripheral blood cells from apparently healthy individuals. Our observed somatic mutation burden is high enough that nearly all analyzed sites carry at least one somatic mutation (including known oncogenic mutations) in ~20,000 cells. Surprisingly, mutation patterns and individual variant allele frequencies are remarkably conserved between individuals in an age-independent manner. As a result, a given mutation, whether oncogenic or not, will exist at similar levels within most individuals. In contrast to the general phenomenon, we have identified two individuals with patterns of somatic mutation that resemble a mismatch repair deficiency, exhibiting mutations that exist at uniformly elevated mutation frequencies. These results demonstrate that somatic mutations, including oncogenic changes, are far more common in healthy human tissue than previously appreciated, and suggest an unappreciated degree of non-randomness within the processes underlying mutagenesis.

73M Phenotypically wild barley plants show evidence of extensive introgression from cultivated barley. *C. Liu*[1], L. Lei[2], C.M. Depies[2], B.J. Steffenson[3], P.L. Morrell[2] 1) Plant & Microbial Biology, University of Minnesota, Saint Paul, MN; 2) Agronomy & Plant Genetics, University of Minnesota, Saint Paul, MN; 3) Plant Pathology, University of Minnesota, Saint Paul, MN.

A number of studies have reported genotypic evidence that phenotypically wild accessions of wild barley (Hordeum ssp. spontaneum) have been subject to introgression from cultivated barley. We use 318 total accessions from the Wild Barley Diversity Collection (WBDC), which is part of the extensive collections of wild barley accessions that are maintained for their utility in crop improvement. With comparable genotype and exome capture resequencing data, we make use of identity by state comparisons between the 318 WBDC accessions and cultivated barley to identify genomic regions that appear to have been subject to introgression from cultivated barley. We found multiple WBDC accessions show evidence of introgression. Using the genomic intervals for well characterized genes involved in domestication and improvement, we examine evidence for introgression at genomic regions potentially important for maintaining a wild phenotype. We also assess the size of runs of identity by state, with early evidence suggesting that most, but not all introgression has occurred very recently.

74M Human prehistoric demography revealed by polymorphic pattern of CpG transitions. *X. Liu* Human Genetics Center, UTHealth School of Public Health, Houston, TX.

Prehistoric demography of human populations is one essential piece of information to illustrate our own evolution. Despite its importance, our knowledge is very limited, even for the relatively recent population dynamics during and around the Holocene. Here we inferred demographic histories from 1 to 40 thousand years ago (kya) for 24 population samples, using a newly developed model-flexible method with 36 million non-coding CpG sites genome-wide. Our results show many population growth events likely due to the Neolithic Revolution[1] (i.e. shifting from hunting and gathering to agriculture and settlement): Han Chinese experienced a dramatic ~10 fold population growth around 8 kya to 12 kya; some South European and South Asian populations also began their long-term population growth around 10 kya but in a more gradual fashion; and British and Western European populations began their takeoff around 6-7 kya. The potential agriculture-associated population growth of Luhya in Webuye, Kenya (LWK) came relatively late (no earlier than 3 kya) compared to other African populations, in contrast to the assumption that they are the direct descendants of Bantu-speaking immigrants from West Africa. We also observed several population growth events dated before the introduction of agriculture. Our results help to paint a clearer picture of human's prehistoric demography, confirming the significant impact of agriculture on population expansion, and provide new hypotheses and directions for future research.

75M Gene expression drives the evolution of dominance. C. Huber[1], A. Durvasula[2], A. Hancock[3], *K. Lohmueller*[1,2] 1) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA; 3) Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany.

Dominance is a fundamental concept in molecular genetics and has implications for understanding patterns of genetic variation, evolution, and complex traits. However, despite its importance, the degree of dominance has yet to be quantified in natural populations. Here we develop a novel composite likelihood approach that leverages genetic variation data from outcrossing and selfing species to co-estimate the distribution of selection coefficients (*s)* and the dominance coefficient (*h)*. Since selection acts immediately on recessive homozygotes in self-fertilizing organisms, the genetic variation data from a selfing species allows us to discriminate between different values of *h.* Application of our approach to amino acid changing mutations in *Arabidopsis* suggests that most mutations are recessive and that more deleterious mutations tend to be more recessive than less deleterious mutations. We next use our data to test the existing models for the mechanism of dominance. For example, Fisher's model suggests that dominance arose via modifier mutations at other loci and that these loci are subject to selection. Wright, and later Kacser and Burns, proposed a metabolic theory model where mutations in enzymes are

predicted to be recessive because the overall flux through a metabolic network is robust to decreasing the amount of one of the enzymes of the pathway by one-half. We find that neither of these models for the evolution of dominance can explain how the inferred relationship between *h* and *s* varies with gene expression level and connectedness of genes. Thus we develop a new model for the evolution of dominance. Our new model predicts that dominance arose as a consequence of the functional importance of genes and their optimal expression levels. Our model matches many of the salient features of the data.

76M Effects of unsampled "ghost" populations on estimation of evolutionary history. *M. Lynch*, A. Sethuraman Biological Sciences, California State University San Marcos, San Marcos, CA.

Signatures of gene flow from unsampled "ghost" populations have been known to affect the estimates of evolutionary history from population genomic data. The presence of "ghost" populations is often not acknowledged in population genetic studies, which could potentially lead to erroneous conclusions about the demographic history of species, unless explicitly accounted for in a model-based framework. This study aims to address: (1) how does the presence of unsampled "ghost" populations affect genomic variation in sampled (observed) populations?, and (2) can accounting for these "ghost" populations improve estimation of evolutionary history? Specifically, we evaluate the performance of commonly utilized summary statistics (heterozygosity, polymorphism, allele frequencies, Fst and Dxy), and demographic parameter estimation under the Isolation with Migration (IM) model using IMa2p (Sethuraman and Hey 2015). We use extensive simulations under an IM model under five scenarios of varying levels of bi- or uni-directional gene flow from "ghost" populations (as described under an island model by Beerly 2001), and empirical data from African Hunter-Gatherer populations (Hadza and Sandawe, sensu Lachance et al. 2012) to understand potential bias and accuracy of estimates, and discuss ways to reduce the error of estimation. Preliminary analyses show that estimates of effective population sizes, and migration rates show increased bias and greater variance with increased levels of migration from the "ghost" population into the sampled populations. Additionally, IM analyses in African Hunter-Gatherer populations indicate greater support for a model with gene flow with a large, unsampled "ghost" population.

77M *Brassica oleracea*: The dog of the plant world. *M.E. Mabry*[1], E.Y. Gallagher[1], S.D. Turner[1], E. Katz[2], G. Ziegler[3,4], I. Baxter[3,4], D.J. Kliebenstein[2], M.A. Gore[5], J.A. Labate[6], J.C. Pires[1] 1) Division of Biological Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO; 2) Department of Plant Biology, University of California Davis, Davis, CA; 3) Donald Danforth Plant Science Center, St. Louis, MO; 4) United States Department of Agriculture–Agricultural Research Service (USDA-ARS), Plant Genetics Research Unit, St. Louis, MO; 5) Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY; 6) USDA-ARS, Plant Genetic Resources Unit, Cornell University, Geneva, NY.

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

78M The evolutionary history and population structure of the *Drosophila nasuta* clade. *Dat Mai*, Matt Nalley, Doris Bachtrog Integrative Biology, University of California Berkeley, Berkeley, CA.

The *Drosophila nasuta* clade is an under-examined species group comprising 11 to 14 young species with divergence times ranging from 120 KYA to 4 MYA, which provides an opportunity to study the early processes of speciation. To elucidate the evolutionary history and population structure of the clade, we first developed a high-quality genome assembly of *D. albomicans*, one of the species in the *D. nasuta* clade.

We performed single molecule, real-time sequencing and Hi-C sequencing, and created a set of contigs using minimap and miniasm and used Juicer and 3D-DNA to scaffold the contigs with the Hi-C reads. The genome was annotated with Maker. This resulted in a high-quality genome assembly consisting of 10 scaffolds with an N50 value of 33,427,555 and BUSCO assessment suggests that 98.1% of the gene set is complete.

To study the evolutionary history and population structure of the *D. nasuta* clade, we sequenced 68 individuals across all species to at least 8x coverage. The reads were mapped to the generated *D. albomicans* assembly with BWA and SNPs were called using GATK. Clustering analyses on these SNPs, i.e. principal component analysis and structure analysis, were able to group together members of the same species. We then developed a phylogeny using RAxML and Astral to infer the evolutionary relationship between all individuals. To estimate the relative divergence time between species, we calculated ks

values between strains of different species and normalized by the ks between *D. albomicans* and *D. nasuta*. We recapitulated relative divergence values similar to ratios of previously published absolute divergence times.

The produced genome assembly and inferred evolutionary history of the *D. nasuta* clade provide a foundation for future studies in this non-model system.

79M Long-Term Population Dynamics of 'Candidatus *Accumulibacter phosphatis*' in Enhanced Biological Phosphorus Removal Sequencing-Batch Reactors. *E.A. McDaniel*[1], F. Moya[2], P. Camejo[2], K.D. McMahon[1,2] 1) Department of Bacteriology, University of Wisconsin-Madison, Madison, WI; 2) Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI.

Enhanced biological phosphorus removal (EBPR) is an environmentally and economically significant process for prevention of eutrophication in freshwater sources. In wastewater treatment systems, specific microbes known as polyphosphate accumulating organisms (PAOs) are responsible for taking up excess inorganic phosphate and storing it as polyphosphate in their biomass. The most dominant PAO in activated sludge is the bacterial lineage 'Candidatus *Accumulibacter phosphatis'*. Investigations using comparative genomic and metagenomic approaches have revealed key metabolic, physiological, and genetic features of the Accumulibacter lineage. However, the temporal formation of the Accumulibacter population structure into similar but diverging subtypes and clades has remained an outstanding question in the context of phosphorus-uptake productivity and ecological significance. Here, we analyze a 9-year metagenomic time-series of EBPR sequencing-batch reactors enriched for Accumulibacter to understand the evolutionary forces governing the dynamics in population structure over time. Additionally, we combined 16S rRNA amplicon sequencing and qPCR of the polyphosphate kinase (*ppk1*) gene to understand the role of flanking microbial members on the abundance of the two main clades of Accumulibacter detected in our reactors. By studying the population structure of the Accumulibacter lineage through time, we aim to make inferences about the ecological significance of the distinguishing subgroups in the context of physiological differences and overall reactor performance.

80M The ash dieback invasion of Europe highlights huge adaptive potential of the causal fungus, *Hymenoscyphus fraxineus*. *Mark McMullan*[1], Maryam Rafiqi[2], Gemy Kaithakottil[1], Bernardo Clavijo[1], David Swarbreck[1], Neil Hall[1], Matt Clark[1], Nornex 1) Earlham Institute, Norwich, UK; 2) Royal Botanic Gardens, London, UK.

The changing environment and accelerating international trade make pathogen spread an increasing concern. Pathogen invasions establish founding populations with reduced genetic diversity but subsequent invasions are genetic. Here, I highlight the importance of understanding pathogen adaptive potential in invasive as well as native ranges. I use the Ash Dieback invasion of Europe to exemplify this and then consider the impact of this process in agricultural pathogens.

Hymenoscyphus fraxineus is the causal agent of ash dieback, a disease to which European common ash (*Fraxinus excelsior*) trees are highly susceptible. The fungus invaded Europe around 20 years ago from Asia and since has moved from Eastern to Western Europe. We have assembled and annotated a draft of the *H. fraxineus* genome which approaches chromosome scale. By re-sequencing 58 isolates of *H. fraxineus* from across its native (Asian) and invasive (European) ranges we find a tight founder effect impacts pathogen genetic diversity across Europe. Allelic divergence at Core Eukaryotic Genes show that the genetic diversity present in the European population is representative of just two divergent haploid individuals and yet, Ash Dieback has the potential to infect and kill 95% of all European ash trees. This disparity between polymorphism and success is known as the genetic paradox of biological invasions and examples of these are common.

Given that the *H. fraxineus* fungus is already prevalent throughout Europe is there any point in continued restrictions to trade? Analysis of European isolates suggests that the source population was large and equivalent to the observed genetic diversity in the Far East. Moreover, adaptive diversity is further preserved in the genes putatively responsible for interacting with the host. Continued invasion (gene flow) from the native range is therefore expected to increase the severity of the disease threat and so, continued restrictions to gene flow are advisable.

Finally, I highlight the importance of these processes in driving pathogen adaptation within wild and agricultural settings to argue that, in order to understand a pathogen's potential, we should consider its genetic diversity in both invasion and source ranges.

81M Estimating the prevalence of Wolbachia across arthropod and nematode taxa. *Paloma Medina*, Jackie Rogers, Russel Corbett-Detig University of California Santa Cruz, Santa Cruz, CA.

BACKGROUND: *Wolbachia* is emerging as a powerful biological control agent for disease vectors. Studies on *Wolbachia* may contribute to public health efforts to eradicate Zika, dengue, and chikungunya viruses. Though researchers have estimated *Wolbachia's* prevalence in specific taxa, like Drosophila and filarial nematodes, the infection frequency across diverse taxa remains largely unknown. *Wolbachia* can be transmitted both vertically, through host eggs, and horizontally both within species and across species boundaries. Here, we quantify the prevalence of *Wolbachia* horizontal transmission and live

infections in Arthropods and Nematodes.
METHODS: Using the NCBI SRA Database, we developed a bioinformatic method to survey more than 10,000 nematode and 40,000 arthropod samples for the presence *Wolbachia*. We used BLAST to find reads potentially originating from *Wolbachia* live-infections, and we identified horizontal gene transfer events using a k-mer counting approach.
RESULTS: Of the nematode and arthropod species sampled, we found that 1% and 22%, respectively, contained live *Wolbachia* infections. Moreover, our results corroborate a previously reported horizontal gene transfer event in the evolutionary history of *Dictyocaulus viviparus*.
SIGNIFICANCE: Our bioinformatic method is useful to detect live infections and horizontal gene transfer events from sequence data. Moreover, studying the diversity of *Wolbachia* live infections has the potential to reveal the evolutionary history of *Wolbachia* biology. With a greater understanding of *Wolbachia* species history, our work contributes to the global effort to understand and use *Wolbachia* as an effective control agent for infectious diseases.

82M Population genomics of copy number variation in a natural population of maize wild relative teosinte. *W. Mei*[1,2], S. Renny-Byfield[1], A. Lorant[1], A. S. Seetharam[3], J. Doebley[4], M. B. Hufford[5], J. Ross-Ibarra[1,2,6] 1) Department of Plant Sciences, University of California, Davis, Davis, CA; 2) Center for Population Biology, University of California, Davis, CA, USA; 3) Genome Informatics Facility, Iowa State University, Ames, USA; 4) Department of Genetics, University of Wisconsin-Madison, WI, USA; 5) Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, USA; 6) Genome Center, University of California, Davis, CA.

Understanding genetic diversity in natural populations has long been a central theme for biologists. However, so far most efforts have focused on single nucleotide polymorphisms to characterize genetic diversity, demography, introgression and gene flow. Yet considerable evidence suggests an important evolutionary role for structural variation in shaping genetic and phenotypic diversity. Here, we use high depth whole genome sequencing data from a single natural population of the maize wild relative teosinte to study the population genomics of copy number variants (CNVs). We find about 50% of the low-copy genome space shows CNVs (either deletion or duplication) in a single population, 98% of which is segregating in the population. Our initial results suggest low frequency deletions are enriched in genic regions and high frequency deletions in non-coding regions. This pattern is attributed to stronger purifying selection on deletions. In addition, we hypothesise that the local genome environment such as open chromatin may affect the rate of CNVs. The pattern of SFS from deletions is skewed towards singletons. Although genome-wide deletions appear largely neutral, we anticipate that some deletions may affect fitness and are currently testing the contribution of CNVs regions to phenotypic variation. Taken together, our results begin to shed light on the importance of structural variation in plant evolution and adaptation.

83M Understanding the Hidden Complexity of Latin American Population Isolates. *J. Mooney*[1], C. Huber[2], S. Service[3], J.H. Sul[4], C. Marsden[2], N. Freimer[3], K. Lohmueller[2], Costa Rica/Colombia Consortium for Genetic Investigation of Bipolar Endophenotypes 1)) Department of Human Genetics, University of California Los Angeles, Los Angeles, California; 2) Department of Ecology & Evolutionary Biology, University of California Los Angeles, Los Angeles, California; 3) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, California; 4) Department of Psychiatry and Biobehavioral Sciences, Semel Center for Informatics and Personalized Genomics, University of California Los Angeles, Los Angeles, California.

Population isolates have played a key role disease-gene mapping studies. However, most population isolates examined to date were founded from a single ancestral population. There is limited knowledge about the demographic history of admixed population isolates. Here we investigate genomic diversity of recently admixed population isolates from Costa Rica and Colombia and compare their diversity to a benchmark population isolate, the Finnish. Our data set consists of whole genome sequence data from 449 individuals, with a mean coverage of approximately 24X. We find that Latin American isolates have increased genetic diversity measured both by the number of variants and the average number of pairwise differences relative to the Finns. However, we also observed an increase in the amount of identity by descent (IBD) segments in the Latin American isolates relative to the Finns. The increase in IBD segments is likely a consequence of a very recent and severe population bottleneck during the founding of the admixed population isolates. The Finnish, on the other hand, have experienced older and longer bottlenecks which reduced genetic diversity relative to the admixed isolates. Furthermore, we found that the proportion of the genome that falls within a long run of homozygosity (ROH) in Costa Rican and Colombian individuals was significantly greater than that in the Finnish. Our simulations and extended pedigree data demonstrate the necessity of recent inbreeding for the generation of long ROH, indicating an increase in recent consanguinity in the Latin American isolates relative to that seen in the Finns. Lastly, we found that recent consanguinity increases the number of deleterious variants found in the homozygous state relative to neutral variants, which is particularly relevant if deleterious variants are recessive. In summary, our study shows that the genetic diversity in isolated populations can be complex, with multiple demographic factors operating on different timescales leaving distinct patterns in the genome. Consequently, it is imperative that we understand the demographic history of each isolate before making predictions about genetic diversity and the distribution of deleterious variation throughout their genomes.

84M Estimating the relative contribution of deleterious and neutral SNPs to agronomic phenotypes. TJK Kono[1], C Liu[1], EE Vonderharr[1], D Koenig[2], JC Fay[3], KP Smith[1], *PL Morrell*[1] 1) Department of Agronomy & Plant Genetics, University of Minnesota, St. Paul, MN; 2) Department of Botany & Plant Sciences, University of California, Riverside; 3) Department of Biology, University of Rochester.

Targeted identification and elimination of segregating deleterious variants has been proposed as a novel approach to plant breeding. This approach is motivated in part by the observation that demographic events and strong selection associated with domestication pose a "cost of domestication." This includes an increase in the proportion of genetic variants at phylogenetically constrained sites, where the mutation is likely to impose a fitness cost. Recent advances in DNA resequencing technology and sequence constraint-based approaches to predict the functional impact of a mutation now permit the identification of putatively deleterious SNPs on a genome-wide scale. Previous surveys have shown that individual crop genomes carry hundreds of putatively deleterious SNPs. However, the contributions of these SNPs to phenotypic variation and the relationship between deleterious variation and recombination rate have received limited empirical evaluation. We use exome resequencing and SNP genotyping in three cycles of an experimental spring six-row barley breeding population to compare the phenotypic contributions of putatively deleterious SNPs to those of neutral SNPs and track their inheritance through a pedigree that was subject to selection. While deleterious SNPs occur across the genome, in this barley population they are more abundant in the chromosome arms, where both gene density and diversity are higher. Across functional classes of SNPs from noncoding to putatively deleterious, putatively deleterious SNPs explain more phenotypic variance for grain yield than other classes of SNPs at the same frequency.

85M Inter-species variation in the gut microbiota controls gene regulation in primates. *A.L. Muehlbauer*[1,2], A.L. Richards[3], A. Alazizi[3], M. Burns[1,2], A. Gomez[5], J. Clayton[6,7], K. Petrzelkova[8], C. Cascardo[3], R. Pique-Regi[3,4], F. Luca[3,4], R. Blekhman[1,2] 1) Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, Minnesota; 2) Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota; 3) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan; 4) Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan; 5) Department of Animal Science, University of Minnesota, Saint Paul, Minnesota; 6) Department of Computer Science, University of Minnesota, Minneapolis, Minnesota; 7) Biotechnology Institute, University of Minnesota, Saint Paul, Minnesota; 8) Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czech Republic.

The human gut microbiome plays an important role by regulating the host's immune system, performing metabolic functions, extracting nutrients, and protecting against disease. In addition, microbiome composition exhibits extreme variation between host species; in primates, this variation clusters along expected host phylogenetic relationships. However, it is unclear how inter-species variation in gut microbiome composition impacts host physiology, and especially how it influences species-specific health conditions. A possible mechanism for the gut microbiome to affect host physiology is by altering gene expression in neighboring colonic epithelial cells. Here, we used a novel in vitro experimental system to assess how human host cells respond to inter-species variation in microbiome composition. We incubated human colonic epithelial cells with live microbiome samples extracted from four primate species (human, chimp, gorilla, and orangutan), and used RNA-seq to identify genes that respond to compositional differences in these microbiomes. We found 3728 genes with a conserved response, showing a similar regulatory response to the microbiomes of all four species. In addition, we identified 554, 388, 129, and 16 genes that respond specifically to human, chimp, gorilla, and orangutan microbiomes, respectively. We found that these genes are associated with GWAS traits, and that genes that respond only to human microbiomes are enriched for genes associated with relevant gut diseases, including colorectal cancer and Crohn's disease. These results provide a first glimpse into a possible coevolutionary relationship between primates and their microbial symbionts, and demonstrate possible mechanisms for microbiome variation driving species-specific health phenotypes.

86M Localizing and classifying adaptive targets with trend filtered regression. *M.R. Mughal*[1], M. DeGiorgio[2,3] 1) Bioinformatics and Genomics at the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA; 2) Departments of Biology and Statistics, Pennsylvania State University, University Park, PA; 3) Institute for CyberScience, Pennsylvania State University, University Park, PA.

Identifying genomic locations of natural selection from sequence data is an ongoing challenge in population genetics. Current methods utilizing information combined from several summary statistics typically assume no correlation of summary statistics regardless of the genomic location from which they are calculated. However, due to linkage disequilibrium, summary statistics calculated at nearby genomic positions are highly correlated. We introduce an approach that accounts for the similarity of statistics calculated from adjacent genomic regions through trend filtering, while reducing the effect of multicollinearity through regularization. Our penalized regression framework has high power to detect sweeps, is capable of classifying sweep regions as either hard or soft, and can be applied to other selection scenarios as well. We find that our method is more robust to missing data than similar current approaches, often has higher power, and is also robust to strong background selection. Application to human genomic data revealed positively selected regions discovered by previous methods such as *LCT* in European populations, as well as novel candidates such as a hard sweep at *ABCC11* in South Asians,

which has been implicated in ear wax consistency in East Asians.

87M Contrasting the genomic consequences of precipitously declining local populations of Florida scrub-jays. *T. Nguyen*[1], N. Chen[5], R. Bowman[4], J. Fitzpatrick[1,3], A. Clark[1,2] 1) Department Ecology and Evolutionary Biology, Cornell University, Ithaca, NY; 2) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 3) Cornell Lab of Ornithology, Cornell University, Ithaca, NY; 4) Archbold Biological Station, Venus, FL; 5) Center for Population Biology and Department of Evolution and Ecology, University of California, Davis, Davis, CA.

Understanding the genomic consequences of declining populations – even while they still appear to be robust – is a dynamic challenge in evolutionary biology, ecology and conservation. In particular, habitat loss and fragmentation may induce extirpations and reduce connectivity between populations, leading to shrinking populations and concomitant declines in genetic diversity. Loss of genetic diversity can have severe consequences for species, including inbreeding depression, reduced resistance to diseases, and increased susceptibility to environmental stochasticity. Thus, describing the relationship between genetics and population size is critical for predicting species' vulnerability to further perturbations and for conservation management. Although ample theoretical knowledge exists about the impacts of declining populations on the genetic structure of populations, empirical evidence still remains scarce, especially in natural populations. We leverage a dataset of 7,834 autosomal single-nucleotide polymorphisms and demographic data for 288 Florida scrub-jays (FSJ) from two populations of varying sizes and trajectories to investigate the genomic consequences of population decline in the wild. We compare the FSJ population at Archbold Biological Station (ABS), which has remained stable owing to intensive management, with a nearby population at Placid Lakes Estates (PLE), which has declined precipitously over the last decade due to housing development. We used PLINK to assess changes in genetic diversity, level of inbreeding and degree of relatedness of individuals within and between these two populations from 2000 to 2008 to characterize in detail the genetic changes accompanying the beginning of this rapid population decline. We found that birds in PLE in 2000 were more heterozygous than birds in ABS in 2000, most likely due to greater initial immigrant rates into that population. Interestingly, we also found that birds in PLE had a higher inbreeding coefficient and degree of relatedness than birds in ABS across both years. Finally, we tested for changes in heterozygosity, level of inbreeding and degree of relatedness from 2000 to 2008 in either population but detected no significant differences. It is possible that not enough time had elapsed over the 8 years to reveal a detectable genetic signature. This study provides the preliminary results for our larger state-wide study to better understand the genomic consequences of population decline over long time scales.

88M Recent environmental adaptation in an invasive species: house mice in the Americas. *M. Phifer-Rixey*[1], K. Bi[2], F. de Mello Martin[3], M.W. Nachman[2] 1) Department of Biology, Monmouth University, West Long Branch, NJ; 2) Museum of Vertebrate Zoology, Department of Integrative Biology, University of California, Berkeley; 3) Department of Ecology and Evolution, Australian National University.

Recently, house mice (*Mus musculus domesticus*) have expanded their range in association with humans, establishing populations in a variety of novel habitats, including most of the Americas. While this expansion has made them notorious as exotic, invasive pests, it also provides a unique opportunity to study the genetic basis of evolutionary change over short time scales in a vertebrate model system. Sampling of populations spanning a latitudinal transect in North America revealed strong evidence of environmental adaptation. Using phenotypic analysis, RNA-seq, and genome scans, we identified traits and genes underlying environmental adaptation and uncovered a key role for regulatory evolution. Here, we build on those results by sampling > 100 mice from populations along a latitudinal transect from southern Argentina to equatorial Brazil, providing a "natural replicate." Exomic sequencing was used to evaluate patterns of genetic variation in these populations and to identify candidate genes contributing to adaptation to environmental variation, including shared and unique targets of selection. Together, these data provide insight into evolutionary processes underlying the expansion of house mice into the Americas.

89M Antiviral enzyme APOBEC3G introduces clustered inherited mutations that fuel adaptation in human populations. *Y. Pinto*[1,2], E. Li[2], E.Y. Levanon[1], A. Keinan[2,3,4,5] 1) Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Cornell Center for Comparative and Population Genomics, Cornell University, Ithaca, NY; 4) Center for Vertebrate Genomics, Cornell University, Ithaca, NY; 5) Center for Enervating Neuroimmune Disease, Cornell University, NY.

The molecular clock is based on the assumption that mutations mostly occur randomly and independent of each other, thereby accumulating at a constant rate over time. We recently discovered a mutational process driven by the activity of a member of the APOBEC family of deaminases. APOBECs can concurrently introduce clusters of mutations in single-stranded DNA (ssDNA), which enables diversifying immunoglobulin genes and inactivating viral DNA. Recently, this gene family has been shown to introduce clusters of somatic mutations caused by cytosine deamination in multiple types of cancers. We hypothesized that APOBEC3 members, following their rapid expansion in primates, may have introduced similar clusters of mutations in the germline that can be inherited. We tested this hypothesis by relying on the unique mutagenesis pattern of

APOBEC3s: In the presence of one of several sequence motifs that include a C, it can mutate this together with many nearby C nucleotides. We traced such activity of APOBEC3G in a comparative genetic analysis of the human, archaic humans (Neanderthal and Denisova) and chimpanzee genomes. We conservatively identified tens of thousands of such clustered nucleotide-specific mutations. We hypothesized that these mutation clusters are less likely to be neutral, as compared to single-nucleotide mutations, with many not being observed as inherited due to being lethal or very deleterious early in conception and development. Combined with them occurring more often in functional elements due to APOBECs acting on ssDNA, the mutation clusters that segregate are more likely to be targeted by positive selection. Hence, we applied several population genetic tests of adaptation for a subset of mutation clusters that are also polymorphic in human populations. Our results point to positive selection on several such clusters of genome-wide significance. Of interest is that most instances of adaptation are population-specific, which seem to correspond to adaptation to local environments, and act on cis-regulatory eQTLs. These results suggest that APOBEC3G-introduced mutations are more likely to have functional consequences—than is already expected by APOBECs' tendency to mutate ssDNA—which are maintained by natural selection. Combined, we provide evidence for exaptation of an antiviral mechanism as a source of genomic variation with population-specific functional consequences in humans.

90T Combining population genomics and fitness QTL to identify the genetics of local adaptation in *Arabidopsis thaliana*. *N. Price*[1], B. Moyers[1], J. Lasky[2], J. Monroe[1], J. Mullen[1], L. Lopez[2], G. Oakley[3], J. Lin[1], J. Ågren[4], D. Schrider[5], A. Kern[5], J. McKay[1] 1) Bioagricultural Sciences & Pest Management, Colorado State University, Fort Collins, CO; 2) Department of Biology, Pennsylvania State University, University Park, PA 16802, USA; 3) Department of Botany and Plant Pathology and Center for Plant Biology, Purdue University, West Lafayette, IN 47907, USA; 4) Department of Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden; 5) Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA.

Evidence for adaptation to different climates in the model species *Arabidopsis thaliana* is seen in reciprocal transplant experiments, but the genetic basis of this adaptation remains poorly understood. A major open question is whether local adaptation results from 1) genetic tradeoffs (GT), where alleles that maximize fitness in the home environment are deleterious in alternative environments or 2) conditionally neutral (CN) alleles that are advantageous in the home environment but neutral in alternative environments. Addressing the above, will help enhance our understanding on how temporally or spatially varying selection maintains genetic variation, which population genetic signatures can be used to identify local adaptation in the genome, and finally the biological process underlying local adaptation.
Field-based quantitative trait loci (QTL) studies provide direct but low-resolution evidence for the genetic basis of local adaptation. Using high-resolution population genomic approaches that included the identification of: (1) recent sweeps; (2) SNPs showing significant allele frequency divergence (*Fst*); and (3) SNPs showing significant correlations to climate, we examine local adaptation along previously identified genetic tradeoff (GT) and conditionally neutral (CN) QTL for fitness between locally adapted Italian and Swedish *A. thaliana* populations (Ågren et al. *PNAS*, 2013). Using permutation tests, we examine whether GT or CN QTL peaks are found significantly close to genomic regions enriched in the aforementioned population genomic signals of local adaptation. We find that genomic regions enriched in high *Fst* SNPs are found significantly close to GT QTL peaks while peaks of CN QTL are close (not significantly) to regions enriched with SNPs showing correlations to climate in Eurasia. Among the high *Fst* regions identified, some show significant correlations to climate in Eurasia and evidence of recent sweeps in Sweden. Examining unfolded site frequency spectra across genes containing high F_{ST} SNPs suggests genetic tradeoffs may be due to more recent adaptation in Sweden than Italy. Finally, we collapse a list of thousands of genes spanning GT QTL to 42 genes that likely underlie the observed genetic tradeoffs and explore potential biological processes driving these tradeoffs.

91M Disambiguating second degree relationship types using identity by descent sharing to third party samples. *Y. Qiao*, Amy Williams Cornell University, Ithaca, NY.

Estimates of genetic relatedness are useful in many contexts, notably pedigree-based analyses, yet relatedness inference methods typically report only the degree of relatedness of a pair of samples and do not uncover their pedigree relationship. Identifying pedigree relationships is simple for first degree relatives—parent-child and full siblings. Yet only one degree more distant relationships, including grandparent-grandchild (GP), half-siblings (HS), and avuncular (AV) pairs, are very challenging to distinguish. This ambiguity hinders efforts to perform pedigree reconstruction, an approach with the potential to enable usage of population-based samples in family-based studies (e.g., studies of de novo recombination and mutation).

Our approach differs from existing methods that attempt to distinguish second degree relatives in that it leverages identical by descent (IBD) sharing between the second degree pair and other third party individuals. These outside individuals are expected to have different sharing rates depending on the second degree relationship type. For example, for second degree relatives x_1 and x_2, we consider another sample y that is related to both x_1 and x_2 but not descended from either one. If x_1 is the grandparent of x_2, x_1 and y are expected to share 4-times as much of their genome IBD as x_2 and y, since y is related to x_1 and x_2 through an ancestor of both. By similar logic, the amount of IBD sharing between x_1 and y is expected to be 2-times or

1-times that of x_2 if x_1 is, respectively, the uncle/aunt or half-sibling of x_2. We implemented this idea, analyzing the amount of DNA shared with samples that are third through sixth degree relatives of the second degree relatives, and used logistic regression to perform classification.

To evaluate this method, we simulated 1,000 pairs of each type (GP, HS, and AV), initially considering a scenario with genotype data for numerous close relatives for each pair: 6 first cousins and 18 second cousins of the older second degree relative. Using the simulated IBD sharing information between these relatives and both x_1 and x_2, our regression model achieved >99% classification accuracy. We then turned to a more realistic setting by using data for only 2 first cousins and 4 second cousins of the older second degree relative. Given this information, the classifier achieved 100% accuracy when identifying GP relatives and also correctly identified 97% of HS and 95% of AV pairs. These results are currently based on exact IBD segment information, so real data is likely to have reduced accuracy, but we anticipate this new method will be able to distinguish second degree relatives with high fidelity and will be applicable on large population datasets.

92T A map of highly constrained coding regions in the human genome. *A.R. Quinlan*, J.M. Havrilla, B.S. Pedersen, R.M. Layer Human Genetics, University of Utah, Salt Lake City, UT.

Deep catalogs of genetic variation collected from many thousands of humans enable the detection of intraspecies constraint by revealing coding regions with a scarcity of variation. While existing metrics such as RVIS and pLI summarize constraint for entire genes, single metrics cannot capture the fine-scale variability in constraint within each protein-coding gene. To provide greater resolution, we have created a detailed map of constrained coding regions (CCRs) in the human genome by leveraging coding variation observed among 123,136 humans from the Genome Aggregation Database (gnomAD). The most constrained coding regions in our map are enriched for both pathogenic variants in ClinVar and de novo mutations underlying developmental disorders. While observed regions of constraint are generally correlated with interspecies conservation, many CCRs are highly constrained in the human lineage but not strongly conserved across species. CCRs also reveal protein domain families under extreme constraint, suggest unannotated or incomplete protein domains, and facilitate the prioritization of previously unseen variation in studies of disease. We explore the tissue specificity of and impact of CCRs on protein-protein interactions. Finally, we demonstrate that a subset of the most constrained CCRs likely exist within genes that cause yet unobserved human phenotypes owing to strong purifying selection.

93M The genomics of invasion: characterization of the red lionfish from its native and introduced range. *E.M. Reed*[1], M.O. Burford Reiskind[1], R.B. Roberts[2] 1) Department of Applied Ecology, North Carolina State University, Raleigh, NC; 2) Department of Biological Sciences, North Carolina State University, Raleigh NC.

Invasive species present a major threat to biodiversity and ecosystem health. Understanding the genetic characteristics of these invaders is key to assessing the risks they pose and informing conservation decisions. The invasion of lionfish in the Atlantic Ocean has had substantial negative impacts on native species, including those of economic importance to fisheries. It is a highly effective invader and has spread since its initial discovery in Florida as far north as New York and into the Gulf of Mexico and Caribbean Sea. Initial reports suggest that two species, *Pterois volitans* and *Pterois miles* make up the invasive populations, and consequently some have credited the invaders' success to potential hybridization. We used population genomic techniques to characterize invasive lionfish and address the question of hybridization between the two potential invading species.

Methods. We sampled red lionfish (*P. volitans*) from five locations in its introduced range and two in its native range. We also collected samples of *P. miles* from Sri Lanka and *P. antennata* from the Yale museum and extracted genomic DNA from fin clips for all individuals (n=83). We built genomic libraries using ddRADseq and identified SNPs using the STACKS *denovo* and *population* pipelines. We measured genetic diversity (H_E), inbreeding coefficient (F_{IS}), and genetic differentiation (pairwise F_{ST}) for all populations in GENEPOP. We investigated genetic structure in the introduced and native range of *P. volitans* in STRUCTURE and with discriminant analysis of principle components (DAPC). Finally, we conducted tests for outlier loci between one native and one invasive population in LOSITAN and DAPC.

Results. We found low genetic diversity in the invasive range compared to the native range of *P. volitans,* though the invasive range also showed comparatively low inbreeding, indicating high migration rates between invasive groups. This was supported by low differentiation and genetic structure in the invasive range. We found no evidence of admixture between *P. volitans* and *P. miles* in the invasive range, and both pairwise F_{ST} and genetic structure indicate that the invasive originated from Taiwan and began in the Bahamas. We therefore used these two populations for our outlier analyses. We identified 1,866 outlier loci using LOSITAN and 71 outliers using DAPC. Of these, 67 loci were present in both. Therefore, there is some evidence that lionfish are undergoing local adaptation in their invasive range.

94M Population genomic analyses of meiotic drive in a species pair. *J.A. Reinhardt*[1], R.H. Baker[3], K.A. Paczolt[2], G.S. Wilkinson[2] 1) Biology, SUNY Geneseo, Geneseo, NY; 2) University of Maryland College Park, College Park, MD; 3) American Museum of Natural History, New York, NY.

Meiotic drive alleles violate the law of equal segregation and as a result, have an advantage over other alleles at the same

locus. In the case of X-linked distorters, the sex ratios produced by male carriers are distorted in favor of females, as Y-bearing sperm fail to develop. In theory, drive loci can be stably maintained as single loci, but they are often found as part of large - even chromosome wide - inversion polymorphisms. We report a comparative analysis of genomes and transcriptomes collected from a family of flies known to carry drive polymorphisms in multiple species. Comparing sequences of hundreds of expressed genes from the testes of males with female biased and equal sex ratios in two species, we found evidence for both a strongly diverged SR X chromosome (T. dalmanni) and in sharp contrast a case (T. whitei) in which drive had little detectable impact on gene expression or sequence divergence. We also identified a single candidate drive locus in T. whitei carrying fixed genetic differences between the standard and sex-ratio individuals. Population genetic analyses using a high quality draft genome of T. dalmanni detected differences in the copy number of repetitive sequences a thousands of fixed polymorphism, along with a large centromere effect on variation in both SR and standard sex chromosomes. We conclude an X-linked sex ratio distorter likely exists as a freely recombining polymorphic locus in T. whitei, in contrast with a large and ancient chromosome-wide distorter haplotype in T. dalmanni, despite similarities and apparent temporal stability of the phenotype.

95M Demographic history influences levels of deleterious variation and explains patterns of inbreeding depression in Channel Island foxes and Isle Royale wolves. *Jacqueline Robinson*[1,2], Kirk Lohmueller[2,3,4], Robert Wayne[2] 1) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 2) Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA ; 3) Interdepartmental Program in Bioinformatics, University of California Los Angeles, Los Angeles, CA ; 4) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA.

As most non-neutral mutations are deleterious, the inability of selection to effectively remove them from small isolated populations incurs a fitness cost. These detrimental consequences have been observed in numerous wild populations, such as Isle Royale gray wolves, which suffer from severe inbreeding depression. However, other populations, such as California's Channel Island foxes, appear healthy and even persist for thousands of generations despite their small population size and isolation. The explanation as to why one small population shows inbreeding depression and the other does not remains elusive. To address this question, we combine morphological assessment, analyses of complete genomes, and population genetic simulations to investigate these two contrasting systems. We find that although island foxes show no canonical signs of inbreeding depression, their genomes harbor an elevated burden of deleterious alleles and exceptionally low levels of diversity due to long-term small population size. In contrast, there is a high prevalence of malformed vertebrae and other anomalies in Isle Royale wolf skeletons, and Isle Royale wolf genomes contain large runs of homozygosity many megabases in length, occasionally spanning entire chromosomes, interspersed with regions of high variation. This pattern is consistent with recent and intense inbreeding since the founding of the Isle Royale population less than a century ago. Simulations under different demographic models provide insight into the fundamental differences in patterns of deleterious variation between these contrasting systems. Simulations suggest that individuals recently descended from historically large populations, such as Isle Royale wolves, will carry more strongly deleterious recessive variants than individuals from historically small populations. As such, recent inbreeding is predicted to create more homozygous recessive deleterious genotypes in the Isle Royale wolves than in a population with small long-term effective size, such as island foxes. More generally, our findings argue that inbreeding depression is caused by strongly deleterious recessive alleles, having implications for the management of populations increasingly threatened by habitat loss and fragmentation.

96M Fine-scale resolution and analysis of inbreeding tracts in domestic dogs. *A.J. Sams*[1], A.R. Boyko[1,2] 1) Embark Veterinary, Inc., Ithaca, NY; 2) Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY.

Inbreeding and consanguinity leave distinct genomic traces, most notably long genomic tracts that are identical by descent and completely homozygous. These tracts can contribute to inbreeding depression if they contain deleterious variants that are (partially) recessive. Several lines of evidence have been used to show that long (>5 Mb) tracts of homozygosity are disproportionately likely to harbor deleterious variation, but the extent to which long versus short tracts actually contribute to homozygosity at loci known to be deleterious and recessive has not been studied.

In domestic dogs, nearly 200 mutations are known to cause recessive diseases, most of which can be easily assayed using SNP arrays. By examining genome-wide data from over 200,000 markers (including 150 recessive disease variants), we built high-resolution inbreeding maps for nearly 3000 dogs, recording homozygosity tracts down to 500 kb. We observed over 500 homozygous deleterious recessive genotypes in the panel, 90% of which overlapped with homozygosity tracts inferred by GERMLINE. Although most of these genotypes were contained in long homozygosity tracts >5Mb, 11% were contained in short 0.5-2Mb tracts, a significant enrichment compared to the genetic

background, suggesting that even short tracts are useful for computing inbreeding metrics like coefficient of inbreeding (COI). In our dataset, COI differed significantly both among breeds and between dogs within a breed, and could be accurately estimated by the kinship coefficient of the dog's parents (r^2 = 0.99). All breeds harbored some regions of reduced genetic diversity due to drift or selective sweeps; nevertheless, COI values could be reduced by 20-40% in each breed if accurate kinship estimates were used in mating decisions.

97M *De novo* gene conversion events are enriched in sample-specific low heterozygosity regions. *J. Sannerud*[1], J. Peralta[2], J. Blangero[2], A. Williams[1] 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) South Texas Diabetes & Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX.

Crossing-over and non-crossover (NCO) gene conversion are two forms of recombination that introduce genetic diversity into humans and sexually reproducing species. Although numerous studies of recombination have revealed variants that alter the frequency and location of these events, much still remains unknown about the placement of recombinations along the chromosome. This is partly complicated by the relatively low frequency of recombinations; it is unlikely that any two people can independently generate the same crossovers, and it is difficult to study large numbers of meiotic products from a single individual to create a personalized genetic map. Nevertheless, analyses of the genetic context in which recombinations occur have the potential to reveal locus-specific properties that alter the locations of recombination events.
Recent studies of non-crossover leveraged sequence data and uncovered numerous events wherein the nearest non-converted markers are very distant from the conversion tract. These events suggest the possibility that local heterozygosity in the proband influences the frequency of non-crossover. Using data from Halldorsson *et al*. (2016), which detailed over five hundred NCOs, we investigated this possibility by comparing the population-level heterozygosity of these regions with that of the proband. Specifically, we calculated the region-specific heterozygosity in 91 British individuals from the 1000 Genomes Project, and computed Z-scores for the heterozygosity of the deCODE samples. The distribution of the resulting Z-scores is significantly skewed, with an excess of very low heterozygosity regions evident when considering events with $|Z| \geq 2$ ($P < 2.2 \times 10^{-16}$, binomial test). The strong left-tailedness of the results implies that a substantial proportion of the NCOs in these individuals occurred in regions with much lower heterozygosity in the proband than the population average. We also compared to the heterozygosity of Simons Genome Diversity Project Icelander subjects, which recapitulated these results ($P < 2.2 \times 10^{-16}$, binomial test).

98M The unreasonable effectiveness of population genetic inference via image recognition. *D.R. Schrider* University of North Carolina, Chapel Hill, NC.

The availability of population-scale genomic datasets has given researchers a new avenue toward answering questions about populations' recent evolutionary histories. To what extent is a population's genomic variation shaped by the interplay between natural selection and recombination? Has the population experienced substantial changes in size? Has the population experienced gene flow from closely related populations/species, and if so, which portions of the genome were affected? In recent decades a host of theoretical and methodological advances have addressed these problems. Typically these methods summarize patterns of genetic variation with a statistic designed to be sensitive to the phenomenon of interest. More recently, approximate Bayesian computation and machine learning approaches have proved successful in simultaneously examining many of these statistics in order to make far more accurate inference. The rationale for these approaches is that any single statistic will capture only a subset of the discriminatory information present in the original data, and thus a set of complementary statistics will perform better. A more fruitful approach would thus be to perform inference directly on the input sequence data rather than digesting it into a set of numbers. Here we attempt to accomplish this by representing a population genetic alignment as an image and using modern deep learning techniques for image processing. We apply this approach to the problems listed above, and find that in each case it matches or exceeds the accuracy of current state-of-the-art methods. Thus, when applied to images of alignments, modern image recognition algorithms outperform expert-derived statistics and even collections thereof. In light of this result, we argue that the rate of progress in evolutionary genetic inference might be improved by devoting greater effort to exploring the myriad possible image representations of sequence alignments and deep learning architectures for processing them, rather than attempting to devise more powerful summary statistics.

99M Rapid, phase-free detection of long identical by descent segments enables fast relationship classification. *Daniel Seidman*[1], Sushila Shenoy[1], Minsoo Kim[2], Amy Williams[1] 1) Cornell University Graduate School, Ithaca, NY; 2) Weill Cornell Medicine, New York, NY.

Identical by descent (IBD) segments are a useful tool in modern genetics for a variety of applications ranging from demographic inference to relative detection. While many algorithms for detecting IBD exist, all prominent methods rely on internal phasing or pre-phasing and therefore require substantial amounts of compute time. As genetic datasets grow in size, methods for inferring IBD that scale well will be critical.

We present an algorithm that rapidly locates IBD segments by identifying long tracts of identical by state (IBS) between

unphased individuals with application to characterizing relatedness in large samples. Our method is extremely fast and viable for large, biobank-scale datasets. The implementation leverages bitwise operations to check IBS of one sample against others with instruction-level parallelism. Other IBD methods typically employ computationally intensive hidden Markov models that are essential for finding short segments, but our results indicate that long stretches of IBS recover true IBD segments. We leverage these long segments to infer relatedness, focusing on segments above a given threshold length for this analysis.

We ran Refined IBD and our method on a sample of 3,000 individuals, with Refined IBD taking 32 CPU days to complete compared to 2.01 hours for our method (a speedup of 382x). In simulations leveraging European haplotypes, we find that inferred 10 cM or longer segments had a false positive rates of 0.3%, suggesting high reliability. Furthermore, for simulated first through third degree relatives, relatedness inference using 10 cM or longer segments gave almost identical results to using the full range of IBD segments. As well, inference accuracy was only slightly reduced for fourth and fifth degree relatives. We are in the process of further optimizing our implementation and anticipate greater speed improvements and application of relatedness inference to wide ranging population groups.

100M Population genomics of a monitored sea anemone (*Condylactis gigantea*) in Florida, U.S. *N. Sheridan*[1,2], S. Seyoum[2], W. Sharp[3], B. Titus[4], M. Daly[5], A. Schrey[6], C. Richards[1] 1) Department of Integrative Biology, University of South Florida, Tampa, FL; 2) Florida Fish and Wildlife Conservation Commission, St. Petersburg, FL; 3) Florida Fish and Wildlife Conservation Commission, Marathon, FL; 4) Division of Invertebrate Zoology, The American Museum of Natural History, New York, NY; 5) Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH; 6) Department of Biology, Georgia Southern University Armstrong Campus, Savannah, GA.

The giant Caribbean sea anemone *Condylactis gigantea* is an ecologically important member of Florida's nearshore seagrass and hard-bottom habitats and offshore reefs. It provides habitat for many species, including obligate shrimp species, and is a cleaning station indicator recognized by reef fishes. Due to declines in abundance, the Florida Fish and Wildlife Conservation Commission (FWC) closed the commercial and recreational harvest of *C. gigantea* in late 2012. *Condylactis gigantea* is considered a single species throughout the Tropical Western Atlantic, though the genetic population structure of this phenotypically diverse anemone has not been investigated throughout most of its range, including Florida. Investigating genetic population structure may reveal subdivided populations that could be evolving independently. Therefore, to assess structure, we collected tentacle samples from 250 individuals at 9 locations and used restriction site associated DNA sequencing (RADseq) and traditional Sanger-sequencing. Initial analyses of the RADseq generated SNP data suggest two clusters and geographic partitioning. Individuals in the eastern Gulf of Mexico were genetically differentiated from individuals in the Middle and Lower Florida Keys. Mitochondrial DNA sequences (12S, 16S, and COIII) were invariant. Ribosomal DNA (rDNA; partial-18S-ITS1-5.8S-ITS2-partial-28S) analysis supports two clusters with admixture, but not similar geographic partitioning. Here, we detected two clusters co-occurring in nearshore and offshore habitats, which is concordant with prior results from Jamaica waters. The discordant results could be due to the RADseq results providing a more robust signal given the increase in loci or insufficient signal in the rDNA. Additionally, to quantify temporal changes in abundance during the closure, we monitored 17 fixed locations annually from 2013 to 2017. We detected an increase in mean annual density at six of the 17 sites, suggesting that no overall meaningful change in abundance has occurred. Expanding our understanding of this species' genetic structure, as well as monitoring its recovery from decline is crucial for the FWC to effectively manage this species. Using both single locus and genomic methods, shows that RADseq is a more powerful approach for investigating genetic structure in natural populations.

101M The Demographic and Migratory History of *Mycobacterium tuberculosis* in Canadian Northwest Territories. *A. Shockey*, C. Pepperell, T. Smith University of Wisconsin Madison, Madison, WI.

Mycobacterium tuberculosis (*M.tb*), the etiological agent of tuberculosis (TB) and leading cause of death due to an infectious disease, is an urgent threat to global public health. Inference of a pathogen's demographic history from its genomic data is a powerful means of gaining insight into the epidemiology of infectious diseases. For obligate human pathogens, such as *M.tb*, population size is constrained by transmission dynamics and the size of host populations. *M.tb* strains circulating in the Aboriginal communities of the Canadian Northwest Territories (NWT) present a unique framework to examine the demographic and migratory history of *M.tb*. A high proportion of TB cases in Aboriginal communities are associated with outbreaks, and TB disproportionally affects Canadian Aboriginal communities: 930/100,000/year versus the national incidence rate of 4.6/100,000/year. Previous studies have shown *M.tb* strains circulating in these communities form three monophyletic clades: ABjm, DS6 and SKc. It was previously determined that DS6 was introduced to Aboriginal communities in Quebec via the Canadian fur trade. In order to infer the demographic history of *M.tb* strains circulating in these communities, we used site-frequency spectrum (SFS) based methods to test for signatures of population size change in whole genome sequence data from 343 *M.tb* strains in the Aboriginal communities of Alberta, Quebec and Saskatchewan. Both the maximum likelihood phylogeny and neighbor-joining network of these isolates clearly delineate the ABjm, SKc and DS6 clades. These clades have short terminal branch lengths consistent with the star-like topologies of expanding populations. The SFS is leptokurtic, and demographic inference implemented in ∂a∂i indicates these populations have undergone an

expansion. In order to estimate changes in population size over time, we will use the Bayesian Skyline Plot model as implemented in BEAST. Additionally, we will perform ancestral state reconstruction at each node in the phylogeny using geographic location as a trait to infer the migratory history of these populations across the NWT. We expect the *M.tb* populations of the NWT to have undergone an expansion and to infer a Quebec origin for the most recent common ancestor. Characterizing the dynamics of *M.tb* strains circulating in these populations is important for understanding the epidemiology of this pathogen as an ongoing threat to global public health.

102M Male infertility is responsible for nearly half of the extinction observed in the mouse Collaborative Cross. *J.R. Shorter*[1], F. Odet[2], D. Aylor[3], W. Pan[2], C. Kao[5], C. Fu[5], A. Morgan[1], S. Greenstein[5], T. Bell[1,4], A. Stevans[2], R. Feathers[2], S. Patel[2], S. Cates[1,4], G. Shaw[1,4], D. Miller[1,4], E. Chesler[6], L. McMillian[5], D. O'Brien[2,4], F. Pardo-Manuel de Villena[1,4] 1) Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Cell Biology and Physiology, University of North Carolina, Chapel Hill, North Carolina 27599; 3) Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina 27695; 4) Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599; 5) Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27599; 6) The Jackson Laboratory, Bar Harbor, Maine 04609.

The goal of the Collaborative Cross (CC) project was to generate and distribute over 1000 independent mouse recombinant inbred strains derived from eight inbred founders. With inbreeding nearly complete, we estimated the extinction rate among CC lines at a remarkable 95%, which is substantially higher than in the derivation of other mouse recombinant inbred populations. Here, we report genome-wide allele frequencies in 347 extinct CC lines. Contrary to expectations, autosomes had equal allelic contributions from the eight founders, but chromosome X had significantly lower allelic contributions from the two inbred founders with underrepresented subspecific origins (PWK/PhJ and CAST/EiJ). By comparing extinct CC lines to living CC strains, we conclude that a complex genetic architecture is driving extinction, and selection pressures are different on the autosomes and chromosome X. Male infertility played a large role in extinction as 47% of extinct lines had males that were infertile. Males from extinct lines had high variability in reproductive organ size, low sperm counts, low sperm motility, and a high rate of vacuolization of seminiferous tubules. We performed QTL mapping and identified nine genomic regions associated with male fertility and reproductive phenotypes. Many of the allelic effects in the QTL were driven by the two founders with underrepresented subspecific origins, including a QTL on chromosome X for infertility that was driven by the PWK/PhJ haplotype. We also performed the first example of cross validation using complementary CC resources to verify the effect of sperm curvilinear velocity from the PWK/PhJ haplotype on chromosome 2 in an independent population across multiple generations. While selection typically constrains the examination of reproductive traits toward the more fertile alleles, the CC extinct lines provided a unique opportunity to study the genetic architecture of fertility in a widely genetically variable population. We hypothesize that incompatibilities between alleles with different subspecific origins is a key driver of infertility. These results help clarify the factors that drove strain extinction in the CC, reveal the genetic regions associated with poor fertility in the CC, and serve as a resource to further study mammalian infertility.

103M Minimal effects of neo-Y chromosomes on transcriptomes of house flies in spite of evidence that selection across ecological habitats maintains stable polygenic sex determination. *Jae Hak Son*[1], Tea Kohlbrenner[2], Svenia Heinze[2], Daniel Bopp[2], Richard Meisel[1] 1) Biology and Biochemistry, University of Houston, Houston, TX; 2) Molecular Life Sciences, University of Zurich, Zurich, Switzerland.

Sex determination (SD) evolves rapidly, with genes at the top of SD pathways differing between closely related species and even variable within species. It is not clear what evolutionary forces maintain stable polygenic SD systems with multiple master SD genes segregating within a population. The house fly, *Musca domestica,* has multiple male-determining neo-Y chromosomes, making it a good model system to identify the selective forces maintaining polygenic SD. The male-determining gene *Mdmd* is most frequently found on the Y chromosome (Y^M) and third chromosome (III^M). Y^M and III^M appear to have different phenotypic effects across ecological habitats, which could explain their stable maintenance. However, previous analyses of the sequences of these neo-Y chromosomes revealed very few differences from their homologous neo-X chromosomes. To address the paradox of ecologically-relevant phenotypic effects yet minimal sequence differences between the neo-Y chromosomes, we performed RNA-seq to determine the phenotypic effects on house flies carrying different neo-Y chromosomes.

First, to determine the effect of the III^M chromosome on gene expression, we performed RNAi knockdown of the SD gene *transformer* in genotypic females, creating sex-reversed males without a neo-Y chromosome. The sex-reversed males have similar gene expression profiles to normal genotypic males, and profoundly different from phenotypic females, in both head and abdomen. This suggests that alleles on the III^M chromosome have minimal effects on male phenotypes. Second, to compare the effects of Y^M and III^M on gene expression, we performed RNA-seq on one Y^M strain and two III^M strains that are otherwise isogenic. We also measured expression in a second nearly isogenic Y^M strain in which the third chromosome was replaced with non-*Mdmd* bearing chromosome III. The most substantial expression differences were observed between the two Y^M strains with different third chromosomes (not between the Y^M and III^M strains), providing further evidence that the III^M chromosome has minor effects on male phenotypes. These gene expression results, along with the previous sequence

analysis, paint a consistent picture that the Y^M and III^M neo-Y chromosomes are minimally differentiated from their neo-X homologs. We therefore conclude that phenotypic effects of the neo-Y chromosomes that are responsible for selective maintenance of polygenic SD depend on ecological contexts that are not assayed in our experiment.

104M Discovery of *Drosophila melanogaster* from Wild African Environments and Genomic Insights into Species History. *Q.D. Sprengelmeyer*[1], S. Mansourian[2], E. Jirle [2], J.D. Lange[1], M.C. Stensmyr[2], J.E. Pool[1] 1) Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; 2) Department of Biology, Lund University, Lund, Sweden.

A long-standing enigma concerns the geographic and ecological origins of the intensively studied vinegar fly, *Drosophila melanogaster,* a globally widespread species which "has invariably appeared to be a strict human commensal". In spite of its sub-Saharan origins, previous efforts to collect the species from wild environments in Africa have failed. Here, we report the first collection of this species from African wilderness, specifically from national parks in Zambia and Zimbabwe. After sequencing the genomes of wild-collected flies from Zambia, we found notable differences between park and town populations in the frequencies of chromosomal inversions, and at genes with developmental and synaptic functions. Combining these new genomes with prior data allowed us to gain novel insights into the history of this species' geographic expansion. Our findings provide a substantially improved model of the species' demographic history that will provide a critical resource for future evolutionary and genomic studies of this key model organism. Furthermore, the opportunity to study wilderness populations of /D. melanogaster/ will open the door for studies on the biological basis of its adaptation to human environments.

105M Identification of *De novo* genes created by chromosomal rearrangements in *Drosophila yakuba*. *Nicholas Stewart*, Rebekah Rogers Bioinformatics, UNC Charlotte, Charlotte, NC.

While analysis investigating population genetic variation has been strong, variation of chromosome structure has been an underappreciated form of genetic diversity. This variation is the result of complex mutations such as duplications, deletions and chromosomal rearrangements. These chromosome rearrangements provide the opportunity for major changes to genes and/or gene expression by creating chimeric genes or creating *de novo* transcripts by attaching previously untranscribed loci with coding regions. We used paired end genomic sequence data and RNA-Seq data of 11 strains of *Drosophila yakuba* to screen for chromosomal rearrangements and the creation of *de novo* and chimeric genes. We have found multiple cases of *de novo* genes created by chromosomal rearrangements within and between chromosomes. These *de novo* genes are expressed more in testes than somatic tissue. Overall there is considerable variability in chromosome structure between the 11 strains. More rearrangements were identified on the X chromosome than any of the autosomes, suggesting the X is more susceptible to chromosome alterations. Here we provide evidence suggesting that chromosomal rearrangements are a source of novel genetic variation, and can be a mechanism to facilitate the formation of *de novo genes*.

106M African ROH Drive Enrichment of Deleterious Alleles in a Sample of Admixed Individuals. *Z.A. Szpiech*[1], A.C.Y. Mak[2], M.J. White[2], D. Hu[2], C. Eng[2], E.G. Burchard[2], R.D. Hernandez[1] 1) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco; 2) Department of Medicine, University of California San Francisco.

Runs of homozygosity (ROH) are important genomic features that manifest when identical-by-descent haplotypes are inherited from parents. Their length distributions are informative about population history, and their genomic locations are useful for mapping recessive loci contributing to both Mendelian and complex disease risk. We have previously shown that ROH, and especially long ROH that are likely the result of recent parental relatedness, are enriched for homozygous deleterious coding variation in a worldwide sample of outbred individuals (Szpiech, et al. 2013). However, the distribution of ROH in admixed populations and their relationship to deleterious homozygous genotypes is understudied.

Here we analyze whole genome sequencing data from 1,441 individuals from African American, Puerto Rican, and Mexican American populations. These populations are three-way admixed between European, African, and Native American ancestries and provide an opportunity to study the distribution of deleterious alleles partitioned by local ancestry and ROH. We re-capitulate previous findings that long ROH are enriched for deleterious variation genome-wide. Then, partitioning by local ancestry, we compare the proportion of deleterious homozygotes in ROH comprised of single ancestry haplotypes to the proportion of benign homozygotes in those ROH. We find that ROH falling in African ancestry tracts are enriched the most followed by European and Native American ROH.

These results suggest that, while ROH on any haplotype background are associated with an inflation of deleterious homozygous variation, African haplotype backgrounds that have experienced recent demographic bottlenecks may play a

particularly important role in the genetic architecture of complex diseases for admixed individuals, highlighting the need for further population genetic study of these populations.

108M Clustering of non-synonymous substitutions provides evidence for widespread epistasis and convergence in protein evolution. *A.M. Taverner*[1], L. Blaine[2], P. Andolfatto[1,3] 1) Lewis-Sigler Institute, Princeton University, Princeton, NJ; 2) Department of Molecular Biology, Princeton University, Princeton, NJ ; 3) Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ.

Evolutionary rate varies among sites within proteins, but the causes of this variation are as yet poorly understood. Here, we compare the rate of autocorrelation among non-synonymous substitutions among four taxa: yeast, *Arabidopsis*, *Drosophila* and primates. Across these diverse groups, we find that non-synonymous substitutions occur in clusters that extend over a 20-30 codon length scale. We confirm that the strength of clustering is dependent on the distance between substitutions in the protein rather than distance in the DNA sequence. Furthermore, we find that non-synonymous substitutions cluster more strongly and exhibit compensatory dynamics within lineages but exhibit patterns of convergent evolution between lineages. Overall, our results support a prevalent role for epistasis and convergent evolution in shaping protein evolution across the tree of life. Suprisingly, simulations of an epistasis model based on changes in protein folding stability (the Stokes-shift model) appear to be unable to explain the spatially-clustered epistasis pattern observed in in proteins.

109M Transfer RNA genes experience exceptionally elevated mutation rates. *B. Thornlow*, J. Hough, J.M. Roger, H. Gong, T.M. Lowe, R.B. Corbett-Detig University of California, Santa Cruz, Santa Cruz, CA.

Transfer RNAs (tRNAs) are a central component for the biological synthesis of proteins, and they are among the most highly conserved and frequently transcribed genes in all living things. Despite their clear significance for fundamental cellular processes, the forces governing tRNA evolution are poorly understood. We present evidence that transcription-associated mutagenesis and strong purifying selection are key determinants of patterns of sequence variation within and surrounding tRNA genes in humans and diverse model organisms. Remarkably, the mutation rate at broadly expressed cytosolic tRNA loci is between 8.7 and 13.8 times greater than the nuclear genome average. Furthermore, evolutionary analyses provide strong evidence that tRNA loci, but not their flanking sequences, experience strong purifying selection, acting in direct opposition to this elevated mutation rate. We also find a strong correlation between tRNA expression levels and the mutation rates in their immediate flanking regions, suggesting a simple new method for estimating individual tRNA gene activity. Collectively, this study illuminates the extreme competing forces in tRNA gene evolution, and implies that mutations at tRNA loci contribute disproportionately to mutational load and have unexplored fitness consequences in human populations.

110M Efficient pedigree recording for fast population genetics simulation. *Kevin Thornton*[1], Jerome Kelleher[2], Jaime Ashander[3], Peter Ralph[4] 1) Ecology & Evolutionary Biology, University of California, Irvine, CA; 2) Big Data Institute, Oxford University, Oxford, England; 3) Ecology & Evolutionary Biology, University of California, Los Angeles, CA; 4) Department of Mathematics, University of Oregon, Eugene, OR.

We present a new approach to recording the entire history of a population during forward-time, individual-based population genetics simulation. Our approach uses extensions of the data structures in the *msprime* software to record the marginal genealogies of all individuals (including ancient samples when needed), eliminating the need to expicitly simulate mutations not affecting fitness. As another benefit, it is easy with these tools to initialize a forward-time simulation with prior history produced by efficient coalescent simulation. We show that recording histories substantially reduces the run-time complexity of two different forward simulation engines, providing run time improvements of up to 50-fold in our tests and the possibility of substantial reductions in memory requirements. We will also describe the API, new in *msprime* 0.5.0, that enables our simulation method. These new features generally enable storage of population-genomic information within the *msprime* data structures from third-party code. The method makes it possible to not only run substantially larger simulations, but also to record and analyze the resulting whole-population genealogies across entire genomes.

111M Patterns of neutral genetic diversity under background selection are impacted across time by population demography. *R. Torres*[1], M. Stetter[2], J. Ross-Ibarra[2], R.D. Hernandez[1] 1) University of California, San Francisco; 2) University of California, Davis.

Neutral genetic diversity across the genome is determined by the complex interplay of mutation, demographic history, and natural selection. While the direct action of natural selection is limited to functional loci across the genome, its impact can have effects on nearby neutral loci due to genetic linkage. These effects of selection at linked sites, referred to as genetic hitchhiking and background selection (BGS), are pervasive across natural populations and have been studied extensively in species such as humans and Drosophila melanogaster. However, only recently has work been focused on the consequences of demography on the patterns of neutral genetic diversity caused by selection at linked sites. Studies by Beissinger et al. 2016 and Torres et al. 2017 investigated the impact of demography on BGS in maize and humans but found contradicting patterns for the impact of demography on genetic diversity, with opposite differences observed between bottlenecked and

non-bottlenecked populations. In order to develop more clear intuition about the expected impact of demography on selection at linked sites and reconcile differences observed in the data, we conducted extensive forward simulations of BGS across a 2 MB region with 8 specific demographic models that incorporated both ancient and recent population bottlenecks with varying population size decreases and recoveries. We measured relative neutral genetic diversity (pi) and relative singleton density from the simulations across time and observed that the window of time in which these statistics are measured can lead to the qualitative differences that are observed in real data. For example, in simulations of ancient bottlenecks, relative neutral genetic diversity has time to recover and become greater in the bottlenecked population when compared to a population of constant size. For simulations of recent bottlenecks, the opposite case is observed, with relative neutral genetic diversity lower for the bottlenecked population when compared to a population of constant size. This demonstrates that demography dominates patterns of relative diversity in regions of BGS for populations recently suffering population bottlenecks. However, for populations where bottlenecks have occurred in the distant past, patterns of relative diversity under BGS approach their expectation at population equilibrium. We also recaptiulated these patterns when conducting simulations of BGS using the inferred demographic history of humans and maize. Together, these simulations illustrate the importance of considering a population's long and short term demography when studying patterns of diversity in regions of selection at linked sites and explain the discrepancies observed in previous studies.

112M space: a tool for dynamic pca exploration. N. Berkowitz[1], *D. Turissini*[1], E. Elyashiv[1], K. Rand[1], O. Schaedel[1], Y. Wang[1], E. Hong[1], C. Ball[1], K. Chahine[2] 1) Ancestry, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

We present SPACE, a PCA exploration tool that can be run from a web browser. Implemented in R using the package Shiny, SPACE allows users to dynamically scale, subset, and transform data with a simple interface. SPACE can incorporate user-defined metadata which can then be examined for a single sample or for a collection of samples solely by pointing and clicking.

113M Inference of demographic history provides insight into domestication and morphotype diversification in *Brassica oleracea*. *Sarah D. Turner*[1], Makenzie E. Mabry[1], J. Chris Pires[1], Timothy M. Beissinger[1,2] 1) Division of Biological Sciences, University of Missouri, Columbia, MO; 2) United States Department of Agriculture, Agricultural Research Service, Columbia, MO.

The vegetables comprising *Brassica oleracea*, commonly known as cole crops, can be categorized into six distinct and diverse morphotypes: kale, cabbage, Brussels sprouts, kohlrabi, broccoli, and cauliflower. In addition to being valued for their flavor, culinary, and nutritional attributes, these crops are an especially interesting model for domestication; wild mustard, a small, unpalatable plant, was selected to enrich different plant organs, producing an extreme range of phenotypic diversity in a single species. However, despite widespread use of these crops to demonstrate the power of human selection on crop domestication, the demographic history of *B. oleracea* remains poorly understood.

Using 119 resequenced genomes that encompass the major morphotypes in *B. oleracea*, we present a demographic analysis of the ancestral population sizes and the timing of population splits. Specifically, we investigated divergence times between leafy kales and diversified (e.g. heading and tuber-forming) morphotypes, and between broccoli and cauliflower, the latter of which is thought to be a relatively recent domesticate.

Consistent with previously published literature and literary references on the domestication of cole crops, we find evidence of a severe reduction in the population size of *B. oleracea* around 3.5-7 kya, corresponding to a likely domestication event. Following domestication, we observe a modest expansion of diversity in all morphotypes, suggesting a demographic trajectory that is similar, though less extreme, than maize. Interestingly, we also identify a split between broccoli and cauliflower in the recent past, ca. 500-1000 ya, matching prior evidence on the introduction of cauliflower to Europe. By elucidating the demographic history of *B. oleracea*, this work expands our understanding of how humans have influenced the evolution of cole crops, provides insight into Brassica evolution and, by observing patterns of recent population expansion, facilitates future crop improvement efforts.

114M Discovering the molecular basis of polygenic local adaptation in the large Scots pine genome. *J.S. Tyrmi*[1,2], T Pyhäjärvi[1,2], O Savolainen[1,2] 1) Ecology and Genetics, University of Oulu, Finland; 2) Biocenter Oulu, University of Oulu, Finland.

Scots pine (Pinus sylvestris) is a key species in many areas throughout its vast distribution range spanning from South-West Europe to North-East Eurasia and it also has significant economic value. Scots pine is an excellent species for studying molecular basis of local adaptation in natural populations, because they are characterized by high levels of genetic and phenotypic variation in their distribution and populations are known to be locally adapted. The genetic diversity of Scots pine has only been studied in the past using a small number of candidate genes as their large and repetitive genomes still prevents whole genome resequencing. Here we have for the first time created a large-scale dataset using targeted sequence capture to examine some 3000 gene areas. 12 populations with total of 120 samples throughout Scots pine distribution range

in Europe and Western Russia have been included in this analysis. The sampling includes two parallel environmental gradients, where temperature, precipitation and other environmental variables vary. Sampling along western gradient spans from South to North Europe and a second gradient in the east spans from South to North Russia. We have then applied multiple methods to uncover the genomic basis of local adaptation, including performing linear regression of population allele frequencies on latitude, landscape genomics methods such as BayEnv and other genome scans to detect selection, including pcadapt and bayescan. We have also searched for covariance patterns between allele frequencies of different loci (i.e. linkage disequilibrium), which is expected to arise if a trait contributing to local adaptation has polygenic architecture, strong diversifying selection and high levels of gene flow. Since our sampling includes two separate clines we have also been able to examine whether the same genes contribute to local adaptation in the different parts of its distribution range. As a result of these analysis we have identified many interesting putative genes contributing to local adaptation of Scots pine, including a several members of pentatricopeptide repeat gene family. Members of this gene family are known to have many important functions in plant physiology and have also been recognized as outliers in selection scans of other forest trees. The results of our study shed light on the relatively poorly understood genetic architecture of local adaptation in plants with extremely large genomes.

115M Inference of selective sweep parameters through supervised learning. *Ian Vasconcellos Caldas*[1], Andrew G. Clark[1,2], Philipp W. Messer[1] 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.

The evolutionary history of most natural populations happens over timescales that are not directly observable. Nevertheless, ongoing and past evolutionary events are expected to leave characteristic signals in patterns of genetic variation that can be used to make inferences about these events. One classic such pattern is a "selective sweep" produced by the increase in frequency of a positively selected allele, which is associated with distinct patterns of heterozygosity and linkage around the selected locus. However, "hard" sweeps that start from a single adaptive mutation and "soft" sweeps that start from standing genetic variation or recurrent adaptive mutations have very distinct genetic footprints. For this reason, most methods that identify one kind of sweep are underpowered for other kinds, and the ideal of the all-encompassing scan for positive selection remains elusive. There remains a need to better characterize different varieties of sweep events in order to treat them as distinct evolutionary processes. Here, we present a method capable of inferring sweep parameters, including the "softness" of sweeps from standing genetic variation and recurrent de novo mutation, as measured by the initial allele frequency when selection started and the rate of adaptive mutation at the locus, respectively. We apply our method to a population of the malaria parasite P. falciparum, sequenced through the Pf3k project. This population has been under strong selective pressure to develop resistance against anti-malarial drugs such as artemisinin and has know resistance loci that have experienced hard and soft sweeps. We use extensive simulations to generate hypothetical evolutionary histories while exploring the space of demographic scenarios and selection parameters. A suite of selection scans is then applied to these simulations and the results are fed to a supervised statistical learning framework. We demonstrate how this method allows us to accurately infer sweep parameters of interest. For example, we perform the estimation of the initial allele frequency of a sweeps from standing genetic variation. Our method gains power by combining the idiosyncrasies of multiple selection scan methods and an extensive set of heterogeneous evolutionary scenarios generated in silico. These findings showcase how supervised statistical learning on population genomic data sets can help us better understand the evolutionary past by observing the present.

116M Phylogenetic analysis and the impact of heavy metal contaminants on wild isolates of the ubiquitous ectomycorrhizal species *Cenococcum geophilum*. *Jessica Velez*[1], Chris Schadt[2], Reese Morris[2] 1) University of Tennessee Knoxville, Knoxville, TN; 2) Oak Ridge National Laboratory, Oak Ridge, TN.

Agriculture has become a multifaceted industry as the production of biofuels and the need for greater sustainability grow in prevalence. There is a finite amount of land available for either biofuel or food crop systems, and land availability is further limited due to the presence of major pollutants including heavy metal contaminants, leading to phytotoxicity and other concerns. Plants serve as hosts to a diverse community of microbes including fungi, which are capable of metabolizing and/or immobilizing soil compounds which the plant cannot. Some mycorrhizal fungi directly interact with heavy metal contaminants within the soil, and a fungal-plant symbiosis may increase a plant's ability to survive in a soil containing high heavy metal concentrations. These relationships are often difficult to disentangle as a single plant species may be associated with hundreds of fungal species and each of these associations can display varying influences on host plant survival and growth that co-vary with the environmental and physical conditions of soil. One such symbiont, the *Cenococcum geophilum*, is ubiquitously distributed across multiple climates, soil types and plant species. The genome of *C. geophilum* is among the largest in the fungal kingdom, with a total estimated size of 178 megabases due to an abundance of transposable elements. This large size and highly repetitive genomic data have historically increased the difficulty of both sequencing and analyzing *C. geophilum*. New technologies, such as restriction enzyme associated DNA sequencing (RADseq) decrease both time and financial investment required for such analyses, and will be used to analyze over 200 wild isolates of *C. geophilum* obtained from soil samples across a range of 296 miles in the United States Pacific Northwest. These isolates will be used to determine

if established host population structure will be reflected in the symbiont genomes, despite the generalist nature of *C. geophilum*. These wild isolates will also be screened to determine phylogenomic related correlates with the uptake and growth effects of heavy metal contaminants lead, strontium, copper, cadmium and zinc using laboratory growth assays. If increased resistance is observed, this type of manipulation of host plant symbionts may increase plant tolerance to soil conditions that would otherwise be toxic to the plant, increasing overall land availability for use in agriculture. This would encourage the planting of biofuel crops within fields that may not be optimized for the growth of crops intended for consumption due to the contamination within the soil, thereby avoiding the “food-for-fuel” tradeoff that has driven agronomic policy concerns in the biofuel industry.

117M Haplotype-Based Method for Analyzing Ancient DNA. *S. Waldman*, S. Carmi, The Consortium for the Genetic Study of ancient Israel Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, IL.

Background: The ability to genotype individuals who have lived thousands of years ago has revolutionized our understanding of human evolution. Several methods are available for the population-genetic analysis of ancient DNA; however, no current method can exploit the information in haplotype sequences. Haplotype-based methods that were developed for modern DNA cannot be applied to ancient DNA due to its typical very low coverage.

Methods: We developed a method to compute haplotype similarity directly from low coverage data, even when only a single allele is reported for each SNP. Our approach is based on *ChromoPainter*, a popular tool for representing haplotypes as mosaics of other chromosomes, currently available only for high-quality genotypes. In our modified *ChromoPainter*, we define a set of modern "donor" populations, and we model the ancient genomes as a mosaic of the donor populations' genomes. To understand the origin of modern populations, we adapted a previously published regression technique, whereby the *ChromoPainter* profile of a target population is modeled as a linear mixture of the profiles of potential source groups. The mixture coefficients are interpreted, in our case, as the proportion of a modern population’s gene pool that was inherited from each ancient population.

Results: We used our method to study the relation between modern Middle-Eastern populations and ancient populations of the Levant, Iran, and the Caucasus. Our data included ~10 unpublished genomes from the Bronze and Iron Age Levant. The results demonstrated a high proportion of Bronze-Age Levant ancestry in all modern Middle-Eastern populations examined.

Conclusions: We developed a new method for the analysis of ancient haplotypes. Our method is expected to be highly informative on recent demographic events connecting ancient and modern populations.

118M Selective Sweeps in an Endemic *Drosophila serrata* Population: Insights from 110 Genomes . *Y. Wang*, A. Reddiex, S. Chenoweth School of Biological Sciences, University of Queensland, Brisbane, Queensland, AU.

Understanding how natural selection operates in natural populations is a fundamental goal of evolutionary genetics. At a population genomic level, when positive Darwinian selection occurs at a variable site, genetic variation is reduced in the surrounding region; a phenomenon known as a selective sweep. Selective sweeps are classified as *hard* or *soft* depending on whether selection acts on new beneficial mutations or standing variation. We have analysed selective sweeps across the genome of the montium *Drosophilid*, *D. serrata*, sampled from an endemic population at the centre of the species Australian distribution. Using a newly created 110 genome reference panel for this species, the DsGRP, we applied iHS and H12 tests to SNP data to understand how natural selection might affect this species. In the analysis, we used *p*iHS = -log[1-2|ø(iHS)-0.5|] as a statistic to perform the iHS scan. After testing 5,984,590 SNPs, we identified 1252 significant SNPs belonging to 362 peaks using threshold of *p*iHS = 4.66. The H12 approach, which is applied at a genomic window, rather than individual SNP resolution detected 32 significant peaks (H12 > 0.022, by simulation) among 117,564 windows. They were distributed on 4 chromosome arms: 14 on 2L, 4 on 2R, 8 on 3L, 6 on 3R. The reported range of H12 values was 0.022-0.081, with the median 0.025. The ratio H2/H1 can be used to infer sweep softness, where H2 is haplotype homozygosity calculated using all but the most frequent haplotype in a predefined window, and H1 is the haplotype homozygosity calculated using all haplotypes. The H12 values of maximum, minimum, median and mean in this study were 0.8913, 0.1468, 0.6796 and 0.6261, respectively, which suggested soft rather than hard sweeps. We recovered 247 genes within these sweeps and GO term enrichment analysis indicated an overrepresented biological process of “imaginal disc-derived appendage development” and “sensory organ development”. We then compared the results of *p*iHS and H12, and found 14 of the 32 H12 peaks (43.8%) containing SNPs with *p*iHS > 4.66, which indicated varying agreement between the two methods. Although we did not detect some of the classic sweeps seen in *D. melanogaster* (e.g. *CYP6g1*), our study has shed light on the potential targets of natural selection in this species.

119M Divergent fine-scale recombination landscapes between populations of threespine stickleback fish. Alice Shanfelter, *Michael White* Genetics, University of Georgia, Athens, GA.

Homologous recombination is a highly conserved process throughout the tree of life that is essential for the proper segregation of chromosomes during meiosis. At a fine-scale, recombination rates vary drastically across the genome, often localized into recombination "hotspots," surrounded by regions of the genome with little to no recombination. How recombination hotspots are targeted across the genome appears to be variable across species. In mammals, recombination hotspots are localized by the rapidly evolving histone methyltransferase, PRDM9, resulting in highly divergent localization of hotspots between closely related species. In many taxa outside of mammals, locations of recombination hotspots are conserved over long evolutionary timescales. Recent surveys suggest rapidly evolving PRDM proteins may exist outside of mammals, but it remains unclear whether these taxa also have hotspots with high rates of turnover. In addition, it is unclear whether high turnover of hotspots only co-occurs in taxa with rapidly evolving PRDM9 proteins. Threespine stickleback (*Gasterosteus aculeatus*) fish are an excellent model system to explore the evolution of recombination hotspots over short evolutionary timescales. Multiple isolated freshwater populations of threespine stickleback fish exist throughout the Northern Hemisphere, allowing for the study of parallel evolution of recombination hotspots over the last 15,000 years of divergence. Here, we apply a linkage disequilibrium-based method to estimate fine-scale recombination rates across the genome of a freshwater and marine population of *G. aculeatus*. We find less than 20% of hotspots are shared between the two populations, which suggests hotspot location may be diverging at a high rate in threespine stickleback fish. Contrary to other systems, we do not find evidence of positive selection within PRDM9. Our results indicate alternative mechanisms may exist that lead to high turnover of recombination hotspots and argue that additional surveys are necessary across taxa to understand the diversity of mechanisms responsible for hotspot targeting.

120M Mitochondrial heteroplasmy through the lens of population phylogenetics. *P.R. Wilton*[1], A. Zaidi[2], K. Anthony[2], B. Arbeithuber[2], A. Nekrutenko[2], K.D. Makova[2], R. Nielsen[2,3] 1) Department of Integrative Biology, University of California, Berkeley, Berkeley, CA; 2) Department of Biology, Penn State University, University Park, PA; 3) Department of Statistics, University of California, Berkeley, Berkeley, CA.

The mitochondria within the body form a network of populations of mitochondrial genomes (mtDNA) undergoing genetic drift, mutation, and natural selection. Acting during different developmental and reproductive stages, these population-genetic forces shape distributions of mtDNA polymorphism within individuals, called heteroplasmy. As the primary mode of inheritance of mitochondrial disease, heteroplasmy plays an important role in human health, and thus it is important to understand the processes by which heteroplasmy is inherited and comes to be distributed throughout the body.

We recently developed a population-phylogenetic model of mtDNA evolution during different reproductive and developmental stages. In our model, heteroplasmy frequencies change due to genetic drift and mutation along different branches of an ontogenetic phylogeny. Applying this model to previously published heteroplasmy data from mother-child pairs, we described a severe effective bottleneck comprised of the genetic drift occurring between the divergence of germline and somatic cells in the mother and the separation of germ layers in the offspring, as well as subsequent, less severe bottlenecks that occur during somatic development.

Here we present a new, larger analysis of heteroplasmy in blood and buccal tissues of 363 individuals from 101 multi-generational pedigrees, including 212 mother-child mtDNA transmissions, using a novel, state-of-the-art approach to detecting low-frequency mtDNA variants. Adapting our model and new variant-detection algorithms to these larger pedigrees, we find we have greater power to pinpoint when during development and oogenesis heteroplasmy frequencies change and can more precisely quantify the population-genetic forces bringing about these changes. We discuss the insights into mitochondrial inheritance and developmental proliferation generated by our analysis, including new theoretical assessments of the probability that a mother will pass on a disease-causing mitochondrial variant to her offspring. We conclude with a discussion of a new application of our ontogenetic phylogeny framework to modeling natural selection of mtDNA within the body.

121M Probing the genomic signatures of insecticide resistance in the malaria vector *Anopheles gambiae* via deep learning. *A. Xue*[1], D. Schrider[2], A. Kern[1] 1) Rutgers University, Piscataway, NJ; 2) University of North Carolina at Chapel Hill.

Large gains have been made over the past two decades in decreasing the rate of malaria transmission through control of its mosquito vector, *Anopheles gambiae*. Unfortunately though, these efforts are in danger of collapse due to the evolution of insecticide resistance. We aim to infer the genetic basis of current adaption to vector control efforts through the development of a novel supervised machine learning method to classify partial, incomplete sweeps from population genomic data. Accurate detection of selective sweeps is an important goal for population genomics, and identification of partial sweeps, including between hard sweeps from *de novo* mutations and soft sweeps on previously standing genetic variants, gives valuable information into ongoing evolutionary responses. Specifically, reliable discovery of partial sweeps should provide greater temporal resolution for recent history as well as possible insight into future population dynamics (*e.g.* the nascent

spread of a resistance allele). Here, we utilized a deep learning approach that uses convolutional neural networks (CNNs), coupled with coalescent simulations that incorporated population-specific demography, in order to distinguish among neutral sequences, completed hard sweeps, completed soft sweeps, partial hard sweeps, partial soft sweeps, and respectively associated linked regions that are not directly selected. First, we performed several simulation experiments to demonstrate the power of our CNN approach as well as assess the accuracy of different CNN architectures within the context of partial sweep classification. Subsequently, we applied our CNN classification to whole genomes from nine mosquito populations sampled across sub-Saharan Africa by the Ag1000G Consortium. This study elucidates loci that have experienced recent selection for insecticide resistance, both in specific geographic regions as well as continent-wide, providing valuable genomic resources that will aid in disease management and public health. More broadly, the success of our machine learning approach introduces a viable method to categorize partial versus completed as well as hard versus soft sweeps under a variety of demographic scenarios. This addresses an increasing demand among population genomicists for powerful selection scan tools as whole-genome data rapidly accumulate for a greater diversity of organisms.

122M Characterizing adaptive Neanderthal introgression using ancient and modern population genomic data. *S. Yair*, K.M. Lee, G.M. Coop Deparment of Evolution and Ecology, University of California Davis, Davis, CA.

Hybridization between humans and Neanderthals has resulted in the persistence of Neanderthal DNA in present day non-African populations. While on average Neanderthal-derived alleles have been selected against, a few exist at higher frequency in present-day non-African populations and reflect cases of adaptive introgression. Previous work has identified the genomic regions and present-day populations in which Neanderthal alleles adaptively introgressed. However, we still know little about the context in which these Neanderthal-derived alleles rose in frequency. Here, we present a model-based framework that uses both ancient and present-day samples of modern humans to elucidate the history of adaptive introgression in present-day populations. We model how a sweep placed along different branches of an admixture graph, which includes ancient populations, acts to modify the variance and covariance in neutral allele frequencies among populations at linked loci, relative to neutral population structure. By incorporating ancient DNA, we can determine in which ancient population(s) the Neanderthal alleles were favored, how migration among populations facilitated the spread of the alleles, and whether selection on a specific Neanderthal allele occurred once or independently across populations. These results allow us to distinguish between selection immediately after introgression and selection on standing introgressed variation. Our flexible method can be extended to any system with a complex admixture and demographic history to understand the spatiotemporal history of adaptation.

123M Mito-nuclear ancestry interactions contribute to mitochondrial DNA copy number variation in admixed populations. *A.A. Zaidi*, K.D. Makova Department of Biology, The Pennsylvania State University, University Park, PA.

Despite the fact that mitochondria carry their own genome (mtDNA), many of the proteins required for mitochondrial function in oxidative phosphorylation, as well those required for transcription and replication of mtDNA, are encoded by the nuclear genome. Thus, a compatibility between the mitochondrial and nuclear genomes is essential for proper mitochondrial function. It is reasonable to expect that mito-nuclear interactions would lead to some degree of mito-nuclear co-evolution across diverging human populations, especially given the high rate of mtDNA evolution compared to the nuclear genome. Thus, discordance in ancestry between mitochondrial and nuclear genes could contribute to phenotypic variation in recently admixed populations. Here, we explore phenotypic consequences of mito-nuclear ancestry interactions in human admixed populations using publicly available data from the 1000 Genomes Project. Based on low-coverage sequence data, we calculated mtDNA copy number in lymphoblastoid cells lines of 361 individuals of mixed West African, European, and Native American ancestry. MtDNA copy number is a cellular phenotype known to be associated with many health-related outcomes. We tested whether the interaction between nuclear ancestry and mtDNA haplogroup has an effect on mtDNA copy number. We find that increasing discordance between nuclear ancestry and mitochondrial ancestry is negatively correlated with mtDNA copy number in admixed populations (Slope = -122.34, t-statistic = -3.72, p-value = 9.80 x 10-05) and the direction of this effect is consistent across different mtDNA haplogroups. This might be indicative of incompatibility between proteins, involved in replication and regulation of mtDNA, which are coded separately by the nuclear and mitochondrial genomes. In addition to the pattern observed with global ancestry, we investigate mito-nuclear ancestry interactions at specific loci in the nuclear genome to identify genes that might be involved in regulation of mtDNA copy number. Our results highlight the importance of studying mito-nuclear interaction to better understand the risks involved in mitochondrial replacement therapy and prevalence of common disease risk in humans.

124M An evolutionary history of the *C. elegans* species. *S. Zdraljevic*[1,2], D.E. Cook[1,2], R.E. Tanny[2], E.C. Andersen[1,2,3] 1) Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA; 2) Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA; 3) Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL 60611, USA.

The genome of *Caenorhabditis elegans* has been shaped by recent chromosome-scale selective sweeps. Genomic evidence for these sweeps exists in 147 of the 152 *C. elegans* strains that have been characterized to date. The high prevalence of these sweeps within the *C. elegans* population has obscured efforts to reveal the origins and evolutionary history of this species. However, through the concentrated efforts of researchers and citizen scientists around the world, we have amassed a larger collection of individuals from the *C. elegans* species. Here, we present an updated collection of 249 *C. elegans* strains, for which we have characterized genetic variation and population structure.

Population genomic analyses suggest that the *C. elegans* species is derived from 10 subpopulations. We find evidence for substantial admixture among these subpopulations, including support for at least four migration events. Analysis of representative non-admixed individuals from each subpopulation suggests that geographic clustering of related individuals exists within the species, which has not previously been identified. Furthermore, one subpopulation that consists of one admixed and seven non-admixed individuals contains 58% of the total genetic variation found in the species. Six individuals from this divergent population were isolated on the Hawaiian Islands, one from California, and one from New Zealand. These results lend further support to the hypothesis that the ancestral state of the *C. elegans* species can be found on the Pacific Rim.

Our characterization of the *C. elegans* subpopulations enables the detection of alleles that could have contributed to the chromosome-scale selective sweeps. To test for signatures of selective sweeps, we performed a genome-wide composite likelihood ratio test on each of the identified subpopulations. This analysis identified 13 genomic regions with strong support for a selective sweep. Interestingly, different combinations of loci showed signs of selective sweeps within subpopulations. One genomic region with a strong signature of selection is at the *sup-35*/*pha-1* locus on chromosome III. A signature of selection at this locus is in agreement with a recent study that identified it to be a selfish genetic element that results in a genomic incompatibility between individuals in the *C. elegans* species. Taken together, these results suggest that population genomic analyses can be performed to identify alleles that have shaped the *C. elegans* genome.

125M The evolutionary history of supergene mimicry in *Papilio* swallowtail butterflies. *W. Zhang*[1,2,5,6], E. Westerman[2,3], E. Nitzany[4], S. Palmer[4], M. R. Kronforst[2] 1) School of Life Sciences, Peking University, Beijing, China; 2) Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA; 3) Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA; 4) Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA; 5) Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China; 6) State Key Laboratory for Protein and Plant Gene Research, Peking University, Beijing, China.

Papilio swallowtail butterflies in the *polytes* species-group display a variety of mimicry forms. The genetic basis of the mimicry supergene in *Papilio polytes* was characterized at a molecular level yet little is known about the evolution of supergene mimicry. To answer this long-standing question, we analyze whole-genome sequencing data from *P. polytes* and other species in the *polytes* group. Our results suggest that sexual dimorphism and female-limited polymorphism are a result of ancient balancing selection together with independent origins of similar morphs in different lineages and secondary loss of polymorphism in other lineages. In addition, both natural selection and genetic drift might have contributed to the secondary loss of polymorphism. We also show that extensive genetic hitchhiking has burdened the mimetic haplotype at the origin of mimicry, which partially explains the coexistence of non-mimetic and mimetic female morphs. Our results change our understanding of mimicry by revealing the origin of supergene mimicry, the connection between sexual dimorphism and polymorphism and the history of supergene evolution.

126M Maintenance and evolutionary potential of antibiotic resistance in a long-term experiment with *E. coli*. *K.J. Card*[1,2,3], T. LaBar[1,2,3], J.B. Gomez[1,4], R.E. Lenski[1,2,3] 1) BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI; 2) Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI; 3) Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, MI; 4) Biomedical Laboratory Diagnostics Program, Michigan State University, East Lansing, MI.

Populations often encounter environmental changes that remove selection for the maintenance of previously essential traits. Adaptation to a new environment can thus affect an organism's fitness in its prior environment. How evolutionary processes drive these correlated responses, and how these traits evolve once populations are introduced back into their original environment, are still not well-understood. We used *E. coli* strains from the long-term evolution experiment (LTEE) to study these questions. LTEE populations have been evolving for 30 years and over 60,000 generations in a medium without antibiotics. Thus, we first addressed how this prolonged period of relaxed selection affects the maintenance of antibiotic resistance traits. To that end, we performed extensive resistance measurements on the ancestral and evolved strains from the LTEE. On balance, resistance traits have tended to decay in the absence of drugs, such that the evolved strains are more susceptible than their common ancestor. Evolution in the drug-free environment of the LTEE thus trades off with fitness in drug-containing environments. Next, to see whether the bacteria could overcome these tradeoffs, we then evolved the

ancestral and more-sensitive LTEE-derived strains in the presence of antibiotics. In general, both the ancestral and derived strains gained resistance, although often to different degrees. Our results suggest that complex traits decay in the absence of selection, but these traits can be re-evolved upon further environmental change back to the ancestral state.

127M Impact of ploidy on pathogenesis depends on host immune status. *M.A. Hickman*, Dorian Feistel, Nancy Nguyen, Rema Elmostafa Biology, Emory Univ, ATLANTA, GA.

The yeast *Candida albicans* is a commensal of humans and opportunistic pathogen. Given that *C. albicans* can transition between benign commensal and harmful pathogen, studying the factors that contribute to virulence is often challenging and frequently depends on many contexts including host immune status and pathogen genetic background. Importantly, clinical isolates, including sets of strains isolated from single individuals vary in ploidy either spatially (from distinct host niches) or temporally (over the course of infection). Here, we utilize the nematode *C. elegans* as a mini-host system to study how pathogen ploidy and genetic background impact fungal infection. In addition to reducing overall survival of nematode hosts, *Candida* infection also results in delayed reproduction, which results in significant reductions in population size over many generations. In healthy host contexts, there is no clear correlations between ploidy and pathogenesis: in some *C. albicans* genetic backgrounds diploids are most virulent and yet in others, tetraploids display higher virulence. Intriguingly, when we evaluate pathogen potential in an immunocompromised host context we observe that some ploidies are now significantly more virulent compared to their infections in healthy hosts. Taken together, our work suggests that there are significant interactions between pathogen ploidy and genotype with host immune function that contribute to pathogenesis.

128M Experimental evolution of collective action despite genetic conflict and free riding in a bacterial biofilm. *neal jahren*, jessica dewey, sarah m hong, michael travisano ecology, evolution & behavior, university of minnesota, saint paul, MN.

Genetic uniformity is sometimes theorized as a requirement for reducing conflict between cells in a multicellular organism, but recent interest in the microbiomes associated with multicellular organisms and the holobiont concept have motivated hypotheses that multicellular organisms evolved from the context of multispecies communities. We have induced the physical clustering of *Pseudomonas fluorescens* cells by subjecting populations to settling selection. Clustering emerges through the production of an extra-cellular matrix (ECM), but the populations evolve towards a long-term coexistence of strains with distinct adaptations rather than a selective sweep of one highly adapted strain. Closer inspection of the clusters reveals the presence of a “smooth” strain that does not produce ECM but is able to survive the settling selection by free riding. Comparison with the ancestral strain found that the ability for smooth strains to insinuate into ECM-mediated clusters is an evolved response. Thus, these clusters provide a system to investigate how collective action can evolve in communities of genetically distinct cells and survive challenges to collective action such as free riding. Previous research into the collective action of *P. fluorescens* has focused on the pellicle that grows on the surface of a static culture. Pellicle function is compromised by free-riding smooth strains and pellicle-forming strains perform poorly in shaking culture. A subset of ECM-producing strains that evolved in shaking culture under settling selection also demonstrate adaptation to static culture by outcompeting a pellicle-forming strain. Evidently, the collective action that emerged at a small scale to survive settling selection also provides an effective strategy for large scale co-operation in a pellicle, possibly by avoiding a trade-off that constrains pellicle-forming strains. These results provide evidence for the existence of a new mode of collective action in bacterial biofilms.

129M Transcriptome evolution of a fungal plant pathogen using an experimental evolution approach. *A. Jallet*, A. Genissel UMR Bioger, Inra, AgroParisTech, Paris-Saclay University, Thiverval-Grignon, France.

One interesting question in evolutionary biology is to understand how genome regulation contributes to environmental adaptation. Towards this goal we performed an experimental evolution by maintaining *in vitro* two clones of the fungal wheat pathogen *Zymoseptoria tritici* under fluctuating or stable temperatures for 50 weeks. Whole transcriptome of ancestors and evolved populations were assessed at the two stable temperatures used during the experimental evolution. Our results show a large effect of genetic background and temperature on the abundance of transcripts, with respectively 72 % and 9 % of the total phenotypic variance explained. A model based on a negative binomial distribution using read counts showed a strong response to the selection, with more than 2000 genes differentially expressed among the regimes (ancestral state, stable, fluctuating). However gene expression profiles in response to selection were very different between the two genetic backgrounds. Results from the first genetic background reveal 2.3 % of the genes (254 genes randomly distributed throughout the genome) were differentially expressed between stable and fluctuating selection. A co-expression analysis using gaussian mixture models revealed 3 large clusters, comprising 665 genes with a different expression profile between fluctuating selection and ancestral state, suggesting a transcriptome rewiring associated with fluctuations. Gene ontology enrichment analysis on these top-three clusters showed enrichment for cell division and cell wall biosynthesis categories, consistent with fluctuating selection having a role in these two biological processes. For the second genetic background, we did not detect any genes with transcriptional profile change showing significant differences with their ancestral state and between stable and fluctuating selection. We observed that the 421 genes with similar transcriptional change for stable and

fluctuating selection were enriched in GO terms related to cell cycle. Finally, an analysis on gene expression level plasticity in response to temperature revealed that more genes under fluctuations than genes under stable selection evolved towards a loss of plasticity when compared to their ancestors. These findings are consistent with models predicting that fluctuating selection can favor genetic canalization. This 'evolve-and-reseq' experiment provides new insights into the effect of fluctuating selection on genome regulation in fungi.

130M Uncovering the genotype-phenotype-fitness map of microbes adapting to novel environments. *Grant Kinsler*, Kerry Geiler-Samerotte, Dmitri Petrov Dept. of Biology, Stanford University, Stanford, CA.

Experimental evolutions using DNA barcodes to track millions of independent evolving lineages have recently quantified the spectrum of unique single mutations that can each help microbes adapt to glucose limitation. But how many unique physiological processes are represented by this large pool of beneficial mutations? How many unique ways are there to improve microbial fitness in glucose-limited conditions, or in other novel environments? Using recent developments in DNA barcoding, we precisely measure the fitness of evolved lineages in the environment they evolved in as well as many environments that only slightly differ from this original condition. These data allow us to understand the pattern of correlations amongst adaptive mutants across subtly differing conditions, estimate the number of unique fitness-relevant phenotypes represented by these mutants, and uncover the genotype-phenotype-fitness map for these adaptive lineages. We find evidence that single genetic mutations affect many phenotypes (pleiotropy). Despite this observation, we show that only a small number traits matter for adaptation to the original evolution condition. In particular, our finding sheds light on how adaptation can proceed despite widespread pleiotropy, the key being that not all phenotypes affected by mutation have fitness effects in the current environment. This has wide-ranging implications on how adaptation proceeds in complex phenotype space, specifically regarding the extent to which adaptation is limited by tradeoffs and how environmental dependencies influence the relationship between phenotype and fitness.

131M Drift robustness and the evolution of genome architecture in small populations. *T. LaBar*[1,2,3], C. Adami[1,2,3] 1) Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI; 2) Ecology, Evolutionary Biology, and Behavior, Michigan State University, East Lansing, MI; 3) BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI.

Adaptation is constrained in small populations by weakened natural selection and the increased loss of small-effect beneficial mutations through genetic drift. In other words, small populations can only undergo sustained adaptation if they are able to maintain the fixation of beneficial mutations. Therefore, small populations will either 1) undergo reduced adaptation compared to large populations, or 2) undergo adaptation through different genetic mechanisms than large populations that allow them to avoid the constraints of weakened selection. Here, we will discuss work on this second trajectory using the digital experimental evolution system Avida. We show that the population-genetic environment caused by small population size leads to the evolution of drift robustness, or a decreased likelihood of fitness loss due to drift, in small populations. Small populations evolve drift robustness by adapting to fitness peaks with a deficit of slightly-deleterious mutations and an excess of neutral and strongly-deleterious mutations. We will also show that this evolutionary drive towards drift robustness shapes genomic architecture in small populations beyond the distribution of fitness effects. Small populations adapt to drift-robust fitness peaks by fixing epistatic beneficial mutations. Additionally, small populations evolve greater genetic complexity than large populations as a mechanism to encode drift robustness. Our findings suggest a novel adaptive mechanism by which small populations can overcome weakened selection and may explain the trend towards certain genomic architectures in populations experiencing strong genetic drift across life.

132M Host-virus genome coevolution in laboratory populations of yeast. S. Buskirk, *G. Lang* Department of Biological Sciences, Lehigh University, Bethlehem, PA.

Nearly all genomes contain genetic parasites that replicate selfishly, often at a cost to the host genome, leading to evolutionary arms races between selfish genetic elements and their hosts. Most genomes, including the human genome, exhibit clear signatures of past intragenomic conflicts. Yet our understanding of intragenomic conflict is limited in that few systems exist to study the coevolution of multiple genomes in high temporal resolution.

The yeast *Saccharomyces cerevisiae* is host to a selfish intracellular "Killer" virus. The Killer virus is an encapsulated double-stranded RNA virus that encodes both a Killer toxin and its corresponding immunity component. An infected host secretes the toxin, which kills non-Killer-containing cells. We recently discovered that our laboratory strain — a strain that we have used extensively in experimental evolution — contains the Killer virus.

Previously we evolved ~600 replicate populations of our Killer-containing laboratory strain for 1,000 generations. We find that Killer toxin production is lost in about half of our populations. Of those populations that lose killing ability, about half develop sensitivity to the toxin. Using RT-PCR, we amplified and sequenced the viral genomes from 70 populations. We correlate changes in Killer phenotype to mutations arising in both the host and the viral genomes. We find that both genomes coevolve

in response to toxin production, acquiring mutations in host genes involved in toxin entry and in the virally-encoded toxin itself. Our results establish a tractable laboratory system for the study of host-virus genome coevolution.

133M Single cell copy number variant detection reveals the dynamics of adaptation. *S. Lauer*, G. Avecilla, P. Spealman, N. Brandt, D. Gresham New York University, New York, NJ.

Copy number variants (CNVs) are an important, but understudied source of genetic variation and adaptive potential. Recent studies have revealed the prevalence of CNVs in natural populations, including their role in human evolution and diseases such as cancer. Interestingly, many adaptive CNVs are important for sensing and responding to changes in the environment. In microbial populations under nutrient limitation, CNVs recurrently occur at loci that include genes encoding high-affinity transporters for the limited nutrient. However, existing methods are not sensitive enough to detect CNVs at low population frequency, and the dynamics with which these CNVs are generated and selected are poorly understood. To overcome this challenge, we developed a fluorescent reporter assay that allows single cell detection of CNVs as they arise *de novo* in evolving populations. We performed long term experimental evolution in *Saccharomyces cerevisiae* limited for various nitrogen and carbon sources and tracked duplication and deletion of the general amino acid permease, *GAP1*, with single cell resolution. We found that early duplication events are highly reproducible in independently evolving populations, but later dynamics are complex and CNVs exhibit variable longterm fates. Using a barcoding approach and fluorescence activated cell sorting (FACS), we quantified the extent of clonal interference among CNV lineages and found that multiple, independent CNVs arise within each evolving population. We observe that CNV lineages compete and replace one another across time, and hypothesize that the degree of clonal interference determines the overall success of CNVs in the population. Whole genome sequencing confirmed *GAP1* CNVs and revealed that these alleles are extremely diverse in copy number and size. To understand how these characteristics influence the observed population dynamics, we aim to quantify the fitness of individual lineages using a pooled approach. Together, our results demonstrate the utility of our novel reporter, which provides a high-resolution and quantitative view of CNV dynamics in evolving populations.

134M Investigating the role of population history and environment on germline mitochondrial mutagenesis. *Tess Leuthner*[1], Nathan Keith[2], Craig Jackson[2], Joel Meyer[1], Joe Shaw[2] 1) Nicholas School of the Environment, Duke University, Durham, NC; 2) School of Public and Environmental Affairs, Indiana University, Bloomington, IN.

Understanding the origin, persistence, and consequences of mitochondrial DNA (mtDNA) mutations are increasingly fascinating in both evolutionary biology and medicine, since recent evidence implicates heritable mtDNA mutations in population level phenotypes as well as individual health (such as cancer, ageing, etc.). We used an experimental evolution approach in the freshwater crustacean, *Daphnia pulex*, to investigate variation in germline mutation rates of two different populations: one population sampled from a mining-devastated region, and the other from a pristine environment. The genotype from the contaminated region is from an adapted population of *D. pulex* that is tolerant to cadmium (Cd) after a century of exposure to iron ore smelting runoff. We take advantage of this variation in population histories to investigate the effect of Cd on mutation rate, which is a known mutagen. Twelve genetically identical individuals of the tolerant ("T") population and a Cd-sensitive reference population ("S") of *D. pulex* were exposed to both control, and chronic, environmentally-relevant cadmium conditions for almost 2,000 generations via a mutation accumulation (MA) experiment, followed by deep-coverage, whole-generation sequencing. There is no effect of population history or Cd on overall mitochondrial genome mutation rates. However, there is a trend towards suppression of C:G > A:T transversion mutations and both A:T > G:C and C:G > A:T transition mutations in the "T" MA experimental lines. Strikingly, there are specific context-dependent mutations that occur at C/G sites in the "S" experimental lines, which are absent in the "T" MA experimental lines. We also observe variation in the frequency and region of insertion and deletion mutations, with "T" indels occurring more randomly throughout the genome, with "S" indels occurring at high frequency in the D-loop, or non-coding region, of the mitochondria. Overall, we see very low frequencies of mutations, ranging from 3 – 50% of the entire mitochondrial population, which suggests that either mutations are recent and not fixed, or that *D. pulex* retain low levels of heteroplasmy. One line of "S" in Cd shows a mutator phenotype, with a mutation rate >100-fold higher than the other experiments, which to our knowledge has not been observed in a mitochondrial genome. Overall, these results suggest variation in the spectrum of mtDNA mutations across different populations under variable environments.

135M Invade and conquer: the spread of *wtf* meiotic drivers in yeast. *Jose F. Lopez Hernandez*[1], SaraH Zanders[1,2] 1) Stowers Institute for Medical Research; 2) University of Kansas Medical Center.

Allele transmission is generally fair such that each allele is present in 50% of the gametes. However, this fairness is disrupted when alleles bias their transmission by subverting gametogenesis; this phenomenon is called meiotic drive. We recently identified the *wtf* poison + antidote meiotic driver genes in fission yeast. The *wtf* genes comprise a family with more than 20 genes per genome. These drivers act by poisoning gametes that do not inherit the driver allele. Gametes that do inherit the *wtf* driver are protected from the poison by an antidote that is also encoded by the *wtf* gene. We are using the *wtf4* gene to mathematically model how meiotic drivers spread in populations over time. We are experimentally testing the model's predictions using a high-throughput assay to measure allele frequencies. Interestingly, *wtf4* spreads to fixation in a

population much faster than predicted. This suggests that *wtf4* provides a competitive advantage in addition to its known role in meiotic drive. This advantage may help explain the expansion of the *wtf* family in fission yeast.

136M Sign of selection on mutation rate modifiers depends on population size. *Y. Raynes*[1], C.S. Wylie[1], P. Sniegowski[2], D. Weinreich[1] 1) Center for Computational molecular Biology, Brown University, Providence, RI; 2) Department of Biology, University of Pennsylvania, Philadelphia, PA.

The influence of population size (*N*) on natural selection acting on alleles that affect fitness has been understood for almost a century. As *N* declines, genetic drift overwhelms selection and alleles with direct fitness effects are rendered neutral. Often, though, alleles experience so called indirect selection, meaning they affect not the fitness of an individual but the fitness distribution of its offspring. Some of the best studied examples of indirect selection include alleles that modify aspects of the genetic system such as recombination and mutation rates. Here we use analytics, simulations and experimental populations of *S. cerevisiae* to examine the influence of *N* on indirect selection acting on alleles that increase the genomic mutation rate (mutators). Mutators experience indirect selection via genomic associations with beneficial and deleterious mutations they generate. We show that as *N* declines, indirect selection driven by linked beneficial mutations is overpowered by drift before drift can neutralize the cost of the deleterious load. As a result, mutators transition from being favored by indirect selection in large populations to being disfavored as *N* declines. This surprising phenomenon of sign inversion in selective effect demonstrates that indirect selection on mutators exhibits a profound and qualitatively novel dependence on *N.*

137M Dynamics of second order lineage competition revealed by recursive lineage tracking. *J.I. Rojas Echenique*, A.N. Nguyen Ba, I. Cvijovic, K. Lawrence, M.M. Desai Harvard University, Cambridge, MA.

In an evolving asexual population, the fate of a new mutant lineage is determined both by its intrinsic capacity to leave more offspring than other lineages and by its offspring's potential to acquire further beneficial mutations. Thus, it's possible that mutations could be selected for their effects on evolvability rather than their direct effects on fitness. Here we develop a lineage tracking technique that allows us to directly observe the dynamics of both of these facets of lineage competition: first order selection and the second order effects of additional beneficial mutations. Our technique records the lineage history of every cell in an evolving population genetically by the sequential addition of random DNA barcodes into a time sorted barcode array; earlier barcodes identifying more distant ancestors. Deep sequencing of this growing array allows us to infer the frequency and evolutionary history of each lineage in the population. We've applied our technique to several controlled laboratory populations of budding yeast, We can trace lineage dynamics over a thousand generations of evolution, and directly observe second order competition between the sub-lineages of competing lineages.

138M Tempo, mode, and fitness effects of mutation in *Caenorhabditis elegans* over 400 generations of minimal selection. *A. Saxena*[1], M. Salomon[2], C. Matsuba[2], S. Yeh[1], C. Baer[1] 1) Zoology, University of Florida, Gainesville, FL; 2) John Wayne Cancer Center Institute, Santa Monica, California.

Variation in the mutational process is ubiquitous, from among individual to among taxa. Yet an empirical understanding of causal factors is rudimentary, even less so is the relative contribution of genetic and environmental factors. Stress or physiological condition has been implicated as a source in variation in microbial mutation rates, and recently in fruit flies.

Deleterious mutations are an intrinsic source of stress. We tested if the mutational process is conditioned on mutational load in two groups of *Caenorhabditis elegans* mutation accumulation (MA) lines. "First-Order MA" (O1MA) lines, propagated through single population bottlenecks across ~250 generations of minimal selection were divided into high-fitness and low-fitness groups. "Second-order MA" (O2MA) lines derived from high or low fitness O1MA ancestors, were maintained for an additional ~150 generations. Genomes of 48 O2MA lines and their O1MA progenitors were sequenced. There is significant variation in base substitution rates among O2MA lines, but no effect of starting fitness, whereas the indel rate in high fitness group is greater. Base substitution rate is positively correlated with recombination rate and negatively correlated with 1-Kb GC content, but multiple logistic regression shows mutability is sufficiently predicted by the three-nucleotide motif without recombination rate or GC content. ~90% of the variance in standing nucleotide variation is explained by mutability. Short tandem repeat mutation rate is observed to be an order of magnitude lower than previous estimates. The total mutation rate increased slightly in the O2MA lines, consistent with the "drift barrier" model of mutation rate evolution. Combined with previous experimental estimates of fitness, we infer that epistasis is synergistic on average.

139M Effects of relaxed selection on experimentally evolved populations of *Drosophila melanogaster*. M. Phillips[1], G Rutledge[1], A Talbott[2], S Matty[2], *P Shahrestani*[2] 1) Biological Science, California State University Fullerton, Fullerton, CA; 2) Ecology and Evolutionary Biology, University of California Irvine, Irvine CA.

Experimental evolution and next generation sequencing studies featuring *Drosophila melanogaster* suggest that adaptation in sexual populations is fueled by selection on standing genetic variation, is largely characterized by a lack of fixation, and is highly repeatable using groups of long-standing and newly derived populations subject to the same selection regimes.

However, there remain questions regarding the role of evolutionary history given these dynamics. Previous studies suggest that phenotypes and patterns of genetic variation are primarily shaped by most recent selection regime, and evolutionary history has little impact. But, while evolution may be predictable at the phenotypic level, differences in where experimental populations ultimately originate may have significant effects on outcomes at the genetic level. Reverse experimental evolution studies suggest that the degree to which populations return to ancestral allele frequencies is at least in part contingent on evolutionary history. Intense selection regimes may lead to greater losses of genetic variation. Therefore, the finding that phenotypes and patterns of genetic variation are almost exclusively shaped by most recent selection regime in previous studies could be due to the fact none of the populations studied were exposed to sufficiently intense selection regimes in their evolutionary histories. Here we seek to test this hypothesis using *D. melanogaster* populations that were once subjected to intense selection for desiccation resistance. We examine patterns of phenotypic and genomic differentiation in two five-fold replicated stocks: populations that were selected for desiccation resistance for 260 generations followed by 200 generations of relaxed selection, compared to control populations. The extreme functional differentiation previously seen between these two groups was achieved using intense selection pressures. Assuming large impacts of evolutionary history are in fact due to exposure to intense selection, we would expect to find lingering phenotypic and genomic differentiation between the formerly desiccation resistant populations and their controls. Instead we find convergence of most phenotypic traits after relaxed selection, with the exception of longevity, and we find convergence of the genomes, with the exception of seventeen loci. The lingering differentiation between these population types cannot be explained by fixation events during selection. On the surface, these findings support the idea that phenotypes and patterns of genetic variation are primarily shaped by most recent selection regime in outbred sexually reproducing populations.

140M Dormancy constrains the rate and direction of adaptive evolution. *William R. Shoemaker*, Jay T. Lennon Department of Biology, Indiana University, Bloomington, IN.

In nature, most populations experience some degree of energy limitation, which contributes to sub-optimal growth and reproduction. Many organisms contend with energy limitation by engaging in dormancy, a life-history strategy that allows individuals to enter a reversible state of reduced metabolic activity. Despite its prevalence, we still have a poor understanding of how dormancy affects evolution. Here we report the results from experimental evolution across varying degrees of energy-limitation using *Bacillus subtilis* populations that differ in their ability to enter a dormant state. Using pooled population sequencing, we identify how interactions between energy-limitation and dormancy alter the rate of evolution. In addition, because there are few available tools for quantifying divergent evolutionary trajectories between treatment groups and the degree of similarity between replicate independently evolving populations, we develop statistical approaches to describe the degree of parallel evolution between independently evolving replicate populations relative to a non-parametric null model. Using this approach, we determine that the ability to enter a dormant state results in a divergent evolutionary trajectory and an altered rate of evolution. These results support recent developments in population-genetic theory and suggest that dormancy constrains the rate and direction of evolution. In addition, because the energetic cost of dormancy suggests that it is likely subject to natural selection, its fitness effect and degree of constraint on the direction of evolution likely depends on the degree of energy-limitation in the environment. We identify the mutations within the set of genes that disproportionately contribute to evolutionary divergence and quantify their fitness effect (i.e., the fitness landscape). Using a combination of experimental evolution, population genetic theory, and advanced statistical approaches, we determine how dormancy as a life-history strategy constrains the rate and direction of evolution.

141M Adaptation of *Mycobacterium tuberculosis* during biofilm growth. *T.M. Smith*, T.D. Mortimer, M.B. O'Neill, C.S. Pepperell University of Wisconsin-Madison, Madison, WI.

Purpose. *Mycobacterium tuberculosis* (*M.tb),* the etiological agent of tuberculosis (TB), is the leading cause of death due to an infectious disease worldwide. Long-term treatment is necessary to cure TB due to bacterial subpopulations that persist during antibiotic treatment. Prior research indicates *M.tb* grows as a pellicle (a biofilm at the air-liquid interface) in open cavities of pulmonary lesions: this is likely to play a role in bacterial persistence. Traditional reverse genetic approaches have been used to identify a handful of genes involved in this biologically relevant growth state. As biofilm growth is a complex phenotype and likely to be controlled by a suite of genes, we have used experimental evolution to investigate the genetic basis of *M.tb* biofilm growth. We hypothesized that clinical isolates of *M.tb* serially passaged under the selective pressure to grow as a biofilm would gain spontaneous mutations in genes important for biofilm growth. **Methods.** We conducted 8 serial pellicle passages for 6 clinical *M.tb* isolates. gDNA was extracted and sequenced every 4 passages. We performed BWA-MEM based reference guided assembly (RGA) and called single nucleotide polymorphisms (SNPs) using GATK. We also used the breseq pipeline for RGA and to identify SNPs segregating within the mixed biofilm population. We then used PoPoolation2 to identify SNPs undergoing marked changes in frequency between passages (F_{ST} values > 0.10). We measured gene expression of candidate alleles in ancestral and evolved populations. We used Pilon to identify potential insertions, deletions, and duplications of interest. **Results.** We identified mutations in a transcription factor binding site affecting expression of the flavoprotein disulfide reductase *lpdA* in 2 distinct isolate lines belonging to the same lineage. In addition, we found a large genomic duplication (~150 kb, encompassing ~140 genes) in 2 isolate lines in different lineages. Excitingly, these are both

examples of convergent evolution. We also identified SNPs in 2-component regulatory systems. **Conclusion.** Our results support a role for alteration of gene dosage and expression in adaptation of *M.tb* to biofilm growth. *M.tb* biofilm development clearly relies on multiple genetic loci and interactions between these loci. The genetic determinants we identified are potential targets for developing therapeutic agents against these drug tolerant, persistent populations.

142M High accuracy haplotype-derived allele frequencies from ultra-low coverage pool-seq samples. *S. Tilk*[1], A. Bergland[1,2], A. Goodman[1], P. Schmidt[3], D. Petrov[1], S. Greenblum[1] 1) Department of Biology, Stanford University, Stanford, CA; 2) Department of Biology, University of Virginia, Charlottesville, VA; 3) Department of Biology, University of Pennsylvania, Philadelphia, PA.

Evolve-and-resequence experiments leverage next-generation sequencing technology to track allele frequency dynamics of populations as they evolve. While previous work has shown that adaptive alleles can be detected by comparing frequency trajectories from many replicate populations, this power comes at the expense of high-coverage (>100x) sequencing of many pooled samples, which can be cost-prohibitive. Here we show that accurate estimates of allele frequencies can be achieved with very shallow sequencing depths (*Drosophila melanogaster*, we show that haplotype inference can improve allele frequency accuracy by orders of magnitude, and that high accuracy is maintained after up to 200 generations of recombination, even in the presence of missing data or incomplete founder knowledge. By reducing sequencing costs without sacrificing accuracy, our method enables analysis of samples from more timepoints and replicates, increasing the statistical power to detect adaptive alleles.

143M Evolution in action: Understanding the phenotypes responsible for adaptation of the *C. elegans* reference strain N2 to its lab environment. *Y. Zhao*, R. Campbell, L. Long, W. Xu, P. McGrath School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA.

Evolution does not stop once we bring wild organisms into the lab for experimental study. An example of this is the N2 reference strain of *C. elegans,* used by nearly all *C. elegans* genetics labs, which was cultured for ~18 years before methods of cryopreservation were developed. Using a second lineage that split from N2 ~7 years after isolation from the wild, we have identified 94 new mutations that have fixed in the N2 lineage, including two in *npr-1* and *glb-5,* which were previously identified for their role in so-called social/solitary behavior differences. To understand the evolutionary forces responsible for their fixation, we used five to seven generation, pairwise competition experiments between the four possible combinations of ancestral/derived alleles of *npr-1* and *glb*-5 to measure the fitness effects of each allele. We found that the derived allele of *npr-1* increased fitness in both *glb-5* backgrounds while the derived allele of *glb-5* only increased fitness in the derived *npr-1* background, an example of Bateson epistasis that suggests the derived allele of *glb-5* fixed after *npr-1.* To test whether the derived solitary behavior of N2 was responsible for these gains in fitness, we performed environmental and genetic manipulations (10% O_2 levels, modifying food distribution, or ablating URX neurons) known to suppress social behaviors. Unexpectedly, we found N2 animals still had a fitness advantage over a strain containing ancestral alleles of *npr-1* and *glb-5*, indicating additional biological traits are responsible for fitness gains. RNAseq analysis of age-synchronized animals at both 10% and 21% O_2 identified a large number of spermatogenesis genes upregulated in N2, suggesting that N2 animals mature faster and produce more sperm regardless of behavioral state, which we confirmed using developmental timing and sperm counting experiments. Based upon the role played by orthologs of *npr-1* (NPY) in feeding, we hypothesized that the fitness gains could be due to increases in food consumption. We developed a new feeding assay to measure food consumption at different O_2 conditions and found that N2 animals increased food intake. Finally, we showed that a loss-of-function mutation in the pheromone metabolism enzyme *daf-22* completely suppressed the increased feeding, faster development, and fitness advantage of N2 animals. We suggest that changes to social/solitary behavior in N2 were a pleiotropic consequence of *npr-1* and *glb-5's* ability to strengthen pheromone-induced changes to feeding rate. Together, our results demonstrate how *C. elegans* can serve as a powerful model for understanding the genetic and phenotypic basis of metazoan evolution.

144M Determining genomic signatures of sexual selection utilizing sex hormone response elements. *A. Anderson*[1], A. Jones[2] 1) Dept of Biology, Texas A&M University, College Station, TX; 2) Department of Biological Sciences, University of Idaho, Moscow, ID.

Sexual selection can drive sex-biased gene expression patterns based on sex-specific hormone signaling. Hormone response elements are *cis* regulatory elements and have motifs that can be detected in the genome. We suspect that greater sexual selection might involve more genes or require greater sex-specific controls which could lead to increase in the number of HREs found in the genome. We chose to investigate primates due to their well-documented sex differences and multiple complete genomes. Using 17 primate genomes, we searched for the number of male biased (androgen) response elements and found correlations between the number of AREs and degree of differences in secondary sex traits. We then matched the location of those AREs to genes that have shown an evolutionary change across the phylogeny to determine if a change in the presence/absence of AREs had an effect on the evolution of those genes.

145M Simulating the role of L1 retrotransposons in the evolving cancer genome. *J. Atallah*, J. LeBien, G. McCollam, M. Panta Biological Sciences, University of New Orleans, New Orleans, LA.

While there is increasing evidence that transposable elements play important roles in both germline and somatic evolution, it is unclear how transposition interacts with other forces, such as natural selection, to shape the evolving genome. Although the activity of specific transposable elements has been analyzed extensively, leading to complex models of target site preferences, many previous simulations of transposable element evolution have been implemented in an artificial setting that does not consider actual genomic sequence. In an attempt to develop a more realistic simulation framework, we considered the specific case of L1 retrotransposition in the cancer genome. Our program simulates L1 insertions in the genomic sequences of an evolving cell population, using probabilistic models derived from empirical data. The selective effects of mutations on different categories of genes are taken into account. During program development, we found that L1 target sites are depleted in exonic regions of the human genome, showing that previous findings of depleted L1 insertions in exons may be either a direct outcome of the mechanism of transposition, somatic selection against exonic insertions, or both. Our simulation shows the levels of selection that are consistent with the empirical data and provides new insight into the role of passenger mutations in cancer genome evolution.

146M Relaxed purifying selection and family-specific transposition bursts drive transposable element dynamics following auto-polyploidization in *Arabidopsis arenosa*. *Pierre Baduel*[1], Leandro Quadrana[1], Kirsten Bomblies[2], Vincent Colot[1] 1) IBENS, École Normale Supérieure, Paris, FR; 2) John Innes Centre, Norwich, UK.

Polyploidization, a recurrent event throughout plant evolution, has been frequently associated with an increase in transposable element (TE) content. Polyploids are over-represented among crops, invasive taxa and colonists of challenging or variable environments. However, whether TE dynamics are altered in polyploids, or causally related to their adaptive potential, remains unknown. Two non-mutually exclusive processes could explain the increased TE content seen in polyploids: a transposition burst induced by genome duplication ("genome-shock"), and a weaker purifying selection against new insertions due to the multiplication of genome copies ("masking").

Arabidopsis arenosa is a well characterized and geographically restricted natural diploid-autotetraploid species that provides an ideal system to assess the contribution of the evolutionary forces underlying TE dynamics in polyploids. Using whole-genome sequencing data for 81 diploid and 82 tetraploid *A. arenosa* individuals from 21 populations distributed across Europe, we identified 25,235 non-reference insertions when compared to the reference genome of the closely related *A. lyrata* species. Most insertions are private or at low frequency, suggesting recent TE activity, and involve only 374 of the 2453 annotated TE families.

Detailed assessment of TE activity both before and after the single polyploidization event revealed tetraploids carried hallmarks of past transposition bursts. However, this remobilization was restricted to a handful of specific TE families, including both DNA- and retro-transposons, consistent with the genome-shock hypothesis. In addition, we detected in tetraploids a significant relaxation of purifying selection against TE insertions within coding sequences. This masking predominantly resulted in an increased load of recent genic insertions of CACTA DNA transposons and Gypsy and Copia LTR retro-elements. This first genome-wide assessment of TE dynamics in a natural diploid-autotetraploid system thus indicates that both family-specific transposition bursts and polysomic masking drive TE accumulation following polyploidization. In what measure these two TE dynamics contribute to the adaptive potential of polyploids or underlie their long-term evolutionary demises remains to be tested.

147M The effects of pathogen emergence on the genome of *Mycobacterium abscessus*. *L. Bohr*, T. Mortimer, C. Pepperell UW Madison, Madison, WI.

Purpose. *Mycobacterium abscessus* is a rapid growing, environmental mycobacteria that can cause serious lung infections. Recently, there have been multiple emergences of dominating circulating clones (DCCs) in both subsp. *massiliense* (MAM) and subsp. *abscessus* (MAA). These clones are globally distributed and can provide insight into the emergence of environmental mycobacteria as human pathogens. **Methods.** We have *de novo* assembled and annotated the pangenomes of 29 MAA and 35 MAM clinical isolates from 3 geographic regions. We characterized patterns of lateral gene transfer (LGT) in the core genome using Gubbins and compared recombinant tracts between DCCs and environmentally acquired isolates (EAIs). Additionally, we compared gene frequency and diversity between MAM and MAA, as well as DCCs and EAIs within each subspecies. To infer LGT mechanism, we identified phage and other mobile genetic elements present in these pangenomes. We calculated the population statistic π to compare nucleotide diversity between DCCs and EAIs. **Results.** Both subspecies contain an extensive pangenome with a significant number of genes found at rare frequencies. Recombination has affected a larger proportion of the core genome in MAM than MAA. Mobile genetic elements are likely the predominant LGT mechanism in *M. abscessus*. Preliminary results show a reduction in nucleotide diversity and increased pseudogenization in the DCCs compared to the EAIs in both subspecies. **Conclusion.** Our preliminary results suggest environmental mycobacterial pathogen emergence is associated with certain population genomic effects, including a reduction in the proportion of recombination affecting the core genome, gene pseudogenization, and a reduction in nucleotide diversity. These results

further our understanding of the bacterial evolutionary pathway from the environmental reservoir to the human pathogenic niche.

148M A "natural" mutation accumulation experiment in laboratory mouse lines provides insight into mutational processes of repetitive sequences. *Emily J. Brown*[1], Yu-Yu Ren[2], Abraham Palmer[2], Daniel A. Barbash[1], Andrew G. Clark[1] 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Department of Psychology, University of California San Diego, San Diego, CA.

C57/BL6 is the most commonly used inbred mouse line in research laboratories worldwide, and its origins trace back to the early 1920's when it was established by C.C. Little. A key feature of mouse line nomenclature is the Laboratory Registration Code, which is added to the name of each mouse line that has been independently maintained by a facility for more than 20 generations and has therefore been genetically isolated from all other lines. In this study, we take advantage of the genealogy of fourteen independent sublines of the BL6 laboratory mouse to identify mutations in repetitive sequences. We used methods that we and others had previously developed to estimate simple satellite sequence content in the mouse genome, which is mostly absent in the assembled genome. We used the multiple divergence times of the mouse lines, ranging from approximately 150 generations to fewer than 50, to estimate mutation rates in these difficult, unassembled regions of the genome. We confirm that mutation rates in repetitive sequences are several orders of magnitude higher than that of point mutations, but there is wide variation in mutation rates of different simple satellite sequences. The multiple divergence times present in our dataset allowed us to capture more dynamic oscillations in abundance than is observable when comparing the end points of a classic mutation accumulation experiment. We find that repeat arrays composed of different kmers show a strong correlation in abundance suggesting coordinated gains and losses among kmers. Additionally, we investigated the relationship between simple satellite content and transposable element (TE) content, annotating TE insertion events that have occurred since the divergence of these mouse sublines fewer than 100 years ago. This work represents a comprehensive quantification of the repetitive content of a mammalian genome, and provides a finer dissection of mutational processes in repetitive regions of the genome as compared to previous studies.

149M Abundant genome structure variation shapes heritable phenotypic variation in *Drosophila*. *M. Chakraborty*[1], J.J. Emerson[1], Stuart MacDonald[2], Anthony Long[1] 1) Department of Ecology and Evolutionary Biology, University of California, Irvine, CA; 2) Department of Molecular Biosciences,University of Kansas, Lawrence,KS.

Deciphering the genetic variation underlying phenotypic evolution is a fundamental question in evolutionary biology. Large scale structural mutations (e.g. duplications, deletions, insertions, etc.) play pivotal roles in genome evolution and the genetic basis of diseases. Generally, high throughput short reads are aligned to a reference genome to find structural variants (SV). Recent results from human and *D. melanogaster* show that such approaches miss 40-80% of the SVs. To determine the functional and evolutionary significance of SVs, we resequenced the founder strains of the Drosophila Synthetic Population Resources (www.flyrils.org) using long read sequencing technology and constructed *de novo* assemblies for each strain. The completeness and contiguity of the assemblies are comparable to or better than the current release of the reference strain, with most the genome represented by contiguous sequences (contigs) measuring 20Mb or longer. Collectively, we discovered thousands of structural variants, including duplicates, transposon insertions, and inversions, many of which are evolving under natural selection. Additionally, comparison of our comprehensive SV map with candidate genes obtained from published QTL mapping studies employing the DSPR unveil segregating SVs at majority of the candidate genes. One such candidate gene for nicotine resistance consists of five SV alleles at a Cytochrome P450 gene (*Cyp28d1*), comprising tandem gene duplications and TE insertions. These results suggest that sizable proportion of the phenotypic variation of complex traits in Drosophila may be due to complex genome structural changes which is shaped by natural selection.

150M Genomic survey of sex determination systems in chichlids reveals an evoluationary hot-spot. *A. Elias*[1], B. Dumont[1,2], K. Coyle[1], J. White[1], N. Roberts[1], R. Roberts[1] 1) Department of Biological Sciences, North Carolina State University, Raleigh, NC; 2) The Jackson Laboratory, Bar Harbor, ME.

Despite sex determination being a fundamental process in sexually reproducing species, there is striking diversity, particularly among fish, in the mechanism. Studying sex chromosomes is essential to understand gene, chromosome, and karyotype evolution, but they present unique challenges for sequencing and assembly. Here we present two comparative genomic strategies to characterize novel sex determination loci and gain insight into how novel sex determination alleles evolve, and how they alter the evolution of the chromosomes they reside upon. In many species where a dominant chromosome determines, sex reversal permits production of individuals homozygous for the sex chromosome, easing analysis. We compared WW and ZZ individuals of the cichlid fish *Metriaclima tarakiki*, characterizing the sex determination system with relatively low coverage sequencing. We readily identified a diverged region of 7.7 Mb corresponding to the ZW locus and cataloged variants unique to the W allele. Our analyses suggest candidate polymorphisms for sex determination. We anticipate that our strategy will be broadly applicable for rapid characterization of nascent sex chromosomes in species where sex reversal is possible.

Polygenic sex determination (PSD) is also common in cichlids, where multiple genetic factors segregate and interact to direct sexual development. While individuals of a species with PSD develop phenotypically as either male or female, multiple genetic types of males and females may exist. Using a PoolSeq strategy, we compared the genomes of four sex genotypes that produce two phenotypic sexes within a stable PSD system, *Astatotilapia burtoni*. We identified two sex determination loci at different locations in the genome, cataloged sequence variation between sex alleles at each loci, and identified possible sex determiners in those refined mapping intervals, which notably lack sex determination genes previously found in other vertebrates. The development of a model PSD system will provide the unique ability to address broad questions about evolutionary transitions and epistatic interactions in gene networks underlying sexual development.

Interestingly, the sex determination region identified in *M. tarakiki* overlaps with one of the sex determination regions found in *A. burtoni*. Thus the same locus has been recruited as a sex chromosome in cichlids repeatedly and independently, suggesting that some chromosomes are more likely to be used for sex determination.

151M Nanopore sequencing to characterize natural variation in *Drosophila melanogaster* repeat content. *C. Ellison*, K. Shah, W. Cao Genetics, Rutgers University, New Brunswick, NJ.

Short-read sequencing technology has made it possible to identify sequence variants across the entire genome for large populations of individuals. However, in most eukaryotes, a significant portion of the genome is composed of repetitive sequences, which are difficult or impossible to accurately characterize using short sequencing reads. Here, we describe a pilot project we have conducted to assess the utility of long-read sequencing for characterizing variation in repeat content among strains of *Drosophila melanogaster*. We conducted Oxford Nanopore sequencing along with Hi-C chromosome conformation capture to generate *de novo* genome assemblies for two strains of *Drosophila melanogaster* from the Drosophila Genetic Resource Panel (DGRP). From these assemblies, we have identified hundreds of novel transposable element insertions, large expansions of satellite arrays, and dozens of structural variants, the majority of which were not identified by short-read approaches. Our results suggest that long-read sequencing combined with Hi-C can produce genome assemblies with high contiguity and completeness, paving the way towards future efforts to characterize the population dynamics of repetitive elements.

152M Evolution of expression level and gene dosage following Whole-Genome Duplications in *Paramecium*. *J.-F. Gout*[1,3], O. Arnaiz[2], P. Johri[1,3], T. Doak[3], M. Lynch[1,3] 1) Arizona State University, Tempe, AZ; 2) Centre de Génétique Moléculaire, Université Paris-Sud, Orsay, France; 3) Indiana University, Bloomington, IN.

Whole-Genome Duplications (WGDs) have been rampant in the history of eukaryotes, deeply affecting the gene repertoires of many lineages. WGDs are typically followed by gradual gene losses, with only a fraction of the duplicated genes maintaining both copies over long evolutionary times. While several models have been proposed to explain the retention of duplicated genes, their respective contributions to post-WGD genome evolution remain unclear. Additionally, most of these models cannot account for the fact that selection initially prevents duplicated gene loss, only to eventually allow for one of the two copies to be eliminated, millions of years after the WGD.
Here, we focus on *Paramecium*, an exceptional unicellular eukaryote with three successive WGDs in its lineage. We obtained genome-wide measurements of expression level in 12 *Paramecium* species that share a common WGD (the *P. aurelia* species complex) and modeled the evolution of gene expression level since this common ancestral WGD. We found that gradual divergence in expression level between duplicated genes promotes the eventual loss of the copy with the lowest expression level. We propose a model where post-WGD duplicates conserve their ancestral function and are retained for dosage constraints. The gradual divergence of expression level creates an increasing imbalance in the relative contributions of the two copies, eventually freeing the lowly expressed copy from selective constraints. Because the step of expression divergence is slow, our model can explain why duplicated genes can be retained for millions of years following a WGD, only to be eventually eliminated from the genome.

153M Sex Chromosome Dosage Compensation in the Monarch Butterfly. *L. Gu*, J Walters Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS.

Evolution of heteromorphic sex chromosomes creates a dosage problem for sex-linked expression that different animals cope differently. The Lepidoptera (moths and butterflies) has the female heterogametic sex determination system (WZ~ZZ), where W chromosome is highly heterochromatinized, rendering the Z-linked loci mono-allelic in females. While all other WZ~ZZ taxa (such as birds and snakes) leaves their expression of Z-linked loci largely unbalanced between sexes, the lepidopteran insects are the only WZ~ZZ species known thus far to equalize sex-linkage expression between sexes, a pattern previously reported exclusively in XX~XY taxa such as mammals, flies and worms. However, in spite of the intriguing patterns of dosage compensation among lepidopterans, the molecular mechanism is barely understood. In order to gain further insights, we used the monarch butterfly (Danaus plexippus) as a model and examined sex-specific profiles of both gene expression using RNAseq and the landscapes of four histone modifications obtained by ChIPseq (chromatin immunoprecipitation followed by sequencing). We showed that while the Z-linked expression levels are similar between

males (ZZ) and females (ZW), H3K27me3, which is a hallmark of gene expression repression, was substantially enriched on the Z chromosome in the male. This indicates that monarch butterfly achieves dosage balance in Z expression by specifically down-regulating bi-allelic Z expression in males. H3K27me3 also involves in mammalian dosage compensation, suggesting an aspect of conserved dosage compensation mechanisms across vastly divergent taxa with opposing sex determination systems.

154M The Genome of the mushroom-feeding fly, *Drosophila innubila*, adapting to parasites and toxin exposure. *T. Hill*, B. Koseva, R Unckless Molecular Biosciences, University of Kansas, Lawrence, KS.

Mycophagous flies in the *Drosophila* subgenus (a group of ~1400 species) have been extensively studied ecologically for ~40 years, and are useful models for the evolution of toxin-tolerance, immunity, host-parasite evolution and behavior. While *Drosophila melanogaster* is a staple system for studying development, evolution, molecular genetics and behaviour, little is known about its ecology in comparison to mycophagous *Drosophila*. Here we sequence and assemble the genome of *D. innubila*, a mycophagous *Drosophila,* interesting due to their frequent infection with a DNA virus, association with a male-killing *Wolbachia*, restricted habitat and potentially toxic food substrate, in attempt to better develop this species as a model for host-pathogen evolution. We also sequenced and assembled the genome of its sister species, *D. falleni*, a generalist with wider host range and lower virus infection frequency. Our genome, comparable in quality to release 6 of *D. melanogaster*, is the first whole genome in this highly diverged *quinaria* group of *Drosophila*. We find several signatures of adaptation supporting the rapid evolution of the immune system, toxin tolerance and recombination machinery, compared to *D. falleni*. Consistent with this we also find a significant upregulation of immunity and detoxification genes, compared to *D. melanogaster* and *virilis*. Finally, we also find extensive gene duplication compared to *D. virilis*, with an expansion of gene families associated behavior and neuron development. We expect this genome will help further develop *D. innubila* and *D. falleni* as models for understanding the association between ecology and genotype.

155M Chromosomal rearrangements during turtle evolution altered the synteny of genes involved in vertebrate sex determination. *L.S. Lee*, E. Montiel, B. Navarro, N. Valenzuela Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA.

Sex determination is the biological process that directs an individual's sexual fate into male or female phenotype. Mechanisms of sex determination (SDMs) range from those directed by genetic constitution (genotypic sex determination, GSD) to those directed by environmental factors such as temperature (temperature-dependent sex determination, TSD). Turtles exhibit both TSD and GSD and possess diverse chromosome numbers, ranging from 26 to 68 chromosomes. Evolutionary transitions in SDMs are associated with changes in chromosome numbers in turtles, potentially because fusion-fission events that alter chromosome number also change the relative position and consequently, the regulation of genes governing sexual development. Here, we examined the chromosomal location of ten gene regulators of gonadal development (*Dax1*, *Dmrt1*, *Fgf9*, *Sf1*, *Sox9*, *Wt1, Fhl2*, *Foxl2*, *Gata4*, and *Rspo1*) via fluorescent *in situ* hybridization (FISH) onto metaphase chromosomes from six TSD and six GSD turtles for the first time. Our results uncovered (a) intra-chromosomal rearrangements within a single chromosome in several taxa, (b) chromosomal fusion or fission event across species that explain the co-localize of two genes in a single chromosome in some taxa and in two chromosomes in others, and (c) a translocation event of two genes from a macro-autosome and a micro-autosome to the sex chromosomes of a GSD turtle, rendering them candidate sex-determining genes that warrant further research. Our findings represent a significant advance in our knowledge of chromosomal rearrangements involving regulators of sexual development in turtles and their contribution to the evolution of sex determination in vertebrates.

156M Identification and visualization of lineage-specific genomic rearrangements with R. *D. Lindtke*, S. Yeaman Biological Sciences, University of Calgary, Calgary, T2N 1N4, CA.

Chromosomal rearrangements are known to impact phenotypic diversity and contribute to adaptation and speciation. Although several methods are available that compare pairs of extant genomes, understanding the contribution of rearrangements to the evolution of a particular lineage requires the reconstruction of its ancestral genome. However, the results of current ancestral genome reconstruction methods are not easily accessible for downstream analyses. We developed an algorithm in R to identify and classify lineage-specific rearrangements given an ancestral genome reconstruction from publicly available software. We then quantify and visualize rearrangements such as inversions, transpositions, and translocations along the genome. Our method provides a valuable tool to determine the genome-wide distribution of rearrangements, and might be used to investigate how structural genomic changes contribute to phenotypic evolution. We apply our approach to *Drosophila* species and show that lineage-specific rearrangements are heterogeneously distributed across the genome and differ in magnitude among species.

157M Characterization of a Genus-Specific Unidentified Open Reading Frame Found within the Mitochondrial Genome of *Fusarium*. *M.C. MacKillop*, H. Hamzah, M. Diazgranados, J.C. Kennell Biology, Saint Louis University, Saint Louis,

MO.

WGS Sequencing of the mitochondrial DNA (mtDNA) in the filamentous fungal genus *Fusarium* revealed a highly variable region (HVR) in the genome located between the *rnl* and *nad2* genes. Prior characterization of this region in species of *Fusarium* has revealed a large unidentified open reading frame (LV-uORF), within all species characterized to date. The LV-uORFs are variable in amino acid size and content between species, but highly conserved within certain species complexes, with the exception of the *Fusarium oxysporum* Species Complex. The LV-uORF ORF1931, encoding a putative polypeptide of 1931 amino acids, was found in the HVR of 32 isolates from four species within the *Fusarium graminearum* Species Complex (FGSC). This LV-uORF is actively transcribed, but the putative polypeptide (1931p) has yet to be detected. Current research aims to identify and localize 1931p within *F. graminearum* PH-1 mitochondria via cellular fractionation and immunoblotting. Unlike the high conservation of the LV-uORFs within the FGSC, isolates from the FOSC exhibit greater variability, ranging from 2200 to 2500 amino acids. The FOSC is an ideal lineage for both bioinformatics analysis and functional assays. Here we examine phylogenetic evidence of the allelic diversity within the LV-uORF of 47 strains from *F. oxysporum f.sp.* cubense.

158M Effecs of reproductive mode and ploidy level on repetitive element evolution. *K. McElroy*[1], J. Boore[2], J. Logsdon Jr.[1], M. Neiman[1] 1) Dept of Biology, University of Iowa, Iowa City, IA; 2) Providence St. Joseph Health and the Institute for Systems Biology, Seattle, WA.

What accounts for the immense variation in genome architecture is a fundamental biological question. Repetitive DNA, including transposable elements (TEs), mobile DNA sequences that replicate and insert themselves throughout a host genome, are important contributors to genome size variation. Understanding the conditions that influence the accumulation of repetitive DNA is thus of central importance to understanding genome evolution. Diploidy and sexual reproduction are hallmark traits of eukaryotes and transitions to polyploidy and asexuality are expected to have significant consequences for genome evolution. Reproductive mode is expected be an important determinant of TE evolution because (1) the reduced efficacy of selection should result in the unchecked proliferation of TEs in asexual lineages, but (2), TEs should be lost to genetic drift in asexual populations because TEs cannot spread to new lineages in the absence of sex. Polyploidy leads to genomic reorganization but how TEs influence/are influenced by this process remains unclear despite expectations of increased activity. Accordingly, otherwise similar lineages that vary in reproductive mode and ploidy level represent a particularly interesting setting to evaluate TE dynamics and genome evolution. Here, we use *Potamopyrgus antipodarum*, a New Zealand freshwater snail characterized by frequent coexistence of closely related coexisting diploid sexual and polyploid asexual lineages, to evaluate how changes in reproductive mode and ploidy level affect the accumulation of TEs. We used whole-genome sequence data from 26 sexual and asexual *P. antipodarum* lineages collected from natural populations to evaluate how reproductive mode and ploidy level influence abundance and genomic distribution of TEs and other repetitive elements. While we see no consistent pattern for overall TE load variation associated with reproductive mode or ploidy level we found that ribosomal DNA and histone sequences have repeatedly co-accumulated in asexual lineages, indicating important changes in genomic architecture associated with reproductive mode and ploidy.

159M Double insertion of transposable elements provides a substrate for the evolution of satellite DNA. *M.P. McGurk*, D.A. Barbash Molecular Biology and Genetics, Cornell University, Ithaca, NY.

Repetitive DNA comprises major portions of eukaryotic genomes, either in the form of transposable elements (TEs) dispersed across chromosomes or as often megabase-sized stretches of tandemly repeated sequence termed satellite arrays. Satellite variation is associated with human disease, reproductive barriers, and genomic conflicts, yet the dynamics of their evolution—particularly how new satellites arise—are poorly understood. Intriguingly, some satellites clearly originated as TEs, though it is unclear how mobile genetic parasites can transition to megabase-sized tandem arrays. The presence of two TE-derived satellites in *Drosophila melanogaster*, however, suggests that a survey of population variation could provide insight into both the mechanism and the time-scales of this process. But, determining how repetitive genomic regions vary across populations is generally considered intractable due to the challenges of interpreting repeat-derived sequencing reads. To circumvent these challenges, we developed a method called ConTExt, which models the alignment patterns of repeat-derived paired-end reads to infer structural variation and identifies a wide range of structures within repetitive sequence, including tandem elements.

Applying our method to a panel of 85 *D. melanogaster* strains from 5 populations reveals the emergence of tandem arrays from TEs in various stages of the process. First, we find that most TE families form tandem dimers and that these are a natural consequence of the insertion site preferences common to many TEs, predisposing TEs to insert multiple times at the same site. Consequently, dimers form rapidly during periods of active transposition, most strikingly evidenced by the abundance of *P-element* dimers we observe despite that *P-element* only invaded *D. melanogaster* in the last century. Demonstrating that these dimers can serve as substrates from which new satellites emerge, we discovered a copy number expansion of the DNA transposon *hobo* to ~16 tandem copies in a single line. We further characterized the two established *R1* and *Bari* tandem arrays, finding considerably less structural variation in the *Bari* array, reflecting either a younger age or selection maintaining its sequence. Overall, our results demonstrate the richness of information available in short-read

sequencing data about these understudied genomic regions and sheds light on an underappreciated consequence of TEs: the very feature that defines them—transposition—generates material from which new satellite arrays can arise.

160M Effects of sex-linked variation and paternal age on recombination rate in house mice. *A. Morgan*, J. Crowley, F. Pardo-Manuel de Villena Department of Genetics, University of North Carolina, Chapel Hill, NC.

Recombination is the defining feature of sexual reproduction. In mammalian meiosis, the physical manifestation of recombination – the formation of crossovers between homologous chromosomes – is critical for the faithful segregation of chromosomes to gametes, and thus for fertility. The rate and distribution of recombination across chromosomes modulate the impact of both genetic drift and natural selection on patterns of genetic diversity in populations. Characterizing the genetic and life-history factors that control the rate of recombination is therefore fundamentally important for understanding the evolution of mammalian genomes.

To this end we analyzed genotypes of 521 progeny from reciprocal F1 males between inbred strains representing the three major subspecies of house mice (*Mus musculus*), a well-established model system for the study of recombination and speciation. Recasting the experiment as a classical diallel design, we apply a hierarchical Bayesian model to estimate the contribution of additive, dominance and parent-of-origin effects to variation in male recombination rate. We show that a paternally-inherited factor from the *M. m. musculus* subspecies – likely the Y chromosome – increases the global recombination rate by 0.5 crossovers per meiosis (4.3%). Furthermore, aged males transmit an average of 0.6 more crossovers per meiosis (5.1%) than young males independent of the genetic background. Our design also permits direct estimation of the strength of crossover interference. We implement a well-known model of interference in a hierarchical Bayesian framework and show that crossover interference in the male germline weakens with increasing age at paternity; the effect is variable across backgrounds. Our results highlight a previously unappreciated role of the Y chromosome in control of recombination in hybrid males and hint at further links between recombination, speciation and the sex chromosomes in mice.

161M Properties of sex-biased gene expression in the absence of sex chromosomes. *J. Pascar*[1], E. Watson[1,2], S. Edmands[1] 1) Department of Biological Sciences, University of Southern California, Los Angeles, CA; 2) Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, CA.

Theory predicts that the accumulation of sexually antagonistic loci and decreased recombination between proto-sex chromosomes results in sex chromosome differentiation. Therefore, it has been suggested that polygenic sex determination (PSD) may be an unstable intermediate in the evolutionary trajectory towards sex chromosome development. Presently, it is unclear how sexually antagonistic genes are distributed throughout the genome and how the nature of sex-biased expression in the absence of sex chromosomes differs in comparison to fixed sex chromosomal systems. Depite claims that PSD may be an intermediate stage, in the intertidal copepod *Tigriopus californicus* PSD is seemingly stable, thus providing a model for investigating sex-biased gene expression in the long-term absence of sex chromosomes. *T. californicus* sex determination is known to involve at least six loci on five different chromosomes in addition to a small environmental influence. Here, we use a microdissection approach to isolate male or female gonadal tissue from single *T. californicus* individuals. Total RNA was extracted from the gonads and carcasses of age-matched individuals and sequenced on two lanes of Illumina HiSeqX. We then compare the absolute expression between sexes and tissue to identify sex-biased transcripts and explore their chromosomal distribution. Further, genes with sex-biased expression were also investigated for overlap with candidate sex-determining QTL. Using sex-specific expression a higher resolution of candidate genes responsible for sex-determination is possible. This study is one of the first to highlight the nature of sex-biased gene expression in a species with PSD. A better understanding of sex-biased loci in the emerging *T. calfornicus* model will help illustrate the evolutionary implication of sex differences in systems lacking sex chromosomes.

162M Identifying sequence heteroplasmy across entire organellar genomes of *Daucus carota* using whole genome sequence data. *Adam Ramsey*[1], Vinthuy Phan[2], Diem-Trang Pham[2], Caroline Melton[1], Bernie Daigle, Jr.[1,2], Jennifer Mandel[1,3] 1) Department of Biological Sciences, University of Memphis, Memphis, TN; 2) Department of Computer Science, University of Memphis, Memphis, TN; 3) W. Harry Feinstone Center for Genomic Research, The University of Memphis, Memphis, TN.

Heteroplasmy, a state in which cells or individuals contain multiple but distinguishable mitochondrial or plastid genomes, is increasingly recognized as a common state of organellar genomes. Heteroplasmy has been found in numerous and diverse taxa including bed bugs, *Drosophila*, humans, *Saccharomyces*, *Arabidopsis*, maize, and carrot. In plants, the genomes of mitochondria and plastids may be found in the heteroplasmic state, although the structural complexity found within mitochondrial genomes tends to make it more prone to heteroplasmy. As the development of molecular techniques have improved over the last several decades, the detection of structural and sequence heteroplasmy has increased. Sanger sequencing, quantitative-PCR, and fragment length analysis, among others, allow for the detection of heteroplasmy as single nucleotide variants, indels, and structural rearrangements. Yet with the explosion of whole genome sequencing, interest has

turned towards discovering sequence heteroplasmy across entire genomes. Bioinformatic pipelines have been developed to do so (e.g., MitoBamAnnotator, MitoChip, MitoRS, and MToolBox), but such programs have limitations (e.g., species-specific and sequence data format requirements), and they do not allow for detection of plastid heteroplasmy. Our interest in genome-wide heteroplasmy extends to the maintenance of heteroplasmy across generations and the evolutionary outcomes of heteroplasmy in the context of cyto-nuclear interactions. Here, we report the results of a new bioinformatic pipeline with the ability to 1) detect heteroplasmy in mitochondrial and plastid genomes of any species and 2) visualize heteroplasmic sites from multiple individuals aligned to a reference genome. Such a pipeline enables us to discover regions of organellar genomes commonly found in the heteroplasmic state and test hypotheses relating to those regions. Further, once identified, heteroplasmic regions can be analyzed for signatures of selection and analyzed for geographic structuring of heteroplasmy. In this study, we carried out whole genome sequencing on 48 carrot individuals from North America, Europe, North Africa, and Southwest Asia. After being processed through our pipeline, we find heteroplasmy extremely common in both organellar genomes, and it is present in coding and non-coding regions. Moreover, coding regions contain heteroplasmic variants which alter the amino acid sequences of proteins.

163M NOT REGISTEREDOf evolutionary tuning knobs: Microsatellites as engines of adaptive evolution in common sunflower. *C. Ranathunge*[1], G. L Wheeler[2], A. D Perkins[3], M. E Welch[1] 1) Department of Biological Sciences, Mississippi State University, Starkville, MS; 2) Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH; 3) Department of Computer Science and Engineering, Mississippi State University, Starkville, MS.

The mechanisms by which natural populations rapidly adapt to their local environments are not completely understood. One such proposed mechanism, the "tuning knob" model, predicts that stepwise changes in microsatellite allele length can lead to stepwise effects on phenotypes. To test the predictions of the "tuning knob" model, we estimated the effect of microsatellite allele length on heritable phenotypic variation at the level of gene expression with natural populations of the common sunflower (*Helianthus annuus* L.). Seeds collected from six populations at two distinct latitudes in Kansas and Oklahoma were planted and grown in a common garden. An RNA-Seq experiment was conducted with 95 of these individuals. Of the 3325 microsatellites genotyped using the RNA-Seq data, 479 showed significant correlation between allele length and gene expression (hereafter termed eSTRs). A second common garden experiment with populations further north and south of the focal populations demonstrated consistent patterns of correlation between allele length and gene expression at some eSTRs. These results are consistent with the hypothesis that a substantial number of transcribed microsatellites can function as "tuning knobs." Further, an extensive population genetic study conducted across the latitudinal range from Saskatchewan to Oklahoma with 672 individuals from 17 populations revealed signatures of strong directional selection on eSTRs compared to anonymous microsatellites which suggests that shorter or longer allele lengths are favored in even more extreme environments. This study provides compelling evidence that a substantial number of transcribed microsatellites can rapidly generate heritable and potentially adaptive genetic variation.

164M Deep taxon sampling reveals the evolutionary dynamics of novel gene families in *Pristionchus* nematodes. Neel Prabh, Waltraud Roeseler, Hanh Witte, Gabi Eberhardt, Ralf Sommer, *Christian Rödelsperger* Evolutionary biology, Max Planck Institute for Developmental Biology, Tübingen, DE.

The widespread identification of genes without detectable homology in related taxa is a hallmark of genome sequencing projects in animals, together with the abundance of gene duplications. Such genes have been called novel, young, taxon-restricted, or orphans, but little is known about the mechanisms accounting for their origin, age and mode of evolution. Phylogenomic studies relying on deep and systematic taxon sampling and employing the comparative method can provide insight into the evolutionary dynamics acting on novel genes. We used a phylogenomic approach for the nematode model organism *Pristionchus pacificus* and sequenced six additional *Pristionchus* and two outgroup species. This resulted in 10 genomes with a ladder-like phylogeny, sequenced in one laboratory using the same platform and analyzed by the same bioinformatic procedures. Our analysis revealed that 68-81% of genes are assignable to orthologous gene families, the majority of which defined nine Age classes with presence/absence patterns that can be explained by single evolutionary events. Contrasting different Age classes, we find that older Age classes are concentrated at chromosome centers whereas novel gene families preferentially arise at the periphery, are lowly expressed, evolve rapidly, and have a high propensity of being lost. Over time, they increase expression and become more constrained. Thus, the unprecedented phylogenetic resolution allowed a comprehensive characterization of the evolutionary dynamics of *Pristionchus* genomes indicating that distribution of Age classes and their associated differences shape chromosomal divergence. This study establishes the *Pristionchus* system for future research on the mechanisms that drive the formation of novel genes.

165M Transposon insertional mutagenesis in *Saccharomyces uvarum* reveals trans-acting effects influencing species dependent essential genes. *M.Rose Sanchez*[1,2], C. Payen[1], F. Cheong[1], B. Hovde[1], S. Bissonnette[3], A. Arkin[5], J. Skerker[5], R. Brem[6], A. Caudy[4], M. Dunham[1] 1) Department of Genome Sciences, University of Washington, Seattle, Washington, United States of America; 2) Molecular and Cellular Biology Program, University of Washington, Seattle, Washington, United States of America; 3) Department of Biological Sciences, California State University, Turlock, California, United States of America; 4)

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada; 5) Energy Biosciences Institute, University of California Berkeley, Berkeley, California United States of America; 6) Molecular and Cellular Biology, University of California Berkeley, Berkeley, California, United States of America.

To understand how complex genetic networks perform and regulate diverse cellular processes, the function of each individual component must be defined. Comprehensive phenotypic studies of mutant alleles have been successful in model organisms in determining what processes depend on the normal function of a gene. These results are often translated to the increasing number of newly sequenced genomes by using sequence homology. However, sequence similarity does not always mean identical function or phenotype, suggesting that new methods are required to functionally annotate newly sequenced species. We have implemented comparative functional analysis by high-throughput experimental testing of gene dispensability in *Saccharomyces uvarum*, a sister species of *S. cerevisiae*. We created haploid and heterozygous diploid Tn7 insertional mutagenesis libraries in S. uvarum to identify species dependent essential genes, with the goal of detecting genes with divergent function. Comprehensive gene dispensability comparisons with *S. cerevisiae* revealed that approximately 12% of conserved orthologs are
predicted to display diverged dispensability, including 22 confirmed differentially essential genes. Surprisingly, despite their differences in essentiality, these genes are capable of cross-species complementation, demonstrating that other trans-acting factors that are background dependent contribute to differential gene essentiality. Furthermore, we identified an instance of swapped essentiality between two paralogs, *CDC25* and *SDC25* between these two species. This data set provides direct experimental evidence of gene function across species, which can inform comparative genomic analyses, improve gene annotation and be applied across a diverse set of microorganisms to further our understanding of gene function evolution.

166M Functional parallels between programmed DNA loss in sea lamprey and Polycomb-mediated silencing. *C. Saraceno*, J. Smith University of Kentucky, Lexington, KY.

The sea lamprey (*Petromyzon marinus*) is a living representative of the agnathans, a jawless lineage of basal vertebrates that diverged from the gnathostomes about 550 million years ago. Unlike most vertebrates, and indeed most organisms, sea lamprey possess two distinct genomes; the somatic genome and the germline genome. This “dual-genome” state is brought about during early embryonic development via a process termed programmed genome rearrangement (PGR) whereby an estimated 20% of the genome is physically eliminated from somatic progenitor cells, resulting in germ cells that possess the full suite of genes and all other cell types possessing a reduced fraction of the germline genome. During anaphase of early embryonic mitoses, DNA slated for elimination forms bridges of lagging chromatin between the spindle poles of the retained chromosomes and is subsequently packaged into discreet subcellular structures (micronuclei) prior to being degraded and eliminated from the cell entirely. Micronuclei are enriched for several repressive epigenetic modifications not present in the primary nuclei, suggesting that these modifications play a functional role in the later stages of PGR. In an attempt to understand the possible biological functions of PGR, we searched for the enrichment of eliminated orthologs among published ChIP experiments and found that a subset of lamprey germline-specific orthologs in mouse ESCs and primordial germ cells are targeted by members of the polycomb repressive complex, a conserved set of proteins that are required to maintain the stable repression of target genes. In order to gain insight into whether polycomb group genes play a functional role during PGR in sea lamprey, we performed CRISPR/Cas9 mediated knockouts of several genes that comprise the core machinery of the polycomb complexes. Here, we show that loss of function of the polycomb gene SUZ12 results in significantly increased frequencies of micronuclei. These results suggest that Polycomb-group proteins may contribute to PGR, either by promoting packaging and clearance of eliminated DNA or permitting the proper targeting of sequences slated for elimination.

167M The genome-wide rate and spectrum of spontaneous mutations differs between haploid and diploid yeast. *Nathaniel Sharp*, Linnea Sandell, Christopher James, Sarah Otto Department of Zoology, University of British Columbia, Vancouver, BC, Canada.

While individual mutations are chance events, the biological processes that produce or prevent mutation have the potential to vary among genetic contexts, leading to systematic biases in the numbers, locations, and types of genetic changes that occur. As a consequence, populations with alternative genome architectures may have access to different kinds of genetic variation. By altering the dynamics of DNA replication and repair, alternative ploidy states may confer distinct benefits and risks in terms of new mutations. We used a mutation accumulation approach to examine the spectrum of spontaneous mutations arising in haploid and diploid forms of the budding yeast *Saccharomyces cerevisiae* under relaxed selection. Characterizing the number, types, locations, and effects of thousands of mutations revealed multiple differences between ploidy levels. Per base pair, diploids were less prone to substitutions, such that their genome-wide mutation rate was only 1.4-times that of haploids, much less than the two-fold expectation. Intriguingly, diploids also showed a reduced mitochondrial mutation rate relative to haploids. However, aneuploidy and other structural changes were more common in diploids, and had detrimental effects on growth. Haploidy is expected to reduce the opportunity for conservative DNA repair involving homologous chromosomes, increasing the insertion-deletion rate. By incorporating a particular gene knock-out in half of our lines, we show that the availability of homologous double-strand break repair does not play a major role in

determining the insertion-deletion rates of each ploidy level. Instead, haploids but not diploids were more susceptible to substitutions in late-replicating genomic regions. This suggests that diploids repaired DNA lesions in late S-phase using a conservative template-switching approach, whereas haploids used a mutagenic translesion synthesis pathway. This replication timing effect contributed to the elevated haploid substitution rate and produced a difference in the substitution spectrum between ploidy levels: the spectrum of possible substitution types was shared between ploidy levels in early-replicating regions, but divergent in late-replicating regions. Our analysis shows that when it comes to mutation diploids are not simply doubled haploids; instead, we find that ploidy level is a source of mutation rate variation, which will affect the dynamics of genome evolution in haploid and diploid populations.

168M Evolution of the chromatin landscape across the young threepine stickleback Y chromosome. *D.E. Shaw*, A.F. Shanfelter, M.A. White Department of Genetics, University of Georgia, Athens, GA.

Young sex chromosomes (i.e. X/Y or Z/W) rapidly diverge from one another once homologous recombination is suppressed. Deleterious mutations can quickly accumulate across the sex-limited chromosome, leading to a loss of functional genes. Although much attention has been focused on the evolution of coding regions across sex chromosomes, it remains unclear whether regulatory regions lose function at similar rates as coding regions and how heterochromatin generally forms across sex chromosomes. As ancient Y chromosomes are often made up of dense heterochromatic regions with very few active genes, young sex chromosome systems are needed to understand how chromatin landscapes shift. The threespine stickleback fish (Gasterosteus aculeatus) is an ideal model system to explore the evolution of chromatin across sex chromosomes, as it has a relatively young X/Y sex chromosome pair (less than 16 million years old). Sequence divergence in coding regions between the X/Y pair indicates there are two differently aged evolutionary strata, where many genes already appear non-functional. Here, we investigate chromatin accessibility across both evolutionary strata of the threespine stickleback Y chromosome by use of Assay for Transposase-Accessible Chromatin (ATAC) sequencing. We compared chromatin accessibility in testis and liver tissues across four males from a marine population of threespine stickleback fish and identified differential chromatin accessibility between the X and Y chromosomes. We are investigating local losses in accessibility that could indicate regulatory regions have become non-functional, as well as regional losses of accessibility that could indicate broad heterochromatin formation. Our results will provide important insights into how quickly chromatin landscapes can evolve on young sex chromosomes.

169M Evolutiuon of X chromosome inactivation profiles in mammals. *A.J. Slavney*, E.J. Cosgrove, A.G. Clark Molecular Biology & Genetics, Cornell University, Ithaca, NY.

In mammals, X chromosome inactivation (XCI) facilitates dosage compensation between XX females and XY males by silencing transcription from one X homolog in female cells. Some X genes exhibit partial expression from the inactive X, and are thus said to “escape” XCI. XCI escape introduces gene expression variation between females and males, and inflates variation among females. The involvement of some human and mouse XCI escapers in X aneuploidy syndrome phenotypes suggests that, at least for some genes, XCI escape may be important for normal biological function. However, it is also possible that XCI escape is merely a consequence of inefficient XCI, with little to no functional impact.

Cross-species comparisons of XCI profiles – i.e. which genes escape XCI or are completely inactivated in a given species – can improve our understanding of the evolutionary forces driving XCI escape in general, and the extent to which XCI escape is or is not functionally important at specific loci. To this end, we used existing XCI profiles from human, mouse, and opossum, and a novel dog XCI profile, in a comparative study of XCI evolution. The dog XCI profile was generated from single-cell RNAseq (10x Genomics) of peripheral blood mononucleocytes from two female F1 crossbreed dogs, and represents the first extensive carnivore XCI profile. While the read coverage of each gene was very low in each cell, the inactivated X could be identified for most cells by aggregating data across the entire X. Genes showing biallelic expression (or monoallelic expression of the inactive allele) at the cell level in multiple cells were inferred to escape XCI, whereas those showing only monoallelic expression across cells were classified as X-inactivated. Using this method, we identified a total of 22 XCI escapers in the canine genome.

Comparison of XCI profiles across species revealed limited conservation of XCI status among one-to-one X orthologs. Despite this, we found a significant level of sharing of gene ontology annotations for XCI escapers and inactivated genes across species, including metabolic and immune system processes. When we compared signatures of selection between XCI escapers and X-inactivated genes, we observed that human XCI escapers show a greater degree of conservation than X-inactivated genes. While this pattern generally held in other species, the magnitude of the difference between the XCI categories varied. These results suggest that despite XCI profiles being largely lineage-specific, they may be shaped by similar selection pressures.

170M The Sea Lamprey (*Petromyzon marinus*) Genome Provides Deep Evolutionary Perspectives on Genome Reprogramming and Stability. *J.J. Smith*[1], N Timoshevskaya[1], V.A. Timoshevskiy[1], C.K.M Waterbury[1], C Saraceno[1], . Lamprey Genome Consortium[2] 1) Department of Biology, University of Kentucky, Lexington, KY; 2) Lamprey Genome Consortium.

The lamprey genome provides unique insights into both the deep evolutionary history of vertebrate genomes and the

maintenance of genome structure/integrity over development. The lamprey lineage diverged from all other vertebrates approximately 500 million years ago. As such, comparisons between lamprey and other vertebrates permit reconstruction of ancient duplication and rearrangement events that defined the fundamental architecture and gene content of all extant vertebrate genomes. Lamprey also undergoes programmatic changes genome structure that result in the physical elimination of ~20% of its genomic DNA (~0.5Gb from a ~2 Gb genome) from all somatic cell lineages during early embryonic development. Here, we outline recent progress in assembly and analysis of the lamprey germline genome, and progress in the development of methods for characterizing the cellular events that mediate DNA elimination. We have integrated information from several sampling approaches and sequencing technologies to achieve a highly contiguous assembly of lamprey genome (including: Illumina fragments/mate pairs, 20X coverage in Pacific Biosciences reads, dense meiotic maps, chromatin contact maps, and optical mapping data). This genome assembly has dramatically improved our ability to dissect the molecular basis and genetic outcomes of programmed genome rearrangements (PGRs), and has improved our understanding of the tempo and mode of large-scale duplications and translocations within the ancestral vertebrate lineage. Analysis of the germline genome identifies several genes that are physically eliminated from all somatic tissues. These eliminated genes correspond to several known oncogenes that are somatically silenced by Polycomb Repressive Complex during early embryogenesis, and identify several other novel oncogene candidates. Complementing this assembly, the development of approaches for *in situ* analysis of 3D preserved cells has revealed that PGR unfolds through a series of dramatic cellular events that involve the programmatic alteration of several fundamental mechanisms of mitosis and epigenetic silencing, including: alignment of chromosomes at metaphase, chromatid cohesion, separation and segregation, nuclear envelope formation, and DNA methylation.

171M To TE, or not to TE, that is the question: transposable element dynamics in hybrid and naïve genomes. *C. Smukowski*, M. Dunham Genome Sciences, University of Washington, Seattle, WA.

Transposable elements (TEs) are repetitive, mobile DNA elements that can have deleterious effects on their hosts by triggering changes in gene expression, chromosomal rearrangements, and genome size expansion. Due to these consequences, TEs and their hosts exhibit an evolutionary arms race scenario in which hosts evolve mechanisms to silence transposition and TEs mutate to maintain transposition, successfully colonizing nearly every organism across the tree of life. The yeast *Saccharomyces uvarum*, a relative of *Saccharomyces cerevisiae*, is therefore quite unusual, as it has expelled its' TEs, leaving a genome with no active TEs and only small fragments of former TEs. One of only a handful of organisms ever observed with this pattern, *S. uvarum* thus represents a unique model for understanding the genetic and environmental factors that regulate transposition and how TEs colonize naïve genomes. Utilizing laboratory experimental evolution, we have found novel TE insertions in evolved *S. cerevisiae* haploids and diploids, but no new insertions after hundreds of generations in *S. cerevisiae* x *S. uvarum* evolved hybrids, suggesting active suppression of transposition. To test this hypothesis, we have directly assayed transposition rate in these and other interspecific hybrids, in addition to *S. uvarum* with an artificially introduced TE. Finally, we propose and test a genetic mechanism that may be responsible for the observed inhibition of transposition. In summary, we shed new light on one of the most important interactions driving genome evolution across every domain of life.

172M Methods for statistical inference of adaptive mutations, genes, and biological pathways. *Lauren A Sugden*[1,2], Sohini Ramachandran[1,2] 1) Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Ecology and Evolutionary Biology, Brown University, Providence, RI.

Two major challenges exist for understanding how positive selection drives the adaptation of populations to novel environments: the first lies in identifying genomic loci with adaptive mutations, while the second lies in understanding how those genetic alterations contribute to adaptive phenotypes. Here we introduce two statistical frameworks that provide insight into adaptation at the level of variants and genes, and phenotypic pathways.

First, we describe a supervised classification framework, SWIF(r), that explicitly models joint distributions of a set of selection statistics under adaptive scenarios of interest, returning calibrated posterior probabilities for each genomic site. In simulations and in 1000 Genomes Project data, SWIF(r) outperforms other machine learning methods at identification of regions and precise variants undergoing hard sweeps. We apply SWIF(r) to genotype data from the ≠Khomani San of southern Africa, an understudied hunter-gatherer population with the largest genetic diversity of any modern population, and find an overrepresentation of SWIF(r) signals in genes associated with metabolism and obesity. This finding provides support for the thrifty gene hypothesis, suggesting that fat storage has played an important role in the evolution of the ≠Khomani San; in exome data, we find candidate adaptive mutations in the form of highly differentiated, functional variants tagged by SWIF(r), including a missense mutation in the gene coding for adiponectin, a regulator of glucose and fatty acid metabolism.

Second, we develop a hidden Markov model (HMM) designed to detect evidence for selection at the genic level. With hidden states corresponding to adaptive, linked neutral, and unlinked neutral sites, we leverage stochastic backtrace to sample thousands of probabilistically representative paths through the HMM. These sampled paths provide measures of uncertainty

surrounding the location of beneficial alleles, and also allow us to calculate gene-level adaptive probabilities by counting the number of paths that pass through the adaptive hidden state. These probabilities facilitate the identification of polygenic selection on biological pathways, furthering our understanding of the evolution of complex traits.

173M Transcriptome turnover during organ evolution. *A. Thompson*[1], M. May[1], B. Liebeskind[2], D. Begun[1], B. Moore[1], A. Kopp[1] 1) Ecology and Evolution, University of California, Davis, CA; 2) Molecular Biosciences, University of Texas, Austin, TX.

An organ's phenotype is determined by its transcriptome. As organs evolve, genes are added to and subtracted from the transcriptome through myriad mutational processes leading to transcriptome turnover. Do organ transcriptomes gain and lose genes predominantly through genomic structural changes such as gene duplication and deletion, or are regulatory changes the principle mechanism of turnover? Answering these questions requires new methods of analyzing transcriptomic data and modeling their evolution.

To investigate transcriptome turnover, an accurate way of measuring the binary expression state of genes (on or off) is needed. Next generation RNA sequencing yields high resolution estimates of relative gene expression in transcriptomes, but the qualitative activity state of genes remains obscure. Technical and biological noise, as well as insufficient sequencing depth, often causes inactive genes to appear as if they are expressed, or contrarily, causes active genes to appear silenced. To overcome this noise, we developed a Bayesian statistical framework to leverage data from replicate libraries to estimate sets of genes that are "on" and "off". Our method calculates the posterior probability that each gene is actively expressed in an organ. We demonstrate that our method provides probability estimates that are consistent with expectations based on other data types such as chromatin-state profiling data.

In addition to this new method, we have developed a model of gene family evolution called the birth-death-regulation model, which parametrizes regulatory changes and structural changes such as duplication and deletion. Together, our methods integrate genomic and transcriptomic data to investigate the mode and tempo of important evolutionary processes that add and remove genes from transcriptomes. Finally, we use this model to analyze a comparative RNA-seq dataset to study transcriptome turnover in the rapidly evolving male sex organs of Drosophila, illustrating the power of our model to provide insights into transcriptome evolution.

174M Integrative Cytogenetics of the Sea Lamprey Chromosome Elimination. *V.A. Timoshevskiy*, N.E. Timoshevskaya, J.J. Smith Department of Biology, University of Kentucky, Lexington, KY.

The sea lamprey (*Petromyzon marinus*) is one of few vertebrate species that reproducibly eliminated larger fractions of its genome during normal embryonic development. These elimination events are initiated at the 6th embryonic cleavage and result in the loss of ~20% of the lamprey's genome from essentially all somatic cell lineages (these same sequences are retained in the germline). Available evidence suggests that DNA elimination acts as a permanent silencing mechanisms preventing the somatic expression of a specific subset of "germline" genes. This germline-specific DNA is lost in the form of large fragments, including entire chromosomes. However, reconstruction of eliminated regions has proven challenging due to the complexity of the lamprey karyotype (84 small pairs of somatic chromosomes and ~100 pairs of germline chromosomes) and the exceedingly high repeat content of the genome and even higher repeat content of eliminated fragments.
We applied integrative approach aimed at further characterization of the large-scale structure of eliminated segments, including: 1) developing DNA-probes that selectively labels eliminated chromosomes by laser capture microdissection; 2) *in silico* identification of germline-enriched repeats; 3) determining the chromosomal location of specific repetitive sequences in germline metaphases, 4) verification of specificity to eliminated chromosomes by 3D DNA-DNA-hybridization on lagging anaphases in embryos. Our integrative approach allowed us identify multiple repetitive elements that are found exclusively on the eliminated (germline-specific) chromosomes and resulted in the identification of 12 chromosomes that appear to be programmatically eliminated during early embryogenesis. These chromosomes differ in size, morphology, and the localization of five germline-specific repetitive elements. The fidelity of germline-specific repetitive elements and their distinctive patterning in elimination anaphases is taken as evidence that these sequences might contribute to the specific targeting of chromosomes for elimination and in molecular interactions that mediate their decelerated poleward movement in chromosome elimination anaphases.

175M Copy number and expression variation in ampliconic genes on the Y chromosomes in humans and other great apes. *R. Vegesna*[1,2,3,6], M. Tomaszkiewicz[4], M. DeGiorgio[1,3,4], P. Medvedev[1,2,3,5], K. Makova[1,3,4] 1) Bioinformatics and Genomics Graduate Program, The Huck Institutes for the Life Sciences, Pennsylvania State University, State College, PA; 2) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 3) The Genome Sciences Institute of the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA; 4) Department of Biology, Pennsylvania State University, University Park, PA; 5) Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA; 6) Computation, Bioinformatics, and Statistics Training Program, The Huck Institutes for the Life Sciences, Pennsylvania State University, University Park, PA.

The male-specific region of the human Y chromosome harbors nine multi-copy ampliconic gene families [1]. The gene copies within each family are frequently 99.9% identical to each other, because most of them occupy opposite arms of palindromes, or massive inverted repeats, on the Y. Ampliconic genes are expressed exclusively in testis and encode proteins functioning during spermatogenesis [2]. Experimental studies demonstrated that there is variation in the number of ampliconic genes per family within and across great ape species [3,4,5]. However, how the variation in ampliconic gene copy numbers affects male fertility and expression levels in testis of humans and other great apes has remained understudied. Here we present a novel method, Ampliconic Copy Number Estimator (AmpliCoNE), that utilizes read depth information to estimate ampliconic gene copy number (CN) per family. We estimated ampliconic gene copy number and expression levels in 23 human samples from the Genotype-Tissue Expression (GTEx) project using AmpliCoNE and Kallisto, respectively. Across the ampliconic gene families, the more copious ones (*TSPY* and *RBMY*) had higher expression levels than less copious ones, but, surprisingly, no significant relationship between copy number estimates and gene expression was found within individual gene families. For the more copious gene families, differences in both gene copy number and expression levels were partially determined by Y haplogroups. In particular, the *TSPY* family in haplogroup E (African) had higher copy number estimates but lower expression levels than in haplogroup R (European). In contrast, the *RBMY* family in haplogroup R had higher copy number estimates and lower gene expression levels than in haplogroup E. We assembled the testis transcriptomes [3] of great apes (bonobo, chimpanzee, gorilla and orangutan) to obtain transcripts representing ampliconic gene families and calculated their expression values from currently available great ape samples. Currently, we are in the process of studying the evolution of ampliconic gene expression in great apes.

1. Skaletsky H., et al. Nature (2003).
2. Vallender, E., Bruce L. BioEssays (2004).
3. Tomaszkiewicz M, et al. Genome Res. (2016).
4. Oetjens, M., et al. GBE (2016).
5. Massaia, A., Yali X. Human genetics (2017).

176M Comparative genomic analysis of programmed DNA elimination in lamprey. *C.K.M. Waterbury*, J.J Smith Biology, University of Kentucky, Lexington , KY.

Nearly every multicellular organism possesses a genome that is invariant across all cell types, although individual cells only express a subset of genes that are encoded in the genome and different cell lineages may differ dramatically with respect to the genes that they express. Differentiation of cell lineages is often mediated by epigenetic silencing events that restrict the expression of genes, including DNA and histone methylation. Notably though, a few species, are known to employ an additional mode of silencing that involves the physical removal of DNA from the genomes of somatic cell lineages during development, a phenomenon generally known as programmed DNA elimination. One of the most extreme examples of DNA elimination is seen within lamprey (*Petromyzon marinus)*, which results in the physical elimination of nearly 0.5 Gb of "germline-specific DNA from all somatic tissues during early embryogenesis. We recently developed a new assembly of the lamprey germline genome, which improved our ability to dissect the molecular basis and genetic outcomes of programmed elimination, as well as our ability to reconstruct the evolutionary history of eliminated regions. Initial analyses of these germline specific regions have shown that the eliminated DNA consists of both repetitive and single copy DNA, and provides further evidence supporting the idea that germline-specific genes are likely deleterious if misexpressed in somatic cell lineages. Moreover, these analyses reveal that many germline-specific genes have undergone local duplication and are often arranged as low-copy tandem or inverted repeats. Comparison of orthologous regions between the sea lamprey and Japanese lamprey (*Lethenteron camtschaticum*) genomes, reveal the presence of ancestral syntenic blocks within eliminated regions and lend further support to the idea that eliminated regions are conserved deep within the lamprey lineage.

177M Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *D. Yin*[1], E. Schwarz[2], C. Thomas[1,3], R. Felde[1], I. Korf[4], A. Cutter[3], C. Schartner[5], E. Ralston[5], B. Meyer[5], E. Haag[1] 1) Department of Biology, University of Maryland, 4094 Campus Drive, College Park, MD; 2) Department of Molecular Biology and Genetics, Biotechnology 351, Cornell University, Ithaca NY; 3) Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, ON; 4) Department of Molecular and Cellular Biology and Genome Center, University of California, 1 Shields Avenue, Davis, CA; 5) Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California, Berkeley, CA.

Evolutionary transitions in mating systems can profoundly alter the genome and reproductive traits it encodes. In *Caenorhabditis* nematodes, species that evolved self-fertile hermaphrodites have smaller genomes and reduced mating success than male-female, obligate outcrossing species. The connections, if any, between sexual traits and genome size, and the speed with which genome size changes, remain unclear. We compared chromosome-scale genome assemblies for the two most closely related *Caenorhabditis* nematodes with alternative sexual modes, the outcrossing *C. nigoni* and the selfing *C. briggsae*. The *C. nigoni* genome is 19% larger and encodes 24% more proteins, resembling other male-female *Caenorhabditis*

and indicating that the *C. briggsae* genome shrank rapidly. The two genomes have similar percentage of repetitive and intronic DNA, while intergenic and coding sequences have become smaller in *C. briggsae. C. nigoni*-specific genes were enriched for those encoding small proteins with male-biased expression. We identified a large family of genes that encode small proteins that show strongly male-biased and germ cell-biased expression. Within this family, the *male short secreted (mss)* genes form a monophyletic clade found only in outcrossing species and lacking homologs in the selfing *C. elegans, C. tropicalis* and *C. briggsae.* In the *C. briggsae* genome, *mss* fragments and a pseudogene indicate recent loss of the *mss* sub-family. Using deglycosylation treatment and epitope tagging, we find that MSS are glycoproteins that enter the secretory pathway and are retained on the sperm surface after activation. In the outcrossing *C. remanei*, *mss*-null males had normal fertility, but their sperm failed to compete with wild-type males. Conversely, restoration of *mss* to *C. briggsae* males was sufficient to render their sperm more competitive than those of wild-type males, and made males more common in mixed-sex populations. These results directly link the reduced mating efficacy of selfing species to the loss of reproductive genes, and highlight the ongoing role that sexual mode plays in shaping the genome.

178M Female x male postcopulatory interactions and the evolution of gametic incompatibilities. *Y. Ahmed*, A. Clark Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.

The contribution of sexual selection to speciation is poorly understood, yet a pervasive pattern in evolution is the rapid divergence of sexually selected traits. Postcopulatory sexual selection in particular can drive formation of new species by establishing barriers to fertilization between diverging lineages. We use the *Drosophila virilis* group as a model to investigate the genetic basis of interspecific gametic incompatibilities and the potential role of sexual selection in this process. This group of species is characterized by strong interspecific incompatibilities at the gamete level (postmating, prezygotic) and is an ideal system for genetic studies. We use several approaches to study these incompatibilities. First, on the male side, we are using RNA-seq and proteomics of adult male reproductive organs to identify proteins that are part of seminal fluid secretions. Seminal fluid proteins (SFPs) affect a variety of post-mating responses in females, and are key determinants of paternal reproductive success. We find that SFPs are rapidly evolving between species, and that the most rapidly evolving SFPs reside within mapped genomic regions that contribute to the paternal component of gametic incompatibility between species. Second, we are studying the regulatory postmating response in females mated to con- and heterospecific males to identify genes that show abnormal postmating responses. Typically, females undergo physiological and behavioral changes after mating, which we hypothesize might be disrupted in interspecific species matings. Indeed, we find that heterospecific matings induce abnormal upregulation of several immune-related genes in the reproductive tract of females. This raises the possibility that females launch an immune response against the ejaculate of foreign males, thus contributing to incapacitation of sperm within the reproductive tract. These results show that rapid evolution of reproductive proteins likely drive incompatibilities between closely related species, and we are now exploring the underlying genetic causes of this divergence.

179M Simulations and genomic data from hybridizing haplodiploids reveal exceptionally heterogeneous genetic differentiation. *Emily E Bendall*[1], Vitor C Sousa[2], Catherine R Linnen[1] 1) Department of Biology, University of Kentucky, Lexington, KY; 2) Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências da Universidade de Lisboa.

When gene flow accompanies speciation, genetic differentiation is expected to be highly heterogeneous across the genome, with loci underlying reproductive isolation exhibiting unusually high levels of differentiation. For this reason, genome-wide scans of genetic differentiation are a potentially powerful tool for uncovering the genetic basis of adaptation and speciation. However, interpretation of genome scans is complicated by the many confounding factors that also shape the genomic landscape of differentiation, including local variation in recombination rates, background selection, and hitchhiking. If any of these factors differs consistently among different taxa, this may give rise to predictable differences in patterns of genomic divergence heterogeneity in nature. For example, because recessive mutations are exposed to selection in haploid males, pervasive linked selection may produce exceptionally heterogeneous genomic differentiation in organisms with haplodiploid sex determination. To test this hypothesis, we combined simulations and genome-wide SNP data from a pair of hybridizing haplodiploids (*Neodiprion lecontei* and *Neodiprion pinetum;* both pine sawflies in the order Hymenoptera). In simulations of divergence with gene flow that incorporated both divergent and background selection, we found that genetic differentiation was more heterogenous in haplodiploids than in diploids. Consistent with these simulation results, our empirical data revealed exceptionally heterogeneous genetic differentiation and a strikingly bimodal F_{ST} distribution in sympatric (hybridizing) populations of *N. lecontei* and *N. pinetum*. We also recovered less pronounced versions of this bimodal distribution in allopatric species/population pairs. Together, these results indicate that haplodiploidy alone can produce exceptionally heterogenous genetic differentiation and that divergence-with-gene-flow increases this heterogeneity further. More generally, our data demonstrate that different taxa can have predictably different patterns of genetic differentiation; these taxon-specific effects should be taken account when interpreting the results of genomic scans.

180M The genomics of behavioral evolution and speciation in the *Drosophila athabasca* species complex. *R.R. Bracewell*, K.M. Wong-Miller, D. Bachtrog Integrative Biology, UC Berkeley, Berkeley, CA.

Drosophila athabasca is a powerful system for exploring behavioral evolution and the genetic/genomic underpinnings since

it is comprised of three very recently formed and partially reproductively isolated semispecies. Males sing semispecies-specific courtship songs and there is behavioral isolation due to female preference. To better understand this speciation event, we took a population genetic and genome scan approach combined with QTL mapping of male song parameters to identify regions of the genome that underlie this trait. We first assembled high-quality genomes of the two most closely related semispecies using PacBio sequencing and Hi-C technologies. The resulting chromosome-level scaffolds were then used in conjunction with high coverage genomic data from 40 individuals to fully characterize structural variation thought to play a role in divergence in this system. We then conducted whole genome sequencing of an additional 300 individuals from 16 collection sites to characterize patterns of diversity and divergence across the genome. We find that genetic diversity and divergence are strongly influenced by large-scale inversions in the three semispecies. Results from our QTL mapping indicate that the X chromosome plays a large role in male courtship song and fixed inversions on the X chromosome were found to strongly limit recombination in hybrids. Genome scans for regions of divergence also indicate the X chromosome is particularly divergent and we identify a strong candidate gene that may underlie male courtship song differences between semispecies. In total, our results show how combining classic crossing experiments and QTL analyses with cutting edge genomic techniques can help us understand the process of speciation.

181M The molecular evolution of pheromone communication and reproductive isolation in orchid bees. *Philipp Brand*[1], Thomas Eltz[2], Santiago Ramirez[1] 1) Department of Evolution and Ecology, UC Davis, Davis, CA; 2) Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Bochum, Germany.

Finding a suitable mating partner is central to evolutionary success. Animal lineages across the tree of life use pheromones to locate and identify mates. Variation in pheromone communication systems has been linked to the origin and maintenance of reproductive isolation. Nonetheless, the underlying genetic and molecular mechanisms that control pheromone variation and signal detection remain poorly understood. Unlike the majority of insects, male orchid bees collect chemical substances from various environmental sources to concoct complex species-specific pheromone mixtures that are subsequently exposed to females during courtship display. As a result, orchid bees rely exclusively on olfactory encoding to accumulate and detect pheromones. Here we tested the hypothesis that the evolution of pheromone composition correlates with divergence in genes underlying the sense of smell and that these genes are involved in the origin and maintenance of reproductive isolation in orchid bees. We conducted a population-level analysis of two recently diverged parapatric sibling species including 366 bees collected throughout the entire distribution range of both species. We identified pheromone compounds that are highly species-specific, while genetic differentiation was low between both species. After correcting for population structure, we identified few highly differentiated genomic regions between species. In particular, we found pronounced differentiation in a tandem array harboring 37 odorant receptor (OR) genes. Although most of these genes exhibited a pattern of low divergence, we identified a selective sweep around OR41 in combination with significantly elevated dN/dS in one species, suggesting that the gene evolved under strong divergent selection. In addition, OR41 exhibited differential expression between sexes, a common characteristic of ORs involved in pheromone communication in other insects. Together, our results indicate that OR41 played an important role in the evolution of pheromone communication in these two recently diverged lineages. In addition, our results support the hypothesis that chemosensory receptor gene divergence may underlie the early steps of reproductive isolation in orchid bees.

182M Selection minimizes introression around incmopatabilitie and regions under strong linked selection in *Capsella*. *Y.J. Brandvain*[1], Tyler Kent[2], Stephen Wright[2], Krzsztof Stankiewicz[3] 1) Plant and Microbial Biology, University of Minnesota, St Paul, MN; 2) Department of Ecology & Evolutionary Biology. University of Toronto. Toronto, ON; 3) Mathematics and BioSciences Group. University of Vienna . Vienna, Austria.

Upon secondary contact some genomic regions rapidly introgress while others remain distinct. We wish to know what determines these outcomes. We examine the density of ancestry from Capsella rubella across 180 genomes ot its sister species, Capsella rubella. We find that a parameter that summarizes the extent of linked selection strongly predicts genome-wide heterogeneity in the extent of introgression introgression. This result suggests that, rather than selecting against a small number of incompatibilities or differential adapted alleles, selection acts to remove introgression across most functional regions of the genome. In the face of this strong, genome wide trend, we also find a dearth of introgression on a chromosome arm in which C. rubella carries a derived incompatibility.

183M High inter- and intraspecific turnover of heterochromatin-associated repeats in great apes. *M. Cechova*, R. S. Harris, M. Tomaszkiewicz, F. Chiaromonte, K. D. Makova Penn State, University Park, PA.

Highly repetitive heterochromatin serves as a structural component of centromeres and telomeres and is known to drive speciation. Indeed, small unbalanced heterochromatic segments can lead to mitotic failure and non-viable offspring in inter-specific hybrids. To bring us closer to understanding the role and turnover of heterochromatin during speciation, we conducted the first detailed genome-wide investigation of satellite repeats in six great ape species (human, chimpanzee,

bonobo, gorilla, Sumatran orangutan, and Bornean orangutan).

Repeats were identified from publicly available short-read (Illumina) sequencing data. Additionally, we developed a novel computational tool NoisyRepeatFinder that can be readily used for annotating repeat arrays in long and noisy PacBio and Oxford Nanopore reads. Indeed, using this tool, we identified >14-kb repeat arrays composed entirely of (GGAAT)n pentamer in human.

We discovered extremely rapid interspecific repeat evolution. Most repeated motifs were species-specific and less than 1% of them were shared by all six species. Overall, great ape genomes harbored only a handful of abundant repeats, some of which are phylogenetically related. These include the (GGAAT)n pentamer that represents more than half of the total repeat content in human and is present in long arrays in all species. Additionally, they include homologous 32-mers in chimpanzee, bonobo, and gorilla. Great ape individuals could be classified into species based on 14 most abundant repeated motifs, opening an intriguing possibility that even low-coverage noisy reads from metagenomic samples (such as stool samples) could possibly be used for rapid species identification.

We also found a large interindividual variability in repeat density within the same species. Repeat density between some individuals differed up to 3-fold Individuals with highest repeat density had the elevation of all abundant repeats. Thus, the repeat density of an individual is explained by a concerted elevation of many repeats, and not by an expansion of a few repeats. Indeed, we found the counts of repeated motifs within individuals to be correlated, predominantly positively.

Finally, we found many of the abundant 32-mers to be more abundant in males than females. We predicted that such repeats should be present on the Y chromosome. Fluorescent In Situ Hybridization (FISH) indicated the sub-telomeric location and confirmed the presence of the 32-mer repeat on the bonobo Y, in line with our expectations.

184M *Prdm9*-controlled meiotic chromosome interactions in hybrids between closely related mouse subspecies. *J. Forejt*[1], S. Gregorova[1], V. Gergelits[1], I. Chvatalova[2], T. Bhattacharyya[3], B. Valiskova[1,4], V. Fotopulosova[1], P. Jansa[1], D. Wiatrowska[1] 1) Institute of Molecular Genetics ASCR, Vestec, Czech Republic; 2) The National Institute of Public Health, Prague, Czech Republic; 3) The Jackson Laboratory, Bar Harbor, ME; 4) Charles University, Faculty of Science, Prague, Czech Republic.

Hybrid sterility creates a barrier to gene flow between related taxa, thus safeguarding integrity of closely related (sub)species and contributing to the process of speciation. Genetic control of hybrid sterility is often polygenic and poorly understood, with only a handful of hybrid sterility genes identified until now. The first vertebrate hybrid sterility gene PR domain containing 9, *Prdm9*, causes failure of meiotic chromosome synapsis and infertility in the male hybrids between mouse strains derived from two mouse subspecies. Within a species, *Prdm9* determines the sites of programmed DNA double-strand breaks and meiotic recombination hotspots.

To inquire into the relationship between *Prdm9*-controlled meiotic recombination, meiotic synapsis of homologous chromosomes and hybrid male sterility we used a panel of intersubspecific chromosome substitution strains to construct intersubspecific F1 hybrid males with random stretches of consubspecific homology on several autosomal pairs. We analyzed their ability to restore synaptonemal complexes and to rescue male fertility. We found that >27 Mb of randomly located consubspecific homology intervals fully restored synapsis in a given autosomal pair and that concerted renewal of meiotic synapsis in several pairs of homologous autosomes led to the restoration of fertility of intersubspecific hybrid males.

We conclude that the chromosomal mechanisms can play an important, so far mostly unexplored role in reproductive isolation between closely related (sub)species, and we speculate that impaired recombination between evolutionarily diverged homologous chromosomes could function as one of the ancient mechanisms of hybrid sterility occurring in various sexually reproducing species.

185M Fine-scale ancestry switching across the genomes of wild hybrid mice. *M. Frayer*[1], J. Hvala[1], L. Turner[2], J. Novembre[3], D. Tautz[4], B. Harr[4], B. Payseur[1] 1) Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; 2) Department of Biology and Biochemistry, Milner Centre for Evolution, University of Bath, Bath, UK; 3) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 4) Department of Evolutionary Genetics, Max-Planck Institute for Evolutionary Biology, Plön, Germany.

Hybrid genomes contain signatures of historical admixture and divergence. In hybrid zones between nascent species, differential gene flow is usually measured by comparing genotypes across individual molecular markers. Focusing instead on local ancestry inference naturally incorporates correlations among markers and enables the use of full genome sequences. We applied this approach to house mice sampled from two populations within a hybrid zone between *Mus musculus domesticus* and *M. m. musculus*. Probabilistic reconstruction of local ancestry from forty standard Illumina (unphased) genome sequences and eight Chromium linked-read (phased) genome sequences revealed fine-scale ancestry switching along chromosomes. This suggests many generations of hybridization despite known reproductive barriers between subspecies

and the restricted geographic range for hybrids. The density of ancestry junctions varied substantially across the genome and among individuals. Certain genomic regions exhibited fewer junctions than expected given local recombination rates, raising the prospect that mutations conferring reproductive isolation are located in these regions.

186M Correlation and constraint of self and interspecific incompatibility across the range of Texas *Phlox*. *R. Hopkins*, F. Roda Organismic and Evolutionary Biology, Harvard University, Boston, MA.

Many plant species have evolved genetic recognition systems between pollen and pistils that identify and reject inappropriate pollen before fertilization. Two of the most important and broadly studied systems involve self-pollen and interspecific-pollen recognition. Important outstanding questions are if and how these two recognition systems are mechanistically pleiotropic and if this pleiotropy could constrain the evolution of incompatibility. The hypothesized mechanistic pleiotropy is motivated by the observed correlation across plant species between self and interspecific pollen-pistil incompatibility. Species that have genetic self-incompatibility also tend to reject pollen of closely related species, where as self-compatible species tend to be compatible with heterospecific pollen. Here we demonstrate a within-species correlation of self and interspecific incompatibility and provide strong evidence that interspecific incompatibility imposes a geographic constraint on the evolution of self-compatibility. Our study characterizes variation in self and interspecific incompatibility in the native Texas wildflower *Phlox drummondii*. This species has heritable, quantitative variation in self-incompatibility ranging from complete self-incompatibility to complete self-compatibility. We demonstrate that this variation in self-incapability is significantly correlated with variation in incompatibility with its close congener *P. cuspidata*. Furthermore, we find that both self and heterospecific incompatibility is due to pollen recognition and rejection at the stigmatic surface, and likely involves pollen adhesion or early pollen-tube germination. Finally, we find that variation in incompatibility is geographically distributed suggesting the evolution of self-compatibility is constrained by selection favoring interspecific -incompatibility. *P. drummondii* and *P. cuspidata* co-occur and hybridize in a broad area of sympatry in eastern Texas. The resulting hybrids are largely sterile indicating selection could favor increased interspecific -incompatibility in sympatry. Consistent with this hypothesis, we find sympatric populations have significantly higher interspecific incompatibility than allopatric populations. Because of the strong correlation between incompatibilities we also find that sympatric populations have significantly higher self-incompatibility than allopatric populations.

187M Pairwise hybrid incompatibilities dominate allopatric speciation for a simple biophysical model of development. *B.S. Khatri*[1,2], R.A. Goldstein[1] 1) Division of Infection and Immunity, University College London, London, United Kingdom; 2) The Francis Crick Institute, London, United Kingdom.

How the processes of evolution give rise to non-interbreeding species is still not well understood. In an empirical search for a genetic basis, protein-DNA binding has commonly been identified as a factor in the development of reproductive isolation. Computational and theoretical models based on the biophysics of protein-DNA binding have provided a mechanistic basis of such incompatibilities between allopatrically evolving populations. However, it is an important open question whether or not such pair-wise interactions are important compared to higher-order interactions in speciation, since gene regulation by such binding events occurs embedded within larger gene regulatory networks. Theoretical arguments based purely on combinatorics suggest that higher-order incompatibilities should arise more easily. Here, we use simulations of a simple biophysical genotype phenotype map of spatial patterning in development, to show that biophysics provides a stronger constraint, leading to pair-wise incompatibilities arising more quickly and being more numerous than higher-order incompatibilities. Our results also point to a number of emergent phenomena, directly as a result of the non-trivial nature of this complex genotype-phenotype map, such as reproductive isolation arising more quickly for smaller populations, clustering of hybrids by sequence entropy constraints and sub-diffusive behaviour of hybrids in trait space at large population sizes. We suggest the balance between sequence entropy and fitness may play a universal role in the growth of incompatibilities in complex gene regulatory systems.

188M Cytoplasmic-nuclear incompatibility between wild-isolates of *Caenorhabditis nouraguensis*. *P. Lamelza*[1,2], J. Young[3], H.S. Malik[1,3], M. Ailion[1,2] 1) Molecular and Cellular Biology Program, University of Washington, Seattle, WA; 2) Department of Biochemistry, University of Washington, Seattle, WA; 3) Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.

Understanding the cause of reproductive barriers that demarcate species (e.g. hybrid lethality and sterility) is essential for understanding the process of speciation. We have discovered a genetic incompatibility between two wild-isolates (JU1825 and NIC59) of *Caenorhabditis nouraguensis* that results in F2 embryonic and larval lethality. Inviability seems to be the result of a maternally inherited cytoplasmic factor from each strain being incompatible with recessive nuclear loci from the other. Furthermore, cytoplasmic-nuclear incompatibility commonly occurs between other wild isolates, indicating that this is a significant reproductive barrier within *C. nouraguensis*. We hypothesize that the maternally inherited factor is the mitochondrial genome and that mitochondrial dysfunction underlies hybrid death.

We have generated genome assemblies of NIC59, JU1825 and JU2079 (an inbred strain derived from JU1825). The JU1825 and NIC59 mitochondrial genomes differ by 95 SNPs, some of which may cause functional differences in two subunits of the

electron transport chain (ND1 and COX1), a cysteine tRNA and a 16s rRNA. We have mapped the nuclear incompatibility loci to single chromosomes by bulk sequencing of viable F2 hybrids. By generating and sequencing recombinant introgression lines, we have more finely mapped one of the nuclear incompatibility loci to a 100 kb region which contains 25 diversified copies of an F-box domain containing gene, 14 diversified copies of a transmembrane domain containing gene that has weak homology to a bacterial protease and six diversified copies of a nuclear hormone receptor (NHR) gene. Interestingly, both F-box and NHR genes are known to be rapidly evolving in *Caenorhabditis* and certain F-box and NHR proteins have been implicated to have mitochondrial functions in mammals. We are currently preparing to perform rescue experiments with fosmid clones to identify the causal variant(s) within the mapped region.

189M Developing *Drosophila melanogaster* as a model for the genetic basis of early-stage reproductive isolation. *M. Lollar* Laboratory of Genetics, The University of Wisconsin-Madison, Madison, WI.

A critical question in evolutionary biology is how distinct populations of the same species progress into reproductively isolated and distinct species. Many researchers have attempted to address these uncertainties by studying largely-isolated *Drosophila* species. However, there are also intrinsic barriers to gene flow that exist between geographically distant *D. melanogaster* populations, particularly those between ancestral African and cosmopolitan European groups. This finding suggests that *D. melanogaster* may be a particularly good model for studying the genetic basis of early stage reproductive isolation. Here, we highlight quantitative and population genetic approaches to finding genes that underlying epistatic incompatibilities between natural populations. First, we provide preliminary results from genomic mapping experiments aimed at uncovering the underlying genetic architecture of F2 lethality, which are consistent with multiple incompatibilities per cross.Second, we extend the population genetic analysis of Pool (2015), which revealed strong evidence of ancestry disequilibrium between unlinked loci in the (admixed) Drosophila Genetics Reference Panel. One explanation of this observed deficiency of African-European allelic combinations may be Bateson- Dobzhansky-Muller incompatibilities, and we thus sought to increase the genomic resolution of these results to reveal new insights into potential epistatic incompatibilities. Single nucleotide gametic disequilibrium between previously identified interactors revealed local peaks of correlation between multiple unlinked neuronal genes, often localizing to specific exonic or intronic gene regions. Many of these loci also displayed high F_{ST} between the African and European source populations, consistent with functional evolution following these popualtions' divergence that may have contributed to the origin of incompatibilities. Taken together, these approaches will yield insights into the genetic basis of reproductive isolation while simultaneously elucidating underlying epistatic interactions at incompatible loci.

190M Maintenance of reproductive isolation in sympatric *Mimulus* species. *S. Mantel*, A. Sweigart Genetics, University of Georgia, Athens, GA.

Understanding how reproductive isolation is maintained in diverged species is integral to understanding species preservation and the evolutionary dynamics of natural sympatric systems. We aim to understand the genetic basis of important traits in sympatric *Mimulus guttatus* and *M. nasutus*, which maintain reproductive isolation. Important such traits include drought response and flowering phenology. Under water-limited conditions we have identified substantial differentiation in survival and seedset between sympatric *M. guttatus* and *M. nasutus* in the greenhouse. Consistent with observed microhabitat differences in the field, *M. guttatus* was affected much more severely under drought conditions. This is indicative of differential evolution of inducible response mechanisms to a water limiting environment in *Mimulus*. We have also shown an association between species-specific genotype at a critical photoperiod candidate gene and flowering phenology. The investigation of these traits holds considerable promise for identifying genomic variation responsible for reproductive isolation in this system.

191M The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *M. Menon*[1], J. Bagley[1], A. Whipple[2], A. Schoettle[3], A. Leal-Saenz[4], C. Wehenkel[4], F. Molina-Freaner[5], L. Flores-Renteria[6], S. Gonzalez-Elizondo[7], R. Sniezko[3], S. Cushman[3], K. Waring[2], A. Eckert[1] 1) Virginia commonwealth University, Richmond, VA; 2) Northern Arizona University; 3) USDA Forest Service; 4) Universidad Juarez del Estado de Durango; 5) Universidad Nacional Autonoma de Mexico; 6) San Diego State University; 7) Instituto Politecnico Nacional.

Interactions between extrinsic factors, such as disruptive selection, and intrinsic factors, such as genetic incompatibilities among loci, can contribute to the maintenance of species boundaries. The relative roles of these factors in the establishment of reproductive isolation can be examined using species pairs characterized by gene flow throughout their divergence history. We investigated the process of speciation and the maintenance of species boundaries between *Pinus strobiformis* and *P. flexilis*. Utilizing genome wide dataset generated from ddRAD-seq, we conducted demographic modeling, and genomic cline analyses to illustrate a history of divergence with continuous gene flow between these species. We found an abundance of advanced generation hybrids, climate associated variation in hybrid index, and a lack of loci exhibiting large allele frequency differences across the hybrid zone. A buildup of intrinsic incompatibilities and of co-adapted gene complexes is also apparent in our results, although these appear to be in the earliest stages of development. This supports previous work in coniferous species demonstrating the importance of extrinsic factors in creating and enforcing species boundaries.

192M Caught in the act? Speciation, divergence and admixture in wild tomatoes. *L.E. Rose*, I. Beddows Institute of Population Genetics, University of Duesseldorf, Duesseldorf, DE.

Hybridization between closely related plant species is widespread, but the long-term outcome of hybridization is not always clear. We have investigated the phylogenetic relationships and the history of hybridization in the wild tomato clade (Solanum sect. Lycopersicon). We sequenced RNA from individuals of 38 different populations and, by combining this with published data, built a comprehensive genomic data set for the entire clade. The data indicate that many taxa are not monophyletic and many individuals are admixed due to repeated hybridization. The most polymorphic species, Solanum peruvianum, has two genetic and geographical subpopulations, while its sister species, S. chilense, shows reduced heterozygosity and much less substructure. Furthermore, we discovered a new set of populations (currently recognized as S. chilense) which are genetically intermediate between S. chilense and S. peruvianum. Based upon molecular, morphological, and crossing data, we tested the hypothesis that these disjunct "S. chilense" populations are an example of recombinational speciation. Recombinational speciation is rarely reported and presents many challenges, both in its unequivocal recognition and difficulty in distinguishing it from other modes of speciation and population history. The discovery of these new cryptic hybrid populations opens new avenues to investigate the genomic outcome of hybridization in plants.

193M Fitness benefits of paternal mitochondrial transmission in intra-species hybrids. Shadi Adineh, *Joseph Ross* Department of Biology, California State University, Fresno, Fresno, CA.

Although mitochondria are predominantly inherited maternally, hybridization often elicits paternal mitochondrial transmission. Hybridization, through separation of co-evolved mito-nuclear epistatic loci, can also cause organismal dysfunction. The molecular phylogeny of the nematode *Caenorhabditis briggsae*, a relative of *C. elegans*, raises the possibility that two clades of populations are beginning to speciate. Based on prior findings of mito-nuclear epistasis existing in inter-clade hybrids, we inquired whether inter-genomic interactions might produce hybrid genetic incompatibility. Paternal mitochondrial transmission occurs in *C. briggsae* during the production of intra-species cytoplasmic-nuclear hybrids (cybrids). We find that cybrids, in which the mitochondrial genome of a maternal population is combined with another's paternal nuclear genome, sustain paternal mitochondrial "leakage" at fertilization. This pattern raises the possibility that, when fitness reduction occurs through separation of co-evolved mitochondrial and nuclear alleles, paternal mitotypes that enter the oocyte at fertilization might restore fitness. Subsequent natural selection for individuals with increasing levels of the paternal mitotype might eventually lead to homoplasmy in the presence of the paternal nuclear background. We show that, when hybridization does not lead to fixation of the paternal mitotype, cybrids retain the maternal nuclear allele at one broad nuclear locus. This result reveals the utility of cybridization as a powerful experimental evolution tool for identifying nuclear intervals that potentially contain loci epistatically interacting with mitotype. Such allele pairs might represent those underlying speciation. Future efforts will focus on better understanding the tempo of paternal mitochondrial transmission and the identification of loci involved in facilitating hybrid paternal mitochondrial transmission. This information will be useful in understanding the evolution of molecular mechanisms controlling uniparental mitochondrial transmission and also identifying alleles that elicit hybrid dysfunction.

194M Investigating hybrid seed lethality between closely related subspecies within the *Mimulus tilingii* complex. *G. Sandstedt*, A. Sweigart Genetics, University of Georgia, Athens, GA.

A major goal in evolutionary biology is to understand how reproductive barriers evolve and prevent gene flow from formerly interbreeding populations. In flowering plants, hybrid seed lethality is a common feature of interploidy and interspecific crosses largely caused by a defective endosperm, a tissue critical in providing nutrients to the embryo, but the evolutionary and genetic mechanisms of this reproductive isolating barrier are largely unknown. To investigate the evolution of hybrid seed lethality at an early stage of divergence, my proposed research focuses on three morphologically distinct subspecies within the *Mimulus tilingii* complex. To quantify reproductive isolation within the species complex, I performed crosses within and reciprocally between each subspecies. I counted the total seeds produced from each cross and scored seed viability based on seed phenotype. I measured viable and inviable seeds to determine differences in overall seed size and capture variation. The results suggest a minor reduction in total seed production, but there is strong evidence of reciprocal differences in hybrid seed viability. Additionally, the variation in hybrid seed phenotypes indicates disruption in endosperm development. The ability to produce successful offspring relies heavily on the parental direction of the hybrid crosses, which suggests evolutionary, developmental, and genetic differences underlying hybrid seed lethality within the *M. tilingii* complex.

195M The genetic and genomic basis of species divergence. *A. Schloop*[1], R. Lyman[1], S. Zhou[1], T. Morozova[1], W. Huang[2], E. Scholl[3], A. Dickey[3], T.F.C. Mackay[1] 1) Department of Biological Sciences and Program in Genetics, North Carolina State University, Raleigh, NC; 2) Bioinformatics Research Center, North Carolina State University, Raleigh, NC; 3) Department of Animal Science, Michigan State University, East Lansing, MI.

Understanding the genetic basis of species divergence remains an unsolved problem in biology. The genetic basis of speciation – the process by which one interbreeding population evolves into two reproductively incompatible populations – is

thought to be due to the accumulation of mutations in each lineage that have deleterious epistatic interactions in the background of the other lineage, leading to hybrid infertility and/or inviability. The species pair *Drosophila melanogaster* and *D. simulans* have been extensively investigated. These species are thought to have diverged 5.4 million years ago. Matings of *D. melanogaster* females to *D. simulans* males yield viable, but sterile females. The reciprocal cross is rarely successful, but when it is, hybrid males are viable but sterile. Two mutations, *Lethal hybrid rescue* and *Hybrid male rescue,* can give viable males in the former cross. Genetic analyses at the level of whole chromosomes indicate that the genetic basis of divergence between these species is more complicated, however. We have recently generated a *D. simulans* Genetic Reference Panel (DSRP) of ~290 lines derived from wild-caught females from the Raleigh, NC Farmer's Market; the DSRP is thus sympatric to the *D. melanogaster* Genetic Reference Panel (DGRP) of 205 inbred, sequenced lines. We observed a wide range of hybrid female viability and aberrant wing phenotypes when the *D. simulans* lines are crossed to multiple *D. melanogaster* strains, indicating that there is natural variation in hybrid performance in this population. We have performed a *de novo* assembly of the *D. simulans* genome using long read PacBio sequencing as well as 100X Illumina sequencing of one DSRP line, and are genotyping 140 sequenced DSRP lines. Using this new assembly and the reference panels, we will perform genome wide association analyses to map genetic variants affecting hybrid viability and wing phenotypes. This will be followed by assessing genome-wide allelic-specific gene expression in hybrids.

196M Using genome data to test the proposed hybrid origin of the bear macaque. *Laurie S. Stevison*[1], Damien Waits[1], Ben Evans[2], Don Melnick[3], Jeff Wall[4] 1) Biological Sciences, Auburn University, Auburn, AL; 2) Biology, McMaster University, Hamilton, ON ; 3) Ecology, Evolution, and Environmental Biology, Columbia University, New York, NY; 4) Institute for Human Genetics, UCSF, San Francisco, CA.

Studies based on phenotypic variation estimated that >10% of primate species naturally hybridize; however, the past 50 years have shown us that genetic analysis often uncovers contradictory results with undetected hybridization. Additionally, Schumer *et al* (2014) found that only four taxa pass their three defined criteria for evidence of hybrid speciation despite many proposed cases found in the literature. Here, we sequenced two endangered macaque species to address the putative hybrid origin of *Macaca arctoides* (the bear macaque). This species has unique phenotypic and behavioral traits, including species-specific genital morphology, making it likely to have evolved under strong selection for reproductive isolation. Based on studies reporting incongruence between mitochondrial and autosomal genealogies, *M. arctoides* was proposed to have evolved via an ancient hybridization event between the ancestor of the modern Fascicularis species group and an ancestor of the Sinica species group. We test the hybrid origin hypothesis using genomic data from five macaque species of both groups, including two sequenced here, *M. arctoides* (21x) and *M. assamensis* (16x), with 5.6 and 5.9 million SNPs, respectively, as compared to the rhesus reference genome. Despite the low relative heterozygosity for *M. arctoides*, this species has a large number of unique variants (>2.1 million) consistent with its proposed hybrid origin. Using these high-quality SNP variants, we conducted a 50kb sliding window analysis for hybridization via *fdM* and D statistics using baboons as an outgroup. These results support a mosaic genome consistent with ancestry from both species groups (*fdM* ranges from -0.72 to 0.68), though the majority of the genome supports Sinica placement (negative values). At increasing window sizes (5kb, 50kb, 100kb, 500kb, and 1Mb), we note a narrower distribution of *fdM* values, but persistent genome mosaicism. We further examine tract lengths of introgression to pinpoint the timing of these migration events and investigate gene categories enriched in outliers supporting placement in each species group. Finally, to investigate whether reproductive isolation evolved as a byproduct of isolation, we examine genes involved in the mechanical barrier, specifically male genitalia. We examine long runs of homozygosity within *M. arctoides* as well as *fdM* values at recently identified mice QTL for baculum size/shape. Therefore, our investigation tests all three defined criteria for hybrid speciation.

197M Male fragility and the genetics of sex-specific hybrid incompatibility in the intertidal copepod *Tigriopus californicus.* *Eric Watson*, Suzanne Edmands Dept. of Biological Sciences, University of Southern California, Los Angeles, CA.

Intrinsic postzygotic isolation critically speeds up the process of speciation, and allows for incipient species to coexist. While many studies investigating the genetics of postzygotic isolation utilize separate species, incompatibilities have been shown to exist as a "polymorphic prelude" in the earlier stages of speciation. Many systems, in which hybrid incompatibility QTL have been identified, indicate a preponderance of hybrid incompatibility phenotypes mapping to the X chromosome in XY (male-heterogametic) sex determination systems. However, in the absence of sex chromosomes, dominance and epistasis are distributed across homologous sets of chromosomes, and it is unclear how this affects the distribution of hybrid incompatibility loci in the genome. Here, we investigate the presence of hybrid incompatibilities between populations of an intertidal copepod*, Tigriopus californicus,* with polygenic sex determination. Using founders from three geographically isolated populations, we constructed F_4 recombinant inbred lines using a sibling mating design. Offspring were genotyped at several thousand loci using ddRADseq, and candidate hybrid incompatibility QTL were identified using marker transmission ratio distortion (TRD). We found many more putative hybrid incompatibility QTL in males than in females, indicating the presence of sex-specific mutation load contributing to hybrid incompatibility. Our results shed new light on theory stating that, while genes responsible for hybrid sterility should differ between the sexes, those responsible for hybrid inviability are likely shared.

198M Genetic Architecture of Hybrid Male Sterility in the Mouse. *S.J. Widmayer*[1], D.L. Aylor[1,2] 1) Department of Biological Sciences, North Carolina State University, Raleigh, NC; 2) Bioinformatics Research Center, North Carolina State University, Raleigh, NC.

The formation of reproductive barriers between diverging groups is a key component of species formation. Hybrid male sterility (HMS) is a reproductive barrier that restricts gene flow between two subspecies of mice, *Mus musculus musculus* and *M. m. domesticus*. Crosses between inbred strains derived from these subspecies recapitulate HMS in the lab. Two major loci have been previously linked to HMS, but we observed wide variation in fertility and reproductive traits among hybrids with identical genotypes at those loci. We characterized this variation in a panel of hybrid males bred by crossing *musculus*-dervied PWK/PhJ strain females to males from four inbred mouse strains of primarily *domesticus* origin. These PWK-derived hybrids ranged from complete sterility to complete fertility. In addition, some hybrids exhibited age-dependent fertility that is characterized by delayed-onset of fertility and premature sterility. Taken together, our results support the hypothesis that HMS in the mouse is a polygenic trait and undiscovered modifiers drive key differences in sterile hybrids. We have begun to employ a novel QTL mapping design to identify the modifier alleles driving these differences in sterility phenotypes between hybrids.

199M Tracking the trends in white nationalist misappropriation of population genetics research. *J. Carlson* Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

For decades, white nationalist movements have attempted to validate their racism by appealing to flawed interpretations of population genetics and evolution research, sometimes with the complicity of the researchers themselves. Emergent white nationalist groups (i.e., the so-called “alt-right”) have rekindled this intense interest in the field, with several prominent white nationalists claiming that “human biodiversity” and “race realism” are cornerstones of their ideology. These pseudoscientific interpretive frameworks give racism a veneer of scientific validity, and many members closely follow the latest research in population and evolutionary genetics to cherry-pick and reinterpret results in a way that supports their agenda.

Many in the academic community and beyond have been concerned about the recent surge in white nationalist activity via highly publicized rallies and speaking events (which can turn violent, such as the “Unite the Right” rally in Charlottesville, Virginia last August), widespread recruitment efforts on college campuses, and omnipresence on social media. These trends have prompted several questions about how population geneticists might mitigate and/or respond to racist misappropriation of their results and educate the public about these issues.

By analyzing data I collected from social media and white nationalist forums, I identify three growing trends of white nationalism as it pertains to the field of population genetics. First, I discuss the behavior of informal “journal clubs” which have arisen in these communities, where members read and debate recent findings in the primary scientific literature. Second, I document the practice of unauthorized editing of figures from the population genetics literature and their dissemination as a white nationalist propaganda device on social media. Third, I describe the increasing antagonism of white nationalists towards researchers (particularly women and minorities) who publicly admonish scientific racism or publish results which are incompatible with white nationalist beliefs. I discuss the implications of these findings for the population genetics community in the context of our individual and collective responsibilities as scientists and stewards of knowledge.

200M Evidence that genotype-by-temperature interactions maintain polygenic sex determination in the housefly. *K. Adhikari*, R. Meisel Department of Biology and Biochemistry, University of Houston, Houston, TX.

Sex determination (SD) systems vary across taxa because they are rapidly evolving. A single master regulatory locus is enough to determine sex in most organisms. However, many organisms have multiple master SD loci in their genome that segregate independently, resulting in polygenic SD. Polygenic SD systems are predicted to be unstable intermediates between monogenic SD systems, and the factors responsible for maintaining polygenic SD are poorly understood. Housefly (*Musca domestica*) has a stable polygenic SD system with multiple male and female determiners segregating in natural populations. The male determining factor (*Mdmd*) is commonly found on the Y chromosome (Y^M) and the third chromosome (III^M). Y^M males are found in colder, northern latitudes whereas III^M males are found in southern, warmer latitudes. This suggests that selection operating on a genotype-by-temperature (GxT) interaction maintains this polymorphism. To test this hypothesis, we raised III^M and Y^M males at high and low temperatures, and we studied the resulting GxT effects on two phenotypes: gene expression and extreme temperature tolerance. Using RNA-seq, we found that most variance in testis gene expression can be attributed to temperature and genotype. We identified 12 genes whose expression in testis and 2 genes whose expression in head depends on GxT interactions. We also found that III^M males raised at high temperature are the most tolerant to extreme heat and Y^M males raised at low temperature are the most cold tolerant, suggesting a GxT interaction that includes developmental acclimation. The direction of this GxT interaction is consistent with the III^M chromosome providing a fitness advantage at warmer temperatures and the Y^M chromosome increasing fitness at lower temperatures. Our results therefore support the hypothesis that GxT interactions maintain polygenic SD in housefly through

temperature-dependent fitness effects of genetic variation on the Y^M and III^M chromosomes. Our RNA-seq data suggest that the GxT interactions could be mediated through effects on gene expression.

201M The genetic basis of morphological and behavioral island syndrome traits in deer mice. *F. Baier*, H. E. Hoekstra Department of Molecular and Cellular Biology, Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, HHMI, Harvard University, Cambridge, MA.

Animal populations on islands often exhibit dramatic differences in morphology and behavior compared to mainland populations, a phenomenon known as the "island syndrome". These differences are likely adaptations to environmental conditions on islands, notably altered resource and population densities. However, the extent to which island traits have a genetic basis, indicative of past selection processes, or whether they instead represent plastic responses to environmental extremes has rarely been investigated. Here, we revisit a classic case of island syndrome in deer mice (*Peromyscus maniculatus*) in British Columbia. Previous studies suggested that Saturna Island mice exhibit several island traits, including larger body weight and reduced aggression relative to their mainland ancestors. First, using historical collections, we show that body weight in Saturna Island mice is representative of neighboring island populations, and ~5g (~35%) greater than mainland populations. We next collected deer mice on Saturna Island and the mainland to establish laboratory colonies, and found that Saturna Island mice are heavier both because they are longer and have disproportionately more lean mass. Importantly, these traits are maintained in captive-born mice raised under common conditions, implying a strong heritable component. In addition, F1 hybrids are heavier when born to island mothers than to mainland mothers, revealing a maternal effect on body weight. Second, using behavioral testing in the lab, we also find that wild-born island mice are less aggressive than mainland mice, consistent with previous studies. Remarkably, however, lab-raised male mice born to these founders do not differ in aggression, even when tested shortly after mating with a receptive female, suggesting this behavior difference is not genetic. Together, our results suggest that rodents respond variously to environmental conditions on islands by evolving heritable changes in offspring and maternal genotypes determining a morphological trait, while expressing non-heritable phenotypic responses in a behavioral trait.

202M Parallel adaptation in two co-distributed species along a latitudinal cline. *M.A. Ballinger*, M.W. Nachman Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, CA.

Understanding how natural selection shapes patterns of genetic and phenotypic variation is one of the central goals of evolutionary biology. One approach is to show the repeated evolution of traits in similar environments. A comparative study between house mice (*Mus musculus*) and deer mice (*Peromyscus maniculatus*; ~25 MY diverged) provides a unique opportunity for studying the genomic basis of environmental adaptation: both species are co-distributed, genetically tractable, and are powerful models for understanding the genetic basis of many traits over short and long evolutionary timescales. To determine the genetic mechanisms underlying environmental adaptation, we collected 191 house mice and deer mice from 5 populations along a latitudinal cline, spanning 20 degrees of latitude from southern Texas to northern Minnesota. Both house mice and deer mice collected from various latitudes differed in body size and other morphological traits, such as relative ear length and tail length. We also identified morphological differences associated with tail length in *Peromyscus* populations, likely underlying differences in various *Peromyscus maniculatus* subspecies. Lastly, we sequenced the *Mus* exome at moderate coverage to identify loci contributing to adaptation. Future directions include sequencing and comparing the *Peromyscus* exome to the *Mus* exome dataset to look for convergent, environmental adaptation.

203M Predation induced adaptation in *Daphnia* populations across time and space. *D. Becker*[1], K. Barnard-Kubow[1], A. Edwards[1], A. Beckerman[2], A. Bergland[1] 1) Department of Biology, University of Virginia, Charlottesville, VA; 2) Department of Animal and Plant Sciences, University of Sheffield, UK.

Adaptive evolution is a basic evolutionary process that is now understood to be a common occurrence. Adaptation affects the extent and nature of standing genetic variation and contributes to phenotypic differences between populations. While adaptive shifts are known to result in distinct phenotypes and genotypes, our understanding is still limited on how environmental heterogeneity contributes to the maintenance of functional genetic variation in natural populations. In the present study, we address these limitations by investigating predation induced adaptation in the freshwater invertebrate *Daphnia,* an ecological and genetic model system, over temporal and spatial scales. Midge predation is a substantial selective force for *Daphnia pulex* populations, in response to which daphnids induce neckteeth, providing a clear fitness benefit due to reduced predation. Midge density, and hence predation pressure, varies over *D. pulex*'s growing season, resulting in temporal heterogeneity in selection pressure for the induction of neckteeth. Using an automated picture analysis pipeline, we examined phenotypic changes to predation pressure (i.e., neckteeth induction and shifts in life-history traits) in daphnids originating from natural populations in the UK. We show that *D. pulex* collected from different ponds and across consecutive years vary in their responses to predation. Specifically, we show that substantial variation for the induction of neckteeth exists within and among populations. We further examine whether seasonal and spatial heterogeneity in selection pressure promotes these adaptive shifts in *D. pulex.* Ultimately, our work will allow to assess how, and to what extent, environmental heterogeneity influence the maintenance of functional genetic diversity in populations across time and space.

204M Expanding functional horizons or evolutionary combat? Evaluating drivers of the rapid evolution of bag of marbles in Drosophila. *J. Bubnell*, C. Aquadro Molecular Biology and Genetics, Cornell University, Ithaca, NY.

In D. melanogaster, bag of marbles (bam) acts as the master switch for germline stem cell differentiation during oogenesis and plays a key role in regulating spermatogenesis. bam is rapidly evolving across the Drosophila genus and shows bursts of positive selection in D. melanogaster and D. simulans but not in D. ananassae, D. pseudoobscura, and D. mojavensis. We have resequenced bam in populations of the three closely related species in the yakuba complex— D. yakuba, D. santomea, and D. teissieri. Our results indicate that like in D. melanogaster and D. simulans, bam in D. yakuba and D. santomea is evolving under positive selection. However, we do not see this signal at bam in D. teissieri. What could be driving the episodic adaptive evolution of a gene essential for stem cell differentiation and fertility? One possibility is that bam's role as a differentiation factor in gametogenesis is novel to the melanogaster subgroup and is now under positive selection in these species. To evaluate this hypothesis, we are using CRISPR-Cas9 to generate bam null alleles in diverse Drosophila species and then testing for conservation of function. We have found that bam's core function in gametogenesis is conserved in D. melanogaster, D. yakuba, and D. ananassae. Together, our functional and population genetic data suggest that bam's function as a stem cell differentiation factor is not driving its adaptive evolution. The episodic signal of positive selection we observe at bam implies adaptation to pressures present in these particular lineages. Wolbachia is a maternally inherited endosymbiotic bacteria that transiently infects Drosophila species, manipulates host reproduction to ensure its propagation, and genetically interacts with bam during oogenesis in D. melanogaster. Episodic natural infections with Wolbachia could result in genetic conflict with bam, leading to an evolutionary arms race for control of oogenesis. We are currently working to define the interaction between D. melanogaster bam and Wolbachia, identify possible genetic interactions between bam hypomorphic mutants and Wolbachia in D. simulans and D. yakuba, and analyze patterns of evolution at bam in species with known Wolbachia infection histories.

205M Genomic basis of adaptive island dwarfism in Boa constrictor snakes. *Daren Card*[1], Richard Adams[1], Drew Schield[1], Giulia Pasquesi[1], Blair Perry[1], Andrew Corbin[1], Kristopher Row[1], Juan Daza[2], Warren Booth[3], Chad Montgomery[4], Scott Boback[5], Todd Castoe[1] 1) Department of Biology, University of Texas, Arlington, Arlington, TX; 2) Instituto de Biologia, Universidad de Antiochia, Medellin, Colombia ; 3) Department of Biological Science, University of Tulsa, Tulsa, OK; 4) Department of Biology, Truman State University, Kirksville, MO; 5) Department of Biology, Dickinson College, Carlisle, PA.

Island fauna present natural experiments that can be used to study evolution and adaptation, especially in instances where the adaptive phenotype is extreme and replicated across independent insular systems. Boa constrictors are typically large bodied snakes found across the Americas, yet multiple insular populations on geographically distant island systems show drastically dwarfed phenotypes. This dwarf phenotype appears to be adaptive and associated with the highly arboreal nature of snakes from these island populations, and common-garden experiments indicate the phenotype is largely driven by genetics. Here we used mitochondrial gene sequences and nuclear genomic variants to infer population genetic structure within boas, focusing primarily on Northern and Central American populations. Our results indicate that the dwarf island phenotype has evolved multiple times on different island clusters off the coast of Central America. We also used a combined dataset that included dense sampling of RADseq markers and whole-genome sequencing to characterize genome-wide variation, and we identified multiple loci that appear to be important for driving island dwarf phenotypes that are also known to play roles in developmental abnormalities and dwarfism phenotypes in mammals. We also used our genome-scale datasets to address the possible role that convergent molecular evolution may have played in driving dwarfism phenotypes in isolated populations. Our work establishes a new model system for interrogating a long-recognized and important evolutionary phenomenon, and also underscores the value that island populations can provide for understanding fundamental questions in evolutionary biology.

206M Does historical admixture predict patterns of introgression in a contemporary hybrid zone? A comparison of recent and ancient admixture in *Lycaeides* butterflies. *S. Chaturvedi*[1,2], L. Lucas[1], Z. Gompert[1,2] 1) Biology, Utah State University, Logan, UT; 2) Ecology Center, Utah State University, Logan, UT.

Studies of replicate hybrid zones have found evidence of consistency and variability in patterns of introgression, but these studies generally focus on hybrid zones of a similar age. Whether there is consistency in patterns of introgression over time (at different stages of hybrid zone formation) is less clear. Consistency across time (stages of hybrid zones) could imply consistency in the genetic and ecological factors contributing to reproductive isolation and increased predictability of evolution, and perhaps speciation, in general. Here, we use relatively old admixed populations of *Lycaeides melissa* and *Lycaeides idas* butterflies (admixture occurred about 14,000 ybp) and populations from a recent, active hybrid zone to ask if evolutionary patterns in old admixed populations can predict evolutionary dynamics in the current hybrid zone. We addressed two specific questions a). How well can we predict genetic regions which are most resistant to gene flow in recent active hybrids from patterns of ancestry in the admixed populations? b). How well can we predict genetic regions resistant to introgression in both old and new hybrids from known QTL for ecologically important traits? We generated genotyping-by-sequencing data from approximately 25 populations (*L. melissa*, *L. idas*, and admixed populations) to characterize patterns of ancestry and introgression in the old admixed populations and current hybrid zone. We then compared these patterns

among populations and asked whether genetic regions with exceptional ancestry frequencies or reduced introgression co-occurred with QTL or genes known to affect traits that contribute to RI in this system (e.g., diapause, host plant preference/performance, or wing pattern).

207M Temperature adaptation of Arabidopsis thaliana rosette growth. *P. Clauw*, J. Jagoda, I. Reichardt-Gomez, M. Nordborg Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna.

Climate variation exerts significant evolutionary pressure on plants. *Arabidopsis thaliana* has a large geographic range and local varieties have acquired specific adaptations to their local environment. As a winter-annual, *A. thaliana* grows during autumn under decreasing temperatures that are usually considerably lower than standard lab conditions. Northern varieties may never experience temperatures above 10ºC until they are about to flower in spring. Although we expect northern strains (natural inbred lines, or "accessions") to be adapted to lower temperatures, the importance of temperature adaptation and the mechanisms involved are not known.
To investigate the importance of adapting growth to lower temperatures we exposed a global set of *A. thaliana* accessions to simulated Swedish autumn temperatures. Over three months we gradually let the temperature decrease from 16 to 0ºC. Growth was phenotyped by daily taking pictures of the rosettes. At 16ºC the accessions originating from warmer regions clearly grew better than accessions from colder climates. When colder temperatures were reached, by the end of the simulated Swedish autumn, accessions from colder climates showed improved growth. We further confirmed this by characterising growth at constant temperatures. We saw that accessions from colder regions were less reduced in growth rates in colder temperatures, relative to their growth in warm conditions.
Phenotypically we have indications that growth may be adapted to local temperatures. Currently, we further investigate what the genetics are behind the temperature adaptation of this highly complex trait.

208M Nematode sampling across the Hawaiian Islands reveals niche preferences for *C. elegans* and a new *Caenorhabditis* species. *T.A. Crombie*[1], S.C. Brady[1,2], D.E. Cook[1,2], K.S. Evans[1,2], S. Hahnel[1], D. Lee[1], B.C. Rodriguez[1], R.E. Tanny[1], S. Zdraljevic[1,2], E.C. Andersen[1,2] 1) Department of Molecular Biosciences, Northwestern University, Evanston, IL; 2) Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL.
Caenorhabditis is a diverse genus of nematodes that includes the well studied model *C. elegans*. Until recently, the ecology and niche preferences for nematodes in this genus were largely understudied. Consequently, many of the existing wild isolates originate from opportunistic sampling of anthropogenic habitats. Could this small, biased sampling effort play a role in the low genetic diversity associated with some *Caenorhabditis* taxa such as *C. elegans*? To address this question and elucidate niche preferences within the genus, we extensively sampled nematodes from the Hawaiian Islands, because many of the most genetically divergent *C. elegans* wild strains come from this location.

We sampled nematodes at 2,263 collection sites across five islands, recording GPS coordinates and various niche parameters, including temperature, moisture, elevation, and substrate type. Because of the large scale of this sampling effort, data collection was streamlined using a customized data-collection application for mobile devices (Fulcrum ®). In total, we isolated 2,531 distinct wild isolate strains to be genotyped in the lab. Wild isolates were genotyped by PCR to identify nematodes in the Rhabditid family, then Rhabditid-positive PCR products were sequenced to distinguish *Caenorhabditis* nematodes at the species level. Of 2,263 total collection sites, 233 (10%) contained Rhabditid nematodes. Of these, 157 collections contained *Caenorhabditis* nematodes, including five separate species; *C. elegans*, *C. briggsae*, *C. tropicalis*, *C. kamaaina*, and a new species, *C. sp. 53*. This new species was typically found on flower substrates (10 of 12 collections) indicating a possible substrate preference. *C. elegans* was commonly found on fruits, nuts and leaf litter, which challenges the hypothesis that it is infrequently isolated from leaf litter. At three collection sites, we found two *Caenorhabditis* species on the same substrate, illustrating some niche overlap within the genus. Analysis of niche parameters suggests that *C. tropicalis* and *C. sp. 53* are restricted to warmer substrates, while *C. elegans* prefers higher elevation and cooler environments than other members of the genus. These findings begin to clarify niche preferences among *Caenorhabditis* species. Furthermore, whole-genome sequencing of the 95 *C. elegans* wild isolates collected in this study will reveal additional genetic diversity and facilitate improved population genomic analyses in this species.

209M The relative importance of phenotypic plasticity, trans-generational effects and selection in heat shock responses of the cactophilic *Drosophila mojavensis*. *Fernando Diaz*[1], Joshua Coleman[2], Nathaniel Talamantes[3], Luciano Matzkin[1,4,5] 1) Department of Entomology, The University of Arizona, Tucson, AZ; 2) Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL; 3) Department of Molecular and Cell Biology, The University of Arizona, Tucson, AZ; 4) Department of Ecology and Evolutionary Biology, The University of Arizona, Tucson, AZ; 5) BIO5 Institute, The University of Arizona, Tucson, AZ.
A wealth of evidence has been collected illustrating how environmental conditions affect phenotype by selection, plasticity

and trans-generational effects. Despite their impact on evolution, the relative importance of these mechanisms, as well as the underlying ecological context for which plasticity and trans-generational effects arise and interact, is poorly understood. In order to test for such interactions, we considered the heat shock response of adults and larva of two genetically and ecologically distinct populations of the cactophilic *Drosophila mojavensis*. The populations differ in their thermal exposure (mean and maximum) and variance (diurnal and annual), whereas populations from Sonora, Mexico experience higher temperatures and variance relative to individuals from Catalina Island, California. Trans-generational effects were tested by subjecting the parents to 36°C for 24 hours prior to oviposition, while plasticity (in both larvae and adults) was determined by acclimating (at 36°C) individuals prior to thermal stress test. Thermal responses, measured by ramping heat shocks up to 40°C, showed substantial variation due to population as well as plasticity and trans-generational effects. Increased performance during acclimation and trans-generational treatments were additive when assessed on the Sonora population, while trans-generational effects were only evident in Catalina in the absence of an acclimation treatment. These findings highlight that the ecological context can shape not only the phenotypic responses to a stress environment, but also the influence of carry-over effects. Selection in a population evolving under high temperatures and variance appears to increase thermal responses by acclimation and trans-generational effects.

210M Evolutionary Adaptation at Ecological Time Scales: Seasonal Allele Frequency Changes among Microhabitats of the Killifish, *Fundulus heteroclitus*. *M.A. Ehrlich*, D.N. Wagner, D.L. Crawford, M.F. Oleksiak Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL.

The relative importance of evolutionary change at ecologically relevant time scales (1 to 10s of generations) is unclear. High prevalence of such rapid genetic change would alter our perspective of how organisms will respond to global climate change and enhance our understanding of conservation genetics and the human condition. To provide insights into the evolutionary processes affecting short-term, ecological adaptation, we investigated the genetic variation among microhabitats within a marine saltmarsh estuary. These microhabitats, which include coastal basins, intertidal creeks and tidal ponds, present distinct environmental parameters: ponds are up to 4°C hotter during daytime, anoxic at night, and subject to greater predation pressure than the less variable basins. We contrast resident *Fundulus heteroclitus* from basin versus ponds in spring and late summer to address seasonal allele frequency changes. We hypothesise that pond and/or basin residents display a seasonal shift in allele frequencies, which explains the previously established genetic divergence between these microhabitats. We employ GBS to determine SNP genotypes from 240 tagged individuals (1 coastal basin and 3 tidal ponds). Our data indicate a change in allele frequencies between spring and fall that is most parsimoniously explained by natural selection. These data support the supposition that a combination of small allele frequency changes at multiple loci can lead to rapid phenotypic shifts that confer a selective advantage at short, ecological time scales.

211M Hybrid swarm-based association mapping and evolution of ovarian diapause in *Drosophila melanogaster*. *P. A. Erickson*, D. Y. Song, H. M. Stone, A. O. Bergland University of Virginia, Charlottesville, VA.

Organisms living in temperate environments utilize seasonal changes in environmental cues such as light and temperature to exploit and avoid favorable and unfavorable times of year. While the diversity of physiological and behavioral strategies to life in seasonal environments has long been appreciated, the genes underlying perception of seasonal turnover are elusive, particularly for insects. Using a novel hybrid swarm-based genome-wide association study, we are examining the genetic basis and evolutionary history of ovarian diapause (an arrest of reproductive development) in *Drosophila melanogaster*. We generated two outbred populations representing eastern North American clinal diversity by intercrossing 68 sequenced, inbred lines for 4-5 generations. We exposed hybrid females to different temperatures and day lengths in Raspberry Pi-controlled chambers and phenotyped 2800 flies for ovary development. We have reconstructed full phased genomes for all phenotyped individuals using low-coverage (~0.5X) Illumina sequencing data, enabling us to perform a genome-wide association study for loci controlling environmentally-induced diapause. While we found little effect of photoperiod on diapause, we have identified multiple loci with strong temperature-dependent effects. We have placed the outbred populations in outdoor mesocosms and will use pooled sequencing to assess overwintering survival and the fitness costs and benefits of loci associated with diapause. We will also examine allele frequency at diapause-associated loci in wild North American populations sampled in throughout spring and fall. These integrative experiments will shed light on the molecular mechanisms insects use to detect seasonal changes and will provide valuable insight into the evolutionary history of this ecologically relevant trait.

212M *Genetics of parallel leaf shape evolution in the Mimulus guttatus species complex*. *Kathleen G. Ferris*, Graham Coop, Johanna Schmitt Center for Population Biology, University of California, Davis, Davis, CA.

Parallel evolution, or the independent evolution of similar phenotypes in organisms occupying similar environments, is strong evidence of adaptation. Whether convergent phenotypes are controlled by the same genetic loci and mutations, and therefore whether evolution is predictable at the molecular level, is a central question in evolutionary biology. To address this question we examine the genetic and adaptive significance of parallel leaf shape evolution across and within species in the *Mimulus guttatus* species complex. The genetic basis of leaf shape has long interested evolutionary biologists because leaf

shape varies extensively across the plant kingdom and is hypothesized to be adaptive. Lobed and narrow leaves have evolved from the entire, round leaves of *Mimulus guttatus* in *M. laciniatus, M. nudatus*, and a polymorphic serpentine *M. guttatus* population. In addition to having divergent leaf shapes, all three of these taxa occur in harsh rocky habitats suggesting that leaf shape variation is adaptive in rocky outcrops. We (1) examined patterns of leaf shape variation across replicated altitudinal clines in *M. laciniatus* to detect within species local adaptation in leaf shape, (2) used reciprocal transplants and phenotypic selection analysis to test whether lobed leaf shape is adaptive in *Mimulus laciniatus'* granite habitat, (3) used quantitative trait locus (QTL) mapping and genome wide association to examine whether leaf shape evolution has a similar genetic basis across three rocky outcrop taxa in the *M. guttatus* species complex. We found that (1) leaf lobing appears to be under parallel selection across multiple altitudinal clines in *M. laciniatus,* (2) lobed leaf shape is adaptive in *M. laciniatus*' granite outcrop habitat, and (3) leaf shape is controlled by overlapping genetic regions in all three species. Therefore we found parallel phenotypic evolution in leaf shape both within and between species in the *M. guttatus* complex suggesting that this is a key adaptive trait. We see parallelism at the genetic level across three separate Mimlus taxa suggesting that among these closely related species adaptive evolution can be predictable at the molecular level.

213M Adaptation of commensal gut bacteria in the human microbiome. *N. Garud*[1,8], B. Good[2,3,5,8], O. Hallatschek[2,4,5], K. Pollard[1,6,7] 1) Gladstone Institutes/UCSF, San Francisco, CA; 2) Department of Physics, University of California, Berkeley, CA; 3) Department of Bioengineering, University of California, Berkeley, CA; 4) Department of Integrative Biology, University of California, Berkeley, CA; 5) Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA; 6) Department of Epidemiology & Biostatistics, University of California, San Francisco, CA; 7) Institute for Human Genetics and Institute for Computational Health Sciences, University of California, San Francisco, CA; 8) Equal Contributors.

The evolutionary dynamics of commensal gut microbiota within and across hosts are still yet to be fully understood. A challenge in analyzing pooled short-read sequencing of a complex microbial community is that it is difficult to resolve evolutionary changes between specific lineages. Here we introduce a model-based framework to confidently identify evolutionary mutations on the background of individual lineages from pooled whole genome shotgun metagenomic data. Within hosts, we find evidence for rapid sweeps of SNPs and genes to high frequency within the 6 month sampling window. By comparing the signatures of sweeping mutations within hosts with the typical between-host differences that have been subject to purifying selection on long time scales, we find evidence that adaptation is largely driven by introgression from existing strains, rather than new mutations. Our results suggest that gut bacteria can evolve on human-relevant timescales, and highlight the feedback between short- and long-term evolution across hosts.

214M Population genomics of experimental transplants identifies the genetic basis of adaptation in Mexican maize landraces. *Daniel Gates*[1], Sarah Hearne[2], Jeffrey Ross-Ibarra[1] 1) Plant Sciences, UC Davis, Davis, CA; 2) International Maize and Wheat Improvement Center (CIMMYT), Carreta Mexico-Veracruz, El Batan, Mexico.

Understanding how plant and animal populations adapt to their environments is fundamental in predicting how organisms will face the challenges of a rapidly changing global climate. Despite the fact that adaptation has been well demonstrated for over a century, only recently has data become available that allows for identification of the genetic elements that contribute to highly polygenic fitness traits. Here we describe how the combination of population genomics and reciprocal transplant fitness experiments can be used to understand the nature and distribution of adaptive alleles across maize landraces in Mexico. We use phenotypic data gathered from reciprocal transplants to show fitness trade-offs between different environments. From this data, we use genome wide association analyses to identify the loci that explain observed fitness changes across environments. We find that many such loci show patterns of antagonistic pleiotropy in that they are beneficial at one end of the temperature spectrum and deleterious at the other end. Furthermore, we see that individuals of both maize and teosinte have higher frequencies of beneficial alleles in environments that are predicted to be beneficial. This suggests that adaptation in maize has likely proceeded from ancestral standing genetic variation and variation is being maintained over time in part by adaptation selecting for alternative allele copies in different environments.

215M Host genetic control of the microbiome in wild primates. *L. Grieneisen*[1], M. Dasari[2], T. Gould[1], J. Bjork[2], J-C. Grenier[3], V. Yotova[3], D. Jansen[2], N. Gottel[4], J. Gilbert[4], L. Barreiro[3], J. Tung[5], E. Archie[2], R. Blekhman[1,6] 1) Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN; 2) Department of Biology, University of Notre Dame, Notre Dame, IN; 3) CHU Sainte Justine Research Center, Université de Montréal, Montreal, Québec, Canada; 4) Department of Surgery, University of Chicago, Chicago, IL; 5) Department of Evolutionary Anthropology, Duke University, Durham, NC; 6) Department of Ecology, Evolution, and Behavior, University of Minnesota, Minneapolis, MN.

Variation in gut microbiome composition—defined as the identity and abundance of different bacterial taxa—is linked to key health outcomes in humans and other vertebrates. Variation in the gut microbiome is affected by many environmental factors, but host genetics also has a significant impact on the microbiome. Understanding the relative roles of genetic and environmental factors is a central goal in human disease research. However, it is often difficult to disentangle genetic from environmental effects because relatives often use the same resources and have similar diets. To address this issue, the goal of our study was to identify components of the gut microbiome that are associated with host genetics—i.e. heritable

microbes—controlling for changes in the host's physical and social environments over time. Specifically, we used an unprecedented dataset of >20,000 microbiome profiles collected over 15 years from a well-studied population of wild savannah baboons (*Papio cynocephalus*) with available host genetic relatedness data. Using a pilot dataset of 557 samples, we found that gut microbial heritability changes over time and across environments. Specifically, the overall microbiome community composition is affected by baboon social group membership and hydrologic season. Additionally, we found that pairs of baboons who were more related have more similar abundances of several common bacterial taxa, including Christensenellaceae, a bacterial family that is also highly heritable in humans, suggesting that common host genetic mechanisms may control microbiome composition across primates. Finally, the relative abundance of heritable taxa changes over time within individuals, emphasizing the importance of longitudinal studies to account for temporal and environmental effects in shaping gut microbial composition. To summarize, we present the first analysis of host genetic control of the gut microbiome in non-human primates, indicating that microbiome heritability is affected by fluctuating environment, and suggesting that heritable microbial taxa may be shared across primate species.

216M Effect of gut microbiota on α-amanitin tolerance in mycophagous *Drosophila*. *L. Griffin*, C. Scott-Chialvo, O. Fish, L. Reed Department of Biological Sciences, University of Alabama, Tuscaloosa, AL.

Novel adaptations can allow organisms to escape from selective pressures, occupy new ecological niches, and undergo adaptive radiation. While the evolution of novel morphological adaptations is a well-studied phenomenon, novel biochemical traits are not as well understood. One such biochemical adaptation can be found in several species of mycophagous *Drosophila*. Some of these flies are tolerant to the amatoxins (including α-amanitin) found in mushrooms of the genus *Amanita*, despite these compounds being lethal to all other eukaryotes. The gut microbiome of many animals, especially insects, is known to be important for many physiological functions including detoxification; thus, it is possible that the microbiome could contribute to the detoxification of α-amanitin in these flies. However, α-amanitin toxicity is believed to be specific to eukaryotes because it binds and inhibits RNA polymerase II. This study assesses the potential role of the microbiome in toxin tolerance in six inbred wild strains of *D. tripunctata*. Normal and axenic samples of six strains were reared on diets with and without α-amanitin, and then scored for survival from the larval stage to adulthood. Our results show host genotype and toxin-specific effects on survival and help to characterize the potential contribution of the microbiome to this unique adaptation.

217M Systems genetics of rice adaptation to drought stress. *Simon Groen*[1], Zoe Joly-Lopez[1], Jae Choi[1], Adrian Platts[1], Irina Calic[2], Steven Franks[2], Rahul Satija[1,3], Amelia Henry[4], Georgina Vergara[5], Michael Purugganan[1] 1) Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY; 2) Department of Biological Sciences, Fordham University, New York, NY; 3) New York Genome Center, New York, NY; 4) Crop and Environmental Sciences Division, the International Rice Research Institute, Manila, Philippines; 5) Plant Breeding, Genetics and Biotechnology Division, the International Rice Research Institute, Manila, Philippines.

Drought stress is one of the most important factors that curb global crop production, and its consequences are project to become more dramatic with population growth and climate change. Efforts have been made to understand plant responses to drought stress, yet these have had limited succes in leading to new, stress-tolerant crop varieties.
Asian rice, the staple food for half of the world's population, is grown in relatively stress-free, irrigated conditions or stress-prone, rainfed environments. The aim of my research is to identify genes associated with adaptation to dry conditions through comparative analyses of traditional rice varieties that have contrasting adaptation to irrigated or rainfed environments. The discovery of more than one adaptive strategy will increase the options rice breeders could implement in their programs.
To obtain a fine-grained picture of the genetic architecture underlying adaptive drought tolerance strategies, we have taken a systems genetics approach. We phenotyped a diversity panel of locally adapted, traditional varieties in irrigated and water-limited fields for thousands of molecular, physiological and morphological traits. I will present progress on explicitly connecting gene regulatory variation to higher-scale phenotypes and grain production (fitness) under stressful conditions.

218M Predator-induced transgenerational plasticity is mediated by underlying changes in gene expression and methylation. *Nicole Hales*, Todd Castoe, Matthew Walsh, Drew Schield, Daren Card, Audra Andrew Biology, UT Arlington, Arlington, TX.

Organisms have the ability to alter heritable phenotypic traits in response to environmental signals without altering the DNA sequence of their genome through epigenetic genome modifications. Epigenetic changes, which are encoded by altering the chemistry of particular DNA bases in the genome, may control the expression of genes, and thus the phenotypes that are determined by such genes, and epigenetic changes can be passed down for multiple generations. We have recently developed a new genome-scale approach to accurately measure such epigenetic changes, and tested this approach on the zooplankton *Daphnia*. Previous work has shown that Daphnia rapidly alter their phenotype if exposed to particular environmental cues associated with predators. In this study we exposed Daphnia to these cues and tested for changes in gene expression and epigenetic state in successive generations of *Daphnia*. We found that there is a significant change in

epigenetic states across generations that experience an environmental cue in only the first generation. These epigenetic and phenotypic shifts also correspond to multi-generational shifts in gene expression. This study provides exciting new evidence for the multigenerational altering of phenotypes based on reception of environmental cues through an epigenetic mechanism.

219M Evolution of gene expression and plasticity in parallel cold-adapted *Drosophila* populations. *Y. Huang*, J. Pool Genetics, University of Wisconsin-Madison, Madison, WI.

Gene expression change has shown to be important for adaptive evolution. But the repeatability of gene expression evolution for adapting to similar environments is largely unknown. Further, whether such expression evolution reinforces or opposes the initial plasticity for new environment is debated. To study these questions, we used three pairs of natural *D. melanogaster* populations that represent three parallel cases of adaptation to colder environments in this ancestrally Afrotropical species; each pair included closely-related warm- and cold-adapted populations. To examine the repeatability of gene expression evolution, we surveyed the transcriptomic abundance of each population at cold condition (15 °C, resembling the natural cold environment) for 3rd instar larvae, pupae and adult females. We found evidence for parallelism for larvae samples in that the genes with strong differentiation between populations for any two warm/cold pairs were more likely to show codirectional change (up or down regulated in both cold-adapted populations). However, the pupal and adult stages did not show such pattern. To study expression plasticity, we also measured the transcriptomes for female adults at warm condition (25°C) and compared that to those of the same genotype at cold condition. We found that the plasticity of all three warm-adapted populations was generally reinforced by subsequent evolution, suggesting that the initial plasticity for the cold condition could be adaptive.

220M Deleterious variants and gene by environment interactions in *Zea mays*. *A. Hudson*, J. Ross-Ibarra Department of Plant Sciences and Center for Population Biology, University of California, Davis, CA.

Although selection should act to remove deleterious variants from populations, they are ubiquitous. Multiple forces maintain the presence of deleterious variants in populations, including drift and mutation. Additionally, deleterious variants may be maintained because they are conditionally advantageous depending on the environment. Such seemingly deleterious variants can be identified by studying their effects on fitness in different environmental conditions. Domesticated maize, *Zea mays mays*, has a higher number of putatively deleterious mutations relative to its wild ancestor. *Zea m. mays* experienced a population bottleneck during domestication, which likely increased the proportion of deleterious variants, but deleterious variants are still present in modern day crop lines despite intense selection during breeding and increased purifying selection during the development of inbred lines. *Z. m. mays.* is grown widely across the world and has adapted to a variety of conditions.

We used a multi-parent advanced generation intercross population derived from sixteen *Z. m. mays* inbreds to identify deleterious variants and assess their role in genotype by environment interactions (GxE). The derived lines have been grown in both optimal and stressed (water- and nitrogen-deficient) conditions and have been phenotyped for several traits relevant to fitness. We identified constrained sites in the maize genome using genomic evolutionary rate profiling (GERP). We then identified SNPs within constrained sites present in our panel as putatively deleterious alleles. We used a mixed effects model to measure GxE for putatively deleterious SNPs and compared them to interactions with SNPs in unconstrained sites.

221M The molecular basis of a difference in expression of a prostate-specific gene between humans and chimpanzees, and its relevance for sexual selection in hominids. S.D. Hergenrother, L.G. Loughner, A.M. Colvin Zielen, P. Chovanec, *M.I. Jensen-Seaman* Biological Sciences, Duquesne University, Pittsburgh, PA.

Phenotypic adaptation can occur through changes to protein-coding sequence, or through gene expression, or both. The prostate-specific gene *ACPP* codes for one of the most abundant proteins in human semen, possessing both phosphatase and protease activities. *ACPP* has been suggested to be evolving under positive selection in primates, potentially as a result of sperm competition. Using two-dimensional gel electrophoresis, shotgun LC-MS/MS, and Western blotting, we show the ACPP protein to be expressed several-fold higher in human compared to chimpanzee semen. In order to uncover the molecular basis of the difference in gene expression, we cloned the putative promoter regions of *ACPP* of both species into luciferase reporter constructs. The expression of these constructs in transfected human prostate cells recapitulates the protein level expression in showing a several-fold higher human expression. Taking a phylogenetic approach, transfection of reporter constructs containing the promoters of other hominids suggests that chimpanzees are derived in their reduced expression. We then produced and tested several additional constructs, identifying a specific mutation common to chimpanzees and bonobos sufficient to produce the reduced expression seen in these species. We suggest that since this change occurred specifically in the lineage with the greatest amount of sperm competition among the hominoids, it may likely be adaptive. We also note that the human-chimp ancestor therefore likely had human-like *ACPP* expression. As such, both the coding sequence and the expression of *ACPP* has changed rapidly in chimpanzees while remaining highly conserved in humans. If this gene is evolving in response to shifting pressures caused by changes in mating systems, the observed patterns of

evolution imply that hominins (humans and our extinct bipedal ancestors) have experienced low levels of sperm competition since their divergence from chimpanzees.

222M Experimental natural selection of big bluestem grass ecotypes across the Great Plains' climate gradient. *Loretta C Johnson*[1], Matthew Galliart[1], Nora Bello[2], Jesse Poland[3], Paul St Amand[4], Mary Knapp[4], Brian Maricle[5], Sara Baer[6] 1) Biology Division, Ecological Genomics Institute, Kansas State University, Manhattan, KS; 2) Statistics, Kansas State University, Manhattan, KS; 3) Plant Pathology, Kansas State University, Manhattan, KS; 4) Agronomy, Kansas State University, Manhattan, KS ; 5) Department of Biological Sciences, Fort Hays State University, Hays, KS; 6) Plant Biology and Center for Ecology, Southern Illinois University, Carbondale IL.

A main goal of evolutionary biology is to understand factors that contribute to population genetic divergence, formation of ecotypes, ultimately leading to new species. Habitats are often temporally and spatially variable, causing different selection pressures across gradients, and resulting in genetic divergence among populations. We investigate if ecotypic variation in a dominant US Great Plains grass *Andropogon gerardii* is a result of local adaptation to climate. It has wide geographic distribution across the Great Plains precipitation gradient from western Kansas (dry) to Illinois (wet). Ecotypes (xeric, mesic, wet) were reciprocally planted as ecological communities in Colby, Hays, and Manhattan, KS, and Carbondale, IL. It is crucial to understand bluestem responses to climate for conservation, restoration, and agricultural cattle production. We tested for evidence of local adaptation over 5 years using single ecotype plots (community plots seeded with other prairie plants) and plots with all three ecotypes mixed together. Planting of ecotypes as a community and over multiple years is rarely done, but offers the most realistic test of local adaptation. We identified SNP markers in both plants of known ecotypes and unknown plants from mixed ecotype plots. PCA and population structure show strong genetic differentiation between xeric and wet ecotypes. Outlier analysis identified 64 markers under divergent selection, including GA1 (a gene known to control internode length and height in plants), in which we observe strong ecotype differences between xeric and wet ecotypes. Single ecotype community plots suggest local adaptation to drought with the plants from central KS having higher cover in Hays, KS and plants from Illinois having greater cover in its home site of Carbondale, IL. To analyze the genetic composition of the mixed ecotype community plots, we used the GBS genotype information from plants of known ecotype, then used these data to train a random forest model that allows us to assign unknown individuals from the mixed plots to one of three ecotypes. These multi-year, community plantings show evidence of local adaptation of dry and wet grass ecotypes in reciprocal gardens across the Great Plains. Ultimately these results will provide recommendations to land managers on which climate-adapted source populations of big bluestem is best suited for conservation and restoration planting in future warmer and drier climates.

223M Adaptation to cadmium protects the germline from cadmium-induced mutations. *Nathan Keith*[1], Craig Jackson[1], Stephen Glaholt[1], Kim Young[2], Joseph Shaw[1] 1) School of Public and Environmental Affairs, Indiana University, Bloomington, IN; 2) Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN.

Because the rate of environmental pollution is increasing, it is important to more completely understand how different levels of organismal susceptibility to pollutants (or, “stressors”) correspond to mutational outcomes in the germline. To address this, we sampled over 100 isolates from eleven *Daphnia pulex* lake populations from the industrial mining-devastated region of Sudbury, Ontario, and the ecologically similar, but relatively pristine and non-metal polluted region of Dorset, Ontario. After establishing cultures in clean laboratory for over 50 generations and exposing all isolates to the major environmental stressor by-products of industrial mining (arsenic, cadmium, copper, lead, nickel, zinc), we found the Sudbury populations were adapted to cadmium (Cd), but no other metals. Cd is a potent inducer of oxidative stress, and is both mutagenic and carcinogenic making it an ideal model for studying stress-induced mutation. Here, we provide the rate and spectrum of single nucleotide mutations (SNMs) obtained from both a nonadapted (from Dorset) and an adapted genotype (from Sudbury) via separate mutation-accumulation (MA) experiments. Each experiment was independently propagated in both laboratory controlled conditions (Control), and chronic Cd exposure. We performed deep-coverage, whole-genome sequencing for both experiments representing the mutational processes from >3,500 total generations in the absence of selection.

In the nonadapted experiment, Cd increased the rate of A->G SNMs relative to Control via Nitrosative deamination of adenine (A->inosine->G). Cd also induced a decrease in the rate of C ->G SNMs, the result of Cd inhibition of germline hydroxymethylation (5-hmC). Because 5-hmC sites are subject to elevated rates of germline C->G, Cd’s reduction of germline 5-hmC resulted in a lower rate of C->G. Additionally, Cd increased the rate of G->T, a result of Cd-induced oxidation of guanine. We found the adapted genotype is protected from the Cd-induced mutations observed in the nonadapted genotype. In contrast to the nonadapted experiment, in the adapted experiment Cd decreased the rate of A->G, increased the rate of C->G, and decreased the rate of G->T SNMs. We investigated the influence of Cd on the rate of SNMs in annotated genome regions (e.g introns, exons, etc.). In the nonadapted experiment, relative to Control, Cd induced a shift in the regions where mutations occur by decreasing the overall SNM rate in exons, and increasing the mutation rate in intergenic regions. However, when comparing Control and Cd exposure in the adapted experiment, no difference in SNM rate were observed in

any genome region. Taken together, we give insight into how exposure to an environmental stressor can change the germline SNM rate and spectrum, and give an example where stressor-induced mutations are buffered by adaptation.

224M A pilot study of genetic structure of *Porphyra umbilicalis* KÜTZING in the Gulf of Maine using SNP markers from RNA-Seq. Yuanyu Cao[1], Lindsay Green-Gavrielidis[2], Renee Eriksen[3], *Anita Klein*[1,3] 1) Genetics Program, Univ New Hampshire, Durham, NH; 2) Dept. of Natural Resources Science, University of Rhode Island, Kingston RI, ; 3) Dept. of Biological Sciences, University of New Hampshire, Durham NH.

In the Northeast Atlantic, the red alga *Porphyra umbilicalis* KÜTZING is dioecious and reproduces both sexually and asexually, while in the Northwest Atlantic only asexual reproduction is observed. In this study, transcriptome mining produced putative single nucleotide polymorphisms (SNP) to examine population variation among and within asexual populations of the alga. A computational pipeline was developed that accounts for the specific characteristics of transcriptome dataset, filtered against the available red alga *Chondrus* genome and *P. umbilicalis* EST library to eliminate microbial contamination. For a single population (Schoodic Point, ME), we identified 549 putative SNPs. Primers were designed to amplify a subset of these SNP-containing loci and used in a pilot study of genetic diversity and population structure of seven *P. umbilicalis* populations within the Gulf of Maine. Results of this pilot study show that the genetic diversity is high in the northern populations (Schoodic Point, ME and Lubec, ME) and one southern population (Fort Stark, NH). The genetic differentiation is highest between Reid State Park, ME, , and the rest of populations, and between an estuarine tidal rapid population from Wiscasset, ME and the rest of populations. The open coastal populations at Reid State Park, Schoodic Point and the estuarine tidal rapid population at Wiscasset have novel genotypes.

225M Eukaryotic acquisition of a bacterial operon identified through comprehensive comparative genomics in the yeast subphylum. *Jacek Kominek*[1,2], Drew Doering[1,3], Dana Opulente[1], Xing-Xing Shen[4], Xiaofan Zhou[4], Jeremy DeVirgilio[5], Amanda Hulfachor[3], Cletus Kurtzman[5], Antonis Rokas[4], Chris Hittinger[1,2,3] 1) Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706, USA; 2) DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, WI 53706, USA; 3) Graduate Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706, USA; 4) Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA; 5) Mycotoxin Prevention and Applied Microbiology Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, US Department of Agriculture, Peoria, IL 61604, USA.

Modern sequencing technologies allow investigations into the genetic basis of phenotypic diversity observed in living organisms, not only in individual organisms or genera, but in entire higher-level taxonomic groups. We undertook a massive sequencing and comparative genomics effort to study the genetic basis of metabolic and ecological diversity observed in the entire fungal subphylum of Saccharomycotina, numbering more than 1000 yeast species (Y1000+ Project, www.y1000plus.org). This taxonomic group exhibits great phenotypic diversity at the lifestyle, metabolism, and ecology levels, and includes organisms of very high scientific, medical, and industrial relevance.

One of our most surprising findings was that of a bacterial siderophore biosynthesis operon present in the genomes of a subset of yeast taxa, a trait long thought to be completely absent from the yeast subphylum. We employed sequence analysis and rigorous phylogenetic hypothesis testing and found that it was acquired though a single ancestral horizontal gene transfer event from an Enterobacterial lineage into the ancestor of those species, followed by genomic rearrangements and integration of other yeast genes into the bacterial operon. We used RNA-seq to demonstrate that the expression of siderophore biosynthesis genes does indeed take place, primarily as monocistronic transcripts but with evidence of some bicistronic transcripts. We further employed reporter dye-based and HPLC assays to experimentally demonstrate that biosynthesis of siderophore molecules does occur *in viv*o. Since the yeast species harboring the operon often co-inhabit insect guts alongside various Enterobacteriaceae, our findings suggest an evolutionary scenario where the transferred bacterial operon was maintained by selection in a eukaryotic organism as means of adaptation to this highly-competitive ecological niche. Although gene transfer from bacteria into eukaryotes is not unheard of, the acquisition and maintenance of a functional, multi-gene bacterial operon in any group of eukaryotes have, to our knowledge, never been reported before. The fact that entire metabolic pathways can be transferred as an operon between bacteria and yeasts expands our understanding of the extent and boundaries of gene flow across kingdoms of life.

226M Environmental association in barley landraces: Identifying the genetic basis of low temperature and drought tolerance. *L. Lei*[1], A.M. Poets[1], C. Liu[1], S.R. Wyant[1], P.J. Hoffman[1], C.K. Carter[1], R.M. Trantow[1], X. Li[1], G.J. Muehlbauer[1], F. Katagiri[2], P.L. Morrell[1] 1) Agronomy and Plant Genetics, University of Minnesota, St.Paul, MN; 2) Plant and Microbiology, University of Minnesota, St.Paul, MN.

Barley is cultivated across a broad latitudinal range. In the Northern Hemisphere, the range of cultivation extends from the equator to inside the Arctic Circle (0 - 66˚ N). This broad range of cultivation is especially noteworthy because the wild progenitor species, *Hordeum vulgare* ssp. *spontaneum* occupies a relatively narrow latitudinal range (~30 - 40˚ N), and typically occurs at low elevation (< 1500 m). Changes in growth habit and induction of flowering have been important to cultivation at

higher latitudes and a number of genes involved in flowering time have been isolated. Cold temperature tolerance has been less extensively explored but is amenable to study through environmental association in barley landraces (primitive cultivars). We report an environmental association study involving 784 landraces genotyped with the 9K iSelect genotyping platform. A subset of these lines has been used for exome resequencing. We identify allele frequency outliers across both elevation and latitudinal gradients and make use of mixed model association analysis relative to bioclimatic variables. Using resequencing data, we test for linkage disequilibrium between the SNPs queried in genotyping and SNPs in neighboring loci. In many cases, patterns of linkage disequilibrium are consistent with the causative variant occurring in the immediate vicinity of the queried SNP. While we identify a number of SNPs that potentially contribute to low-temperature tolerance, variants potentially associated with drought tolerance are more abundant in our study.

227M The effects of oxidative stress on *Tigriopus californicus* sex-specific gene expression revealed by single individual RNA-seq. *N. Li*, N. Arief, S. Edmands Department of Biological Sciences, University of Southern California, Los Angeles, CA.

Oxidative stress reflects the imbalance of pro-oxidants and antioxidants. Prolonged oxidative stress would induce cell damage, disease, and aging, and can cause sex-specific effects on phenotypes. *Tigriopus californicus* has recently been proposed as an alternative model system for sex-specific studies due to the absence of sex chromosomes. In this study, we used comparative transcriptomic analysis to assess sex-specific transcriptional responses to oxidative stress in order to elucidate molecular mechanisms of stress tolerance. Male and female *Tigriopus californicus* were maintained in either 1) control conditions, 2) pro-oxidant (H2O2) conditions or 3) decreased antioxidant conditions (reduced carotenoid due to a yeast diet) from birth through six weeks of age. Single individual RNA-seq was then conducted in both males and females by Ligation Mediated RNA sequencing (LM-seq). Twenty-four libraries (4 replicates * (2 treatments + 1 control) * 2 sexes) were sequenced by Illumina Hiseq 4000. Sequence analysis focused on effects of individual variation, gender and stress treatment on gene expression. As the first study to apply single individual RNA-seq in copepods, results will contribute to a better understanding of sex-specific response mechanisms to oxidative stress in the absence of sex chromosomes.

228M Mitochondrial-encoded genes contribute to thermal divergence between *Saccharomyces* species. *X. Li*[1], D. Peris[3], C. Hittinger[3], E. Sia[4], J. Fay[2,4] 1) Molecular Genetics and Genomics Program, Washington University, St. Louis, MO; 2) Department of Genetics, Washington University, St. Louis, MO; 3) Department of Genetics, University of Wisconsin–Madison, Madison, WI; 4) Department of Biology, University of Rochester, Rochester, NY.

The genetic basis of phenotypic evolution is one of the central questions in evolutionary biology. It still remains unresolved how many mutations underlie species' divergence and what is the distribution of their effect sizes. Here, we performed a genome-wide non-complementation screen to systematically dissect the genetic architecture of thermal divergence between two yeast species, *S. cerevisiae* and *S. uvarum*, the former being heat tolerant while the latter being heat sensitive and cold tolerant. Unexpectedly, the screen revealed few nuclear-encoded genes diverged in heat resistance, but a large effect of mitochondrial DNA (mtDNA). Furthermore, we found *S. cerevisiae* or *S. uvarum* mtDNA confers heat- or cold- resistant respiratory growth in S. *cerevisiae* x *S. uvarum* diploid hybrid, respectively. In order to identify the causal genes within mtDNA, we crossed *S. uvarum* to individual mitochondrial gene knockouts of *S. cerevisiae* and found that multiple genes in mtDNA may be involved in the thermal divergence, with a *COX1*-linked region showing the largest effect. Ongoing work aims to characterize the effects of single genes in the linked region by allele replacement in *S. cerevisiae*, using biolistic mitochondrial transformation. Our dissection of the large-effect mtDNA locus highlights the complex genetic architecture underlying species' differences. Given the known role of mito-nuclear incompatibility in speciation of *S. cerevisiae* and *S. uvarum*, our findings also present opportunities for understanding genetic changes at the intersection of speciation and phenotypic evolution.

229M The signature of two centuries of anthropogenic change in genomes of Iberian *Arabidopsis thaliana.* *L. Lopez Perez*[1], S. Marciniak[2], G. Perry[2], J. Lasky[1] 1) Biology, Pennsylvania State University, State College, PA; 2) Anthropology, Pennsylvania State University, State College, PA.

Scientists have observed dramatic phenotypic changes in wild populations in response to anthropogenic and global climate change. Typically, we make inferences about past events based on static genetic patterns in contemporary populations. However, these single time-point observations provide limited information on underlying evolutionary events because multiple processes can generate the same patterns. Therefore, the contributions of evolution versus plasticity to phenotypic changes in wild populations in response to sustained anthropogenic pressures are poorly known. Here we combine historical with current samples to directly track genetic changes of the model species *Arabidopsis thaliana* associated with land use and climate change in the Iberian Peninsula. We sequenced genomes of herbarium samples from the past two centuries and leveraged extensive published data on genetic variation in *Arabidopsis* phenotypes. Ancient samples are known to have low amounts of highly fragmented DNA hampering the genetic data gathering. It is also known that the DNA decay pattern changes over time. Since we have a combination of ancient and historical samples we developed a hybrid wet-lab approach where each sample was individually assessed in order to obtain a consistent number of raw reads among samples and to

maximize the captured proportion of the sequenced genome. Our set of herbarium samples were compared with published recently collected *A. thaliana* genomes of the 1001genomes consortium. Our results indicate that large scale population patterns are retained over time as herbarium and current samples from the same geographic region are more similar among them than with samples from other geographic regions regardless of their collection date. However, despite the genetic similarity within regions across time, fine scale population differences can be appreciated with ancient samples clearly distanced from current ones within geographic regions. We explored several possibilities that might have caused that divergence as for example, changes in allele frequencies that may suggest life history adaptation to disturbed environments or new climates. Furthermore, we assess whether genetic diversity has declined with anthropogenic change.

230M Divergent patterns of copy number variation in natural populations of house mice (*Mus musculus domesticus*) along an environmental gradient. *K.L. Mack*[1], M.A. Ballinger[1], M. Phifer-Rixey[2], M. Nachman[1] 1) Integrative Biology/Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, CA; 2) Biology, Monmouth University, West Long Branch, NJ.

Copy number variants (CNVs) are thought to account for a substantial proportion of total genetic variation and have been associated with phenotypic differences between individuals that can impact fitness. Despite this, there are still few examples of copy number variants that contribute to local adaptation. We apply a read-depth based approach to characterize copy number variation using low-coverage whole genome data in wild-caught individuals of house mice (*Mus musculus domesticus*) collected from five populations along a latitudinal cline in the eastern United States. Consistent with a role for CNVs in local adaptation, we identified two regions where copy number is significantly correlated with latitude. These two regions overlap with 7 genes, whose functions include immunity and cold reception. One of these genes, *Trpm8*, has previously been shown to affect physiological responses to environmental cold in mice, ground squirrels, and hamsters. These results suggest that copy number variation is a significant contributor to genetic variation in North American populations and plays an important role in environmental adaptation.

231M Genetic Architecture of Pollution Resistance in Parallel Populations of Killifish. *J.Thomas. Miller*[1], B. Clark[2], D. Champlin[2], D Nacci[2], A. Whitehead[1] 1) Environmental Toxicology, University of California, Davis, Davis, CA; 2) US EPA Office of Research and Development, Narragansett, RI.

Multiple populations of Atlantic Killifish (*Fundulus heteroclitus)* that reside in heavily polluted habitats along the Atlantic coast of North America have repeatedly and rapidly evolved resistance to highly toxic dioxin-like pollutants (DLCs), which act through the aryl hydrocarbon receptor (AHR) signaling pathway. Quantitative Trait Locus (QTL) mapping has been used to identify some of the genetic markers associated with DLC resistance in a single resistant killifish population. Multi-population genome-wide scans and gene expression profiling also showed that wild DLC-resistant killifish populations have some shared and some unique genomic regions under selection, including some identified in the QTL study as associated with DLC resistance. These studies and others suggest that resistance to the toxic effects of DLCs, such as cardiovascular and other developmental deformities, is extreme but not identical in resistant killifish populations. Here, we extended the QTL approach to all four known resistant killifish populations and employed whole genome scans to compare and contrast genetic regions associated with DLC resistance in these diverse populations. Our results show that genomic regions that associate with DLC resistance in all four populations include regions that contain components of the AHR pathway, *yet some population-specific associations within the AHR pathway were also identified.* These results suggest unique evolutionary trajectories to evolved DLC resistance in wild killifish, and support the development of genotype-to-phenotype maps underpinning adaptations that may permit certain species and populations to persist under rapidly changing environmental conditions.

232M Identification of isoforms and expression profile of Dmrt1 in the painted turtle, Chrysemys picta. *B. Mizoguchi*, N. Valenzuela Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA.

The doublesex and mab-3 related transcription factor 1 (Dmrt1) is a highly conserved gene across phyla, both at the nucleotide and protein sequence level as well as in its function. Dmrt1 encodes a zinc-finger domain, the DM domain, which is involved in sex determination and differentiation in disparate taxa including vertebrates. Non-mammalian vertebrates lack the gene Sry (sex-determining region of the Y chromosome), the male-switch gene in therian mammals. Dmrt1 and homologs work as the male-switch gene in sex determination of birds, the frog Xenopus laevis, the medaka fish (Oryzias latipes) and, recently reported, of two turtle species: Trachemys scripta and Pelodiscus sinensis. Consistently, Dmrt1 is upregulated in males during gonadal development in many species, including the the painted turtle, Chrysemys picta, a species with temperature-dependent sex determination (TSD). Dmrt1 is reported to play different roles during sex determination and differentiation, such as cell fate determination, postnatal Sertoli cells and primordial germ cells maintenance, yet whether these functions are under the control of distinct Dmrt1 transcribed isoforms remains unclear. Dmrt1 isoforms have been characterized in various taxa, including mice, zebrafish, rice field eel and also in the crocodilian Indian mugger, a TSD reptile. However, Dmrt1 isoforms have not yet been identified in turtles.

We address this question using the painted turtle, with the goal of identifying novel Dmrt1 isoforms that may be present during gonadogenesis in C. picta by PCR and profiling their expression by RNA-Seq analysis across 5 embryonic stages , validated qPCR using transcript-specific labeled probes. We discuss the similarities and differences in the expression of the canonical Dmrt1 and isoform transcripts at male- and female-producing temperatures in the painted turtle, and contrast our findings with those in from other vertebrates
This study provides the first insight about Dmrt1 transcriptional diversity in turtles and opens the door for further functional studies of alternative Dmrt1 transcripts uncovered here.

233M Divergence and convergence in Yellowstone monkeyflowers: familiar phenotypes with a new genomic basis. *T. Nelson*[1], F. Finseth[2], P. Breigenzer[1], K. Kolis[1], L. Fishman[1] 1) Division of Biological Sciences, University of Montana, Missoula, MT; 2) W.M. Keck Science Department, The Claremont Colleges, Claremont, CA.

Parallel phenotypic evolution provides a window into the evolutionary constraints generated by genomic architectures and selection regimes. Across its range in western North America, *Mimulus guttatus* (yellow monkeyflower) populations vary in life history strategy, from obligately annual in highly seasonal habitats to entirely perennial in habitats with year-round water availability. Widespread annuals and perennials are differentiated by two diagnostic inversions, which package hundreds of genes into complex alternative strategies. In Yellowstone National Park (YNP), *M. guttatus* is one of a very few species adapted to live near the park's geothermal features, where soils desiccate and can reach 60 ºC in the summer but remain unfrozen and irrigated by snowmelt in winter and spring. Despite close physical proximity to perennial populations — sometimes mere meters — derived annual populations remain differentiated for multiple fitness-related traits, suggesting strong divergent selection. Here, we describe the genetic and genomic basis of the perennial-to-annual transition in YNP. Combining pooled population sequencing with QTL mapping, we find that a highly differentiated region of chromosome 6 (o*ut6*) is a major quantitative trait locus for production of for the life-history-diagnostic trait of rhizome production. We then use whole-genome sequencing of annual and perennials plants to identify the genomic extent, haplotype structure, and relative age of the annual o*ut6* variant. Our results reveal an independent origin of annuality with a simple genetic basis. We interpret these results in the context of range-wide population structure of *M. guttatus* and potentially unique selective regimes in the geothermal habitats of Yellowstone.

234M The adaptive landscapes of Transposable elements in grasses. *Jason Pienaar*[1], Michael R. McKain[1,2], Kristina Zudock[2,3], Taylor AuBuchon[2], Saman Saeidi[2], Rémy Pasquet[5], Daniel J. Layton[2,4], Cassiano A. D. Welker[2,6], Watchara Art-han[7], Paweena Traiperm[7], Christine McAllister[8], Elizabeth A. Kellogg[2] 1) Biological Sciences, The University of Alabama, Tuscaloosa, AL; 2) Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132 ; 3) Department of Computer Science, Washington University, St. Louis, MO; 4) Indiana University, Bloomington, IN ; 5) Institute de Recherche et Development, Cedex, France ; 6) Universidad de Uberlandia, Minas Gerais, Brazil ; 7) Mahidol University, Bangkok, Thailand; 8) Principia College, Elsah, IL.

Multiple adaptive hypotheses for variation in transposable element (TE) abundance are found in the literature, but to date there has not been a large enough data set to test the generality of these hypotheses. Here we assembled a data set of 47, low coverage shotgun sequenced grass accessions, where we estimated genome size and transposable element abundance for each individual. Using voucher information, we gathered data on their sampling location (altitude, longitude, latitude) and bioclimatic variables, which we combined with ploidy estimation and duration (annual vs. herbaceous perennial). Using a well-resolved plastome phylogeny and a suite of phylogenetic comparative methods that use an Ornstein-Uhlenbeck process for modeling stabilizing selection on adaptive landscapes, we tested all combinations of the above-mentioned variables as alternative hypotheses for factors that influence the adaptive landscape for the abundance of each transposable element superfamily. We used Aikake Information Criteria to select amongst the various adaptive evolution hypotheses and neutral hypotheses (modeled as a Brownian motion on the phylogeny). We find that only microsatellites evolve neutrally, whereas the larger satellites are constrained by stabilizing selection around a single, global optimum. Adaptive landscapes of the remaining transposable elements are all affected by residual genome size (genome size minus all transposable elements) in combination with either bioclimatic variable 3 (isothermality) for Copia and Gypsy retroelements, helitrons and MuDR or duration for MuDR and retroelements (where perennials have higher TE abundance). In most cases where bioclimatic variables affect TEs, these variables involve temperature or temperature variation. Most of the evidence supports the hypothesis that temperature stability is the main environmental determinant of TE abundance, where abundance is higher in more stable temperature environments.

235M Rapid evolution and the genomic consequences of selection against interspecific mating. *M.Burford Reiskind*[1], P. Labadie[2], I Bargielowski[3], L Lounibos[3], MH Reiskind[2] 1) Applied Ecology, North Carolina State University, Raleigh, NC; 2) Entomology, North Carlina State University, Raleigh, NC; 3) Florida Medical Entomology Lab, University of Florida, Vero Beach, FL.

While few introduced species become invasive, those that do provide critical information on both the ecological and evolutionary processes that make a particular species an effective invader. *Aedes albopictus* (the Asian tiger mosquito), was

introduced into the naturalized range of *Aedes aegypti* (the yellow fever mosquito) in the United States in the mid-1980s, resulting in the rapid displacement of *Ae. aegypti* in much of the southeastern United States. This displacement was likely due to the superior competitive ability of *Ae. albopictus* as larvae and asymmetric mating interference competition, in which male *Ae. albopictus* mate with and sterilize *Ae. aegypti* females (satyrization). The goal of this study was to examine the genome of *Ae. aegypti* for signatures of selection to resist satyrization in the few remnant populations of *Ae. aegypti* in the SE United States that may respond to the interaction with *Ae. albopictus*. We used a double-digest RAD sequencing method to generate a large number of SNPs used to analyze outlier loci between selected and control of laboratory-reared *Ae. aegypti* females from two populations naive to *Ae. albopictus* (Tucson, AZ and Key West, FL) and with wild caught females that either resisted or mated with *Ae albopictus* males from four populations in Florida. We found significant outlier loci comparing selected and control lines and the resisted and mated wild samples. Overall, outlier loci were unique among the different groups, suggesting there are different genomic mechanisms to overcome strong selection against females mating interspecifically.

236M The genetic basis of gut evolution across trophic levels in Malawi cichlid fish. Kaitlin Coyle, Aldo Carmona-Baez, Patrick Ciccotto, Emily Moore, *Reade Roberts* Biological Sciences, North Carolina State University, Raleigh, NC.

Dietary adaptation is a major source of phenotypic and species diversity across the tree of life, yet there is also surprising convergence of trophic phenotypes among distant taxa. At the level of the gut, intestine length and gut microbiota vary in a surprisingly consistent manner with trophic level: carnivores tend to have short gut lengths, while herbivores generally have longer digestive tracts and more complex microbial communities in order to extract nutrients from less rich or digestible food sources. Despite being fundamental axes of adaptation to diet, the evolved genetic basis of variation in gut length and host-microbiota interactions across trophic levels remains largely unstudied. We are now using the adaptive radiation of Lake Malawi cichlids as a model system to explore the genetic architecture of gut variation. When raised in the lab on a common diet, Malawi cichlid species retain stereotypical differences in gut length and gut microbiota by trophic level, supporting a genetic basis for gut variation. Notably, the recent evolutionary divergence of these species permits production of interspecific crosses, including crosses spanning trophic levels. Here, we describe quantitative trait loci (QTL) mapping in a hybrid cross between *Metriaclima mbenjii* (omnivore, long gut) and *Aulonocara koningsi* (carnivore, short gut), where we identify a few loci of moderate effect on gut length, and displaying interesting interactions with sex. We additionally use comparative genome scans by trophic level to explore potential functional divergence within and outside of QTL intervals. We identify a number of intriguing candidate polymorphisms relating to gastrointestinal evolution, including complex evolution of pancreatic amylase haplotypes that may parallel amylase evolution found in other taxa. Ultimately, we aim to integrate these and other analyses to catalog the genetic changes used by a species to evolve from one trophic level to another.

237T Temperature effects on Mendelian inheritance in intra-species hybrids. Rania Haddad, *Joseph Ross* Department of Biology, California State University, Fresno, Fresno, CA.

To address the question of how genetics might limit the ability of populations to respond to climate change, we sought to ascertain the genomic architecture of adaptation to temperature in *Caenorhabditis briggsae*, a relative of *C. elegans*. The molecular phylogeny of this nematode species reveals two clades of populations that appear to be divided by geographic distribution, specifically by latitude. Using temperature as a proxy for lattitude, we inquired which alleles (temperate or tropical) would be retained in multi-generational inter-clade hybrids produced under both temperature regimes. We genotyped replicate, reciprocal advanced-intercross recombinant inbred lines (AI-RIL) created in both temperate and tropical conditions, Nuclear inheritance patterns revealed deviation from Mendelian expectations as a result of two epistatic effects: GxG (nuclear-mitochondrial) as well as GxGxE (nuclear-mitochondrial-temperature). We interpret these results to suggest that population-specific allele co-evolution in the nuclear and mitochondrial genomes has occurred and that some of these epistatic interactions are temperature-dependent. Such loci might represent a complex genetic basis of climate adaptation. Although such cases of environment-dependent epistasis are rare, particularly in experimentally tractable systems, the potential involvement of mitotype in climate adaptation has a strong precedent. Future efforts involve pinpointing the alleles involved in this phenomenon by fine-mapping as well as exploring the phylogenetic distribution of epistatic alleles by creating additional hybrid lines from other temperate-tropical pairs.

238T Local genetic adaptation in United States *Bos taurus* beef cattle. *T.N. Rowan*[1], H.J. Durbin[1], S.M. Nilson[1], C.M. Seabury[2], R.D. Schnabel[1], J.E. Decker[1] 1) Division of Animal Sciences, University of Missouri, Columbia, MO; 2) Department of Veterinary Pathobiology, Texas A&M University, College Station, TX.

Animals that are poorly adapted to their local environments cost the beef industry more than a billion dollars each year. Technical advances in whole-genome sequencing and array-based genotyping have allowed the beef industry to leverage genomic data as a selection tool. This project aims to categorize genomic variation that contributes to successful animals in particular environments, as well as to better understand gene-by-environment interactions in cattle. Using an evolutionary approach to study artificial selection, we create deliverable results for the beef industry while providing important insights to adaptation biology.

Genotypes were phased with Eagle v2.4 and imputed with IMPUTE2 up to ~850,000 SNPs for over 13,000 Gelbvieh, 12,000 Simmental, and 22,000 Red Angus cattle from 9 distinct regional environments of the United States. We performed FLK and TreeSelect selection scans to identify variants under strong selection that are potentially responsible for detectable regional differences. In addition, we used 30-year mean temperature, precipitation, and elevation values as phenotypes in univariate and multivariate genome-wide association analyses to identify loci associated with these climate variables, and thus under selection. To rule out between-family selection, we use the SmartPCA algorithm and the PC-based test for selection to identify loci responsible for familial, but not regional selection differences in these populations. After accounting for familial selection, 334 loci appear to be under strong region-specific selection in Gelbvieh cattle. Loci from regional selection scans will be used to enhance region-specific genomic predictions. These predictions will provide beef producers and breed associations with a new breeding and selection tool to maximize efficiency specific to different environmental stressors throughout the United States.

239T Analysis of *C. elegans* population dynamics to investigate the ecological significance of aging. *A. Scharf*, H. Jin, J. Mitteldorf, K. Kornfeld Developmental Biology, Washington University in Saint Louis, Saint Louis, MO.

The golden goal in biomedical aging research is to develop strategies to prolong life or retard age-related changes. Therefore, most studies are focused on how individuals age. Such studies in model organisms such as *C. elegans* have identified collections of mutants with altered aging characteristics, including extended lifespan. However, aging could also be viewed as one life history trait that has the potential to influence the number of individuals and the age-structure in a population. This ecological dimension of aging has received less attention, although, it is fundamental to understand aging in the context of system biology.

To analyze the effects of aging on the emergent property of population dynamics, we developed a laboratory ecosystem with the nematode *C. elegans* and its bacterial food source so that we can directly measure population dynamics. Worms are cultured in liquid medium with a controlled nutrient influx of *E. coli* and a predefined predation-rate. Automated worm counters are used to monitor population size. In parallel, we developed a computational simulation that corresponds to the laboratory ecosystem and allows us to systematically analyze the relationship of aging and other life history traits on population dynamics in high throughput. Our long term goal is to understand how individual life history traits influence the emergent properties of populations, such as population dynamics. The presented data will discuss the ecological impact of aging in the context of system biology.

240T Dissecting Natural Variation of Stress Signaling in *Saccharomyces cerevisiae*. *A. Scholes*, T. Stuecker, C. Locke, J. Lewis Biological Sciences, University of Arkansas, Fayetteville, AR.

An individual's physiological response to different environmental conditions often depends on their individual genetics. These so-called "gene-environment interactions" broadly impact many fields of biology, ranging from evolutionary genetics to personalized medicine. We have been exploiting natural variation in the budding yeast *Saccharomyces cerevisiae* to understand the role of gene-environment interactions in yeast stress responses. We have been leveraging extensive natural variation in a unique phenotype called acquired stress resistance, where cells that are pre-treated with a mild, sub-lethal dose of stress survive an otherwise lethal high dose of stress. We have found that while our commonly used lab strain of yeast can acquire further oxidative stress resistance when given a mild stress treatment, this depends on the identity of the mild stressor. For example, mild salt induces high levels of hydrogen peroxide resistance, while mild ethanol stress cannot. This is in contrast to vast majority of wild yeast strain that we tested, which can acquire peroxide resistance when pre-treated with a mild dose of ethanol. Because salt-induced acquired peroxide resistance requires catalase activity in the lab strain, we tested whether catalase was necessary for acquired peroxide resistance in a wild oak strain. Indeed, ethanol-induced acquired peroxide resistance was catalase dependent. Surprisingly, when we performed our "control" experiment with salt-induced acquired peroxide resistance, we found that this was largely catalase independent in the wild oak strain, suggesting that the mechanisms of acquired stress resistance differ depending on strain background. We have now tested over a dozen diverse yeast strains, and find a wide range of catalase dependency for acquired peroxide resistance with either ethanol or salt as the mild pretreatment. We hypothesize that these phenotypic differences are due to gene expression variation. We are currently testing this hypothesis by performing transcriptional profiling on these yeast strains as they respond to either mild salt or ethanol stresses. We will use network analysis to identify potential signaling pathways responsible for the expression divergence, as well as the compensatory oxidative stress defense proteins that are responsible for the different mechanisms of acquired stress resistance across strain backgrounds. This approach highlights the power of using natural variation to uncover novel aspects of signaling networks, which may play a large role in gene-environment interactions.

241T Characterizing a novel biochemical trait, mushroom toxin tolerance, in a *Drosophila tripunctata* population. *C. Scott Chialvo*, L. Griffin, O. Sorrell, O. Fish, L. Reed Biological Sciences, University of Alabama, Tuscaloosa, AL.

The evolution and diversification of complex, novel adaptations have long fascinated evolutionary biologists. While much work has been done on understanding the evolution of novel morphological adaptations, far less is known about the evolution of complex, novel biochemical and metabolic traits. In this study, we examine tolerance to a deadly mushroom-

toxin, α-amanitin, in a population of *Drosophila tripunctata*. We used 10 inbred strains of *D. tripunctata* from Alabama to assess genetic variation in toxin tolerance within a population and quantify the roles of two families of detoxification genes (*Cytochrome P450s* and *Glutathione S-Transferases*) in α-amanitin tolerance. Differential gene expression was examined in a single strain that showed an increase of fitness on a diet containing a natural concentration of the toxin. Our results demonstrate significant genetic variation on a natural α-amanitin concentration. Inhibition of either detoxification gene family did not result in a loss of tolerance. Furthermore, our results highlighted the physiological impacts of the diets on the adult flies and identified genes that may contribute to this novel adaptation. Thus, this study expands our understanding of the genetic architecture and physiological mechanisms that underlie a novel biochemical adaptation and provide testable hypotheses for future studies of the evolution of and mechanisms for this trait.

242T Repeated adaptation of *Mimulus guttatus* to harsh serpentine soils. *J. Selby*, J. Willis Biology, Duke University, Durham, NC.

Organisms that live in habitats characterized by harsh abiotic variables provide particularly vivid examples of how habitat-mediated divergent selection creates biological diversity. Plant adaptation to harsh soils has long fascinated evolutionary biologist interested in investigating the genetic and physiological basis of adaptation because selection in these habitats can be quite strong and, in some cases, the abiotic stress is known and can be manipulated in lab and/or field studies. Furthermore, plants have repeatedly adapted to many of these stressful habitat types, providing opportunities to investigate the degree of parallel trait evolution and whether it is due to parallel changes at the genetic level. In this study we investigate the genetic basis of repeated adaptation to serpentine soils in *Mimulus guttatus* using complementary approaches of QTL mapping and whole-genome resequencing of native populations. Serpentine soils pose a unique set of ecological and physiological challenges to plants such that most plant species are unable to grow in these habitats. However, some species, such as *M. guttatus* can be found in both serpentine and non-serpentine soils. Reciprocal transplant studies show that populations of *M. guttatus* are locally adapted to soil habitat with plants from non-serpentine populations being unable to survive past the juvenile stage when planted in serpentine soils. Initial QTL mapping studies in a single on/off pair identified one major QTL underlying these survival differences. Lab-based hydroponic studies show this QTL to be responsible for tolerance to the low calcium to magnesium levels which define serpentine soils. However, *M. guttatus* can be found growing on patchily distributed serpentine soils across much of its range and these soils can vary substantially in their physiochemical properties. Replicated QTL mapping and population resequencing of eleven different on/off serpentine pairs of *M. guttatus* from throughout California and Oregon reveal widespread sharing of this major QTL however this locus has variable effects in the different populations. Furthermore, we find additional loci contributing to serpentine adaptation that show varying degrees of sharing across the populations. Plant adaptation to harsh soil habitats is a classical study system and the work presented here is one of the first studies to elucidate the loci involved in serpentine adaptation and connect those to fitness effects in the field as well as to understand the distribution of these loci across the landscape.

243T Assessing the genetic diversity of native and non-native *Phragmites* (common reed) in Wisconsin. R.L. Bautzmann, J.D. Pesch, B.J. Murphy, *N.P. Tippery* Department of Biological Sciences, University of Wisconsin - Whitewater, Whitewater, WI.

Invasive species threaten the health of ecosystems worldwide, where they can outcompete and exclude native species. *Phragmites australis* (European common reed) is an invasive plant from Eurasia that has impacted wetlands throughout North America. A closely related native plant, *P. americanus* (American common reed), grows in similar habitats and is in danger of being outcompeted by the more aggressive European common reed. In order to better understand the two *Phragmites* species in Wisconsin, we set out to study their genetic variability and to assess geographical and ecological factors that may influence their respective distributions. We obtained plant material from 21 *P. americanus* populations and 22 *P. australis* populations throughout the state and quantified their genetic diversity using eight previously designed microsatellite markers. We evaluated genotypes for 442 individuals in 20 counties across the state, using the R package *adegenet*. Between two and five alleles were recovered for each marker. All markers were genetically variable within *P. australis*, and five of the eight markers were variable in *P. americanus*. Out of 30 total alleles, all but four were private to one of the two species. In *P. americanus*, two markers showed significant departure from Hardy-Weinberg equilibrium, whereas in *P. australis* seven markers were significantly different from equilibrium. Surprisingly we found greater genetic diversity in the non-native *P. australis* (expected heterozygosity [H_e] 0.102–0.695) than in the native *P. americanus* (H_e 0.000–0.390), a pattern that potentially resulted from multiple anthropogenic introductions. Our results are consistent with natural gene flow among populations of *P. americanus*, and rapid anthropogenic expansion of *P. australis* populations.

244T The predictability of genomic signatures of specialization in a generalist/specialist species pair. *Kim Vertacnik*[1], Scott Geib[2], Hugh Robertson[3], Catherine Linnen[1] 1) University of Kentucky; 2) USDA-ARS: Pacific Basin Agricultural Research Center; 3) University of Illinois at Urbana-Champaign.

Plant-feeding insects are excellent models for evaluating the predictability of evolution because of extensive and replicated host-use variation. Herbivory has evolved independently across multiple insect orders, different insect taxa have repeatedly

colonized the same plant groups, and many independent diet breadth expansions and contractions have evolved.

In taxa that have evolved novel host associations and reduced niche breadths (specialization on a novel host), relaxed purifying selection on ancestral adaptations enables the accumulation of loss-of-function mutations, and ultimately pseudogenization and gene loss. Meanwhile, positive selection on mutations beneficial in the new environment can increase molecular evolution. Together, the changes in selection that accompany ecological specialization predict a genome-wide pattern of widespread gene loss and elevated rates of non-synonymous substitution in genes associated with habitat use.

These predicted patterns have been detected in the chemoreceptor gene families of a handful of insect taxa. But to robustly test the hypothesis that changes in diet breadth are accompanied by predictable genomic signatures, diverse generalist/specialist comparisons are needed. To this end, we evaluated rates of gene loss and molecular evolution in sister species of pine sawflies (*Neodiprion*) that differ in diet breadth. While both species feed on pine (*Pinus*) foliage, the specialist (*N. pinetum*) solely uses white pine (*P. strobus*) and the generalist (*N. lecontei*) is found on ~15 native and introduced pine species (but not white pine).

First, via a series of laboratory preference and performance assays, we confirmed the specialist's improved preference, oviposition performance, and larval feeding performance on white pine. Second, to characterize chemosensory gene families in each species, we assembled both genomes *de novo* and manually annotated the olfactory and gustatory receptor gene families. Third, we determined if rates of pseudogenization and non-synonymous substitutions were higher in the specialist. While the patterns of molecular evolution we observed were consistent with other generalist/specialist comparisons, the magnitude was much less pronounced. Together, results to date suggest that although the genomic consequences of ecological specialization are to some extent predictable, the strength of this signal is variable and likely affected by ecological, demographic, and historical factors.

246T Discerning the historical and genetic relationship between the endosymbiotic bacteria *Wolbachia* and the *Drosophila* germline stem cell gene *bag of marbles*. *M. Wenzel*, C.F. Aquadro Molecular Biology and Genetics, Cornell University, Ithaca, NY.

The *Drosophila* protein coding gene *bag of marbles* (*bam*) plays a key role in early male and female gametogenesis by regulating the differentiation of germline stem cells (GSCs). Although the regulation of GSC gene function is essential for reproduction, *bam* shows strong and episodic bursts of protein sequence diversification in many *Drosophila* lineages. Evolutionary conflicts that drive rapid evolution of reproductive genes often are found between males and females, yet these conflicts are not present in the GSC. Previous work in our lab suggests that one potential evolutionary driver of *bam* is the bacterial endosymbiont *Wolbachia. Wolbachia* is estimated to inhabit 40% of arthropods and can manipulate the reproduction of its host. *Wolbachia* partially rescues the reduced fertility of a *bam* hypomorphic mutant in female *D. melanogaster* as well as a hypomorphic mutant of another GSC gene, *sex lethal* (*sxl*). These results have led us to hypothesize that *Wolbachia* may directly contribute to *bam*'s rapid evolution. One approach we are taking to evaluate this hypothesis is in investigating the historical interaction between *Wolbachia* and *Drosophila. Wolbachia* can undergo lateral gene transfer, inserting pieces of its genome into its host's genome. Over time, these insertions neutrally evolve and such "genomic fossils" can act as a proxy to identify past *Wolbachia* infections. To find such fossils, we are employing a k-mer matching program that has been useful in detecting evidence of past viral and transposable element infections in mammals. Thus far, we have run simulations to mimic a historic insertion event and preliminary results suggest we can detect historic *Wolbachia* insertions up to 7 million years ago, with consideration of the insertion size. Furthermore, we found that a k-mer size of 19 base pairs is sufficient to provide a signature with limited noise. A second approach we are taking to evaluate the role of *Wolbachia* as an evolutionary driver of *bam* is in elucidating the manner in which *Wolbachia* interacts with *bam* hypomorphs. The only existing hypomorph, which is rescued by *Wolbachia* infection in females, is located in a region where Bam binds to a key partner Bgcn. Using polymorphism and divergence data we have identified putative new hypomorphic alleles of *bam.* We are currently using CRISPR/Cas9 to generate these mutants in the w1118 background and test for *Wolbachia* rescue to determine if the *Wolbachia* interaction is specific to certain Bam regions.

247T Genomics of parallel adaptation at two timescales in *Drosophila*. *Li Zhao*[1], David J Begun[2] 1) Laboratory of Evolutionary Genetics and Genomics, Rockefeller University, New York, NY; 2) Department of Evolution and Ecology, University of California Davis, Davis, CA.

Understanding the process of adaptation is a fundamental question in evolutionary biology. Both local adaptation on short timescales and the long-term accumulation of adaptive differences between species have recently been investigated using comparative genomic and population genomic approaches in several species. However, the repeatability of adaptive evolution at the genetic level is largely unknown. We studied this problem by comparing patterns of long and short-term adaptation in *Drosophila melanogaster* to patterns of adaptation on two timescales in a highly diverged congener, *Drosophila hydei*. *D*. *hydei* and *D*. *melanogaster* shared a common ancestor about 50 million years ago and have highly diverged ecologies,

mating systems, and ancestral geographic ranges. While the recent spread of *D. hydei* to a cosmopolitan distribution is not as well understood as that of *D. melanogaster*, the colonization of high temperate regions in North America by *D. hydei* is likely to be recent, similar to the history inferred for *D. melanogaster*. We sequenced and annotated the genome, transcriptomes, studied the population genomics of *D. hydei*, and then compared the population genetic signature of the *repleta* species group and *melanogaster* species group. We found despite the fact that these species diverged from a common ancestor roughly 50 million years ago, the population genomics of latitudinal allele frequency differentiation shows that there is a substantial shared set of genes likely playing a role in the short term adaptive divergence of populations in both species. Analyses of longer-term adaptive protein divergence for the *repleta* and *melanogaster* clades reveal a striking level of parallel adaptation. This parallelism includes not only the specific genes/proteins that exhibit adaptive evolution, but extends even to the magnitudes of the selective effects on interspecific protein differences. Using functional genetic data, we provide novel insight into the selection regimes of recurrent parallel adaptation on the several gene families.

248T Reversal of dominance is the most powerful driver of stable polymorphism in fluctuating environments, but boom-bust cycles and the storage effect are more likely to stabilize many loci of large effect. *J. Bertram*, J. Masel Ecology and Evoltutionary Biology, University of Arizona, Tucson, AZ.

It has long been hypothesized that much of the abundant genetic variation observed in nature is a result of stabilizing selection. This “balance” hypothesis has recently gained new impetus with the discovery in orchard *Drosophila* populations of hundreds of polymorphic loci which exhibit large seasonal cycles in allele frequency, and which appear to be ancient. The apparent pervasiveness of seasonally-balanced polymorphisms in *Drosophila* has resurrected an old theoretical problem: temporally-variable selection has generally been regarded as too weak to be an important stabilizer of polymorphism. This has renewed attention on one notable exception, Gillespie's "concave fitness function" in diploids (which implies reversal of dominance to favor currently advantageous alleles). However, in general the strength of stabilization depends on the fitness effects of the stabilized alleles as well the specific biological mechanisms involved. Moreover, different mechanisms can act in unison. It therefore important that we understand the differences between fluctuation-dependent stabilization mechanisms, as well as their interplay. Here we contrast three broad classes of mechanism that are likely to be widespread (and apply in particular to *Drosophila)*: (1) reversal of dominance / segregation lift; (2) protection from selection (including the classical storage effect as can stem from overlapping generations), and (3) boom-bust demography where genotypes experience more generations of advantage when they are rare. We confirm the classical finding of Gillespie that reversal of dominance is a potent stabilizing force compared to other mechanisms, particularly for small-effect alleles. However, boom-bust demography and protection from selection are shown to also be effective stabilizers for alleles with larger effect sizes. This acts as a sort of filter on small-effect alleles - a filter which is absent for reversal of dominance. We argue that this might explain the recent simulation results of Wittmann et al. (Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift, *PNAS*, 2017), which showed that reversal of dominance could only stabilize a relatively small number of larger-effect polymorphic loci in the presence of diminishing returns epistasis. Counter-intuitively, it is the very potency of reversal of dominance that might be inducing this behavior by stabilizing too many alleles of small effect.

249T On the heterozygosity of an admixed population. *S. M. Boca*[1], L. Huang[2], N. A. Rosenberg[3] 1) Georgetown University Medical Center, Washington, DC; 2) University of Michigan, Ann Arbor, MI; 3) Stanford University, Stanford, CA.

Admixed populations are formed by the amalgamation of multiple source groups. As a result of the fusion of source populations with different allele frequencies or allelic compositions, admixed populations can exhibit high levels of genetic variation, reflecting the varied contributions of their multiple ancestral groups. The goal of the present work is to consider the level of genetic variation in an admixed population as measured by heterozygosity. For a model of an admixed population derived from K source groups, we obtain a relationship between the heterozygosity, H_{adm}, and the proportions of admixture derived from the various source populations. We show that the heterozygosity of the admixed population is at least as great as that of the least heterozygous source population, and that it sometimes exceeds the heterozygosities of all of the source populations. We examine the special case of K=2 source populations in detail, characterizing the maximal heterozygosity across the space of possible admixture coefficients in terms of the heterozygosities of the two source populations and the value of F_{st} between them. We present both theoretical results and applications to simulated data as well as to data from admixed human populations from the Americas, providing information useful for interpreting the properties of genetic variability in admixed populations.

250T The relative importance of hemiplasy and homoplasy in trait evolution. *R. Guerrero*[1,2], M. Hahn[1,2] 1) Dept of Biology, Indiana University, Bloomington, IN; 2) Dept of Computer Science, Indiana University, Bloomington, IN.

Convergent evolution is often inferred when the trait of interest is incongruent with the assumed species tree. Incongruence, however, can also arise from incomplete lineage sorting (i.e., hemiplasy). Hemiplasy is rarely taken into account in studies of convergent evolution, and the relative probability of these two events has not been formally described. Here, we study the relative probabilities of homoplasy and hemiplasy for an incongruent trait. We derive expressions for the probabilities of the two events on a three-taxon phylogeny, and explore parameter space of mutation and times in the tree.

We find that hemiplasy is roughly as likely, or more likely, than homoplasy, for a wide range of conditions. We present an R package to calculate the ratio of these two probabilities (the “hemiplasy risk factor") along branches of a phylogeny of arbitrary length. Such calculations can be applied to any tree in order to identify when and where incongruent traits may be more likely to be due to hemiplasy than homoplasy.

251T Self-incompatibility haplotypes can diversify through sequential mutation and gene conversion. *A. Harkness*, Y. Brandvain, E. Goldberg Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN.

The origin of novel complex regulatory mechanisms is a central problem in evolutionary biology. The RNase-based system of self pollen rejection in the Solanacese is a prime example of this problem. In this case, an S-haplotype consists of a poison RNase expressed in the style tightly linked to multiple SLF variants which act as antidotes expressed in pollen. The SLF that detoxifies a given RNase is not found on that haplotype and thus self pollen is rejected, while pollen from other S-haplotypes can be accepted. Clearly, there is no benefit to a stylar poison without an antidote, and there is similarly no benefit to an SLF antidote without a toxin. How then can a new incompatibility haplotype spread? Here we develop a population genetic model of the evolution of a novel S-haplotype. We show that a novel RNase poison is initially favored because it reduces the fitness of haplotypes lacking the antidote to it, and a new equilibrium of RNases is achieved. We then show the conditions allowing the SLF variant that detoxifies this RNase to spread through the population by gene conversion before S-haplotypes lacking this antidote are lost. Our results show that a parsimonious model including only gene conversion and known incompatibility relations is sufficient to explain the possibility of diversification of S-locus haplotypes, though additional factors probably determine the rate of RNase diversification in nature.

252T Examining the impact of mutation, selection and drift on gene expression evolution. *M. Hill*, P. Wittkopp Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI.

Variation in gene expression is a fundamental driver of phenotypic diversity both within and between species. New mutations with cis- and trans-acting effects on gene expression generate this variation, but it is also shaped by the action of natural selection. The relative influence of mutation, drift, and selection in shaping gene expression evolution remains difficult to disentangle. A realistic neutral model describing the variation in gene expression expected in the absence of selection is a key step toward achieving this goal. Generating such models requires the understanding of the distribution of mutational effects on gene expression and the impact of this expression variation on fitness. A body of recent empirical work in *Saccharomyces cerevisiae* systematically characterizing the relative frequency and expression effects of thousands of *cis*- and *trans*-regulatory mutations altering expression of the TDH3 gene as well as the impact of TDH3 expression variation on fitness in multiple environments provides just this type of data. Here, we use this information to build a realistic model of gene expression evolution, with individuals evolving on real fitness landscapes and mutational effects sampled from empirically derived distributions. By using this model to simulate the evolution of gene expression, we reveal the impact of biases in the mutational spectrum, neutral processes, and natural selection on this quantitative trait.

253T A noval understanding of Linkage Disequilibrium and Past Effective Population Size Estimation. *Tin-Yu Hui*, Austin Burt Silwood Park Campus, Imperial College London, Berkshire, UK, SL5 7PY.

By combining linkage disequilibrium (LD) information from tightly and loosely linked loci it is possible to infer the past effective population size (*Ne*). While similar ideas have been applied to populations of humans and other species, many details behind these studies remain unclear, or the assumptions are often too restrictive. This work revisits some existing methods and more importantly generalises the use of LD signals in historical *Ne* estimation. First, we analytically derive the equations relating LD to historical *Ne* and recombination rates, based on the classical coalescent theory. Second, a likelihood framework is proposed to fit population models to the observed LD curves. Subsequent analyses, including interval estimation and goodness-of-fit, can also be inferred from the likelihood function. Third, the method is verified through simulations and examples from real datasets. The results demonstrate the method’s ability to provide information about *Ne* from a few to hundreds of generations ago, depending on the quality and quantity of the data. Recent changes in *Ne*, such as population expansion, reduction, and bottleneck, are also detectable. These theoretical and empirical advances encourage the future use of the method in historical *Ne* estimation.

254T Updating an efficient pipeline for local ancestry inference. *E.M. Jewett*, G.D. Poznik, K.F. McManus, K. Bryc, J.L. Mountain, A. Auton, 23andMe Research Team 23andMe, Mountain View, CA.

The goal of local ancestry inference is to identify the population of origin at each base in the genome of an individual. At 23andMe, inference is performed by an efficient pipeline known as Ancestry Composition. Ancestry Composition employs support vector machines (SVM) to assign local ancestries to windows along the genome, as well as a hidden Markov model to combine information across window-level SVM calls. The Ancestry Composition pipeline is computationally efficient and has demonstrated low error rates on both simulated and real data. Here, we explore methods for improving the window-level classification component of the pipeline by introducing a method based on the positional Burrows-Wheeler transform

(PBWT). This adaptation of Ancestry Composition improves precision and recall, enables window-free local ancestry estimation, and improves computational efficiency.

255T The rate of fitness valley crossing in rapidly adapting populations. *T.A. Kessinger*, J. Van Cleve Department of Biology, University of Kentucky, Lexington, KY.

Ruggedness is a common feature of fitness landscapes. To traverse such landscapes, populations must often cross regions of lower fitness, known as fitness valleys, in order to find higher fitness peaks: the existence of a valley means that multiple mutations, which can be neutral or even deleterious on their own, are necessary to reach the peak. This is one mechanism by which "irreducibly complex" adaptations can arise. Valleys can be crossed by sequential fixation of mutations, multiple mutations occurring simultaneously, or "tunneling", a process in which deleterious mutant subpopulations persist for short times, affording a time window in which a complex adaptation can appear, establish, and ultimately fix.

Previous studies of valley crossing have focused on populations that exhibit little fitness variation. In such populations, the dynamics of neutral variation, as well as the fate of selected variation at low frequencies, are governed by genetic drift. However, some populations of interest, including many pathogens, exhibit a large fitness variance due to many loci of weak effect. As a result, the fate of effectively neutral mutations is determined by the genetic background on which they appear, and their dynamics are governed not by genetic drift but by genetic draft, a qualitatively different process. Such populations adapt rapidly due to their high fitness variance.

Using a combination of direct simulation and analytical methods, we characterize the rate at which rapidly adapting asexual populations cross fitness valleys. We find that the time to cross a shallow fitness valley can be orders of magnitude longer than in the low fitness variance case. However, the crossing time almost does not depend on either the fitness disadvantage of the intermediate or the advantage associated with the full complex adaptation (the depth of the valley and the height of the peak, respectively). Accordingly, rapidly adapting populations can cross deep valleys at a very high rate compared to populations with low fitness variance. In these populations, the evolution of complex adaptations by jumping across rugged fitness landscapes may be the rule rather than the exception.

256T Simple model of the number of independent originations of recurrent mutant in very large populations. *B.S. Khatri*, A Burt Dept of Life Sciences, Imperial College London, Ascot, Berkshire, GB.

We present a simple analytically tractable model for the number and distribution of independent originations in a sample, due to recurrent mutation of a mutant under positive selection. For very large populations, to a good approximation, once a mutant allele establishes its growth will be nearly deterministic; such a recognition allows simplified calculations of many quantities of interest using semi-deterministic theory. Each mutation is assumed to arise de novo and either establish in the population or not; this gives a non-homogeneous Poisson process of establishing mutants in the population as the frequency of the wild type diminishes over time. The average number of origins is then simply an integral of this rate up to the latest possible time a mutant could establish and then grow deterministically to reach a critical frequency for it to be sampled. This time can be simply calculated by assuming that once a mutant is established it grows logistically, remaining in fixed proportion to the overall mutant population. Our predictions compare very favourably to Wright-Fisher population genetic simulations, and so this provides a computationally inexpensive method to calculate the likelihood of population genetic parameters, such as effective population size and selection coefficient; we discuss application to data from the Anopheles gambiae 1000 Genomes Consortium, where our technique suggests a population size two orders of magnitude larger than implied from nucleotide diversity.

257T Deleterious variation mimics signatures of genomic incompatibility and adaptive introgression. *Bernard Kim*, Christian Huber, Kirk Lohmueller Ecology and Evolutionary Biology, UCLA, Los Angeles, CA.

It is appreciated that population size changes can impact patterns of deleterious variation in natural populations. Then, differences in demography between populations should cause differences in the distribution of deleterious variation between populations. Outside of the study of archaic ancestry in humans, little attention has been paid to how differences in the distribution of deleterious variation between admixing populations might influence levels of introgression. Using simulations with biologically feasible parameters drawn from studies on humans and *Arabidopsis thaliana*, we show that gene flow between populations can temporarily reduce the genetic load of smaller populations. Additionally, when new deleterious mutations have dominance coefficients inversely related to the selection coefficients or are fully recessive, between-population differences in the sites at which deleterious variants exist create heterosis in admixed individuals. These effects can lead to large increase in the frequency of introgressed ancestry, particularly when recombination rates are low and gene density is high. Under certain scenarios, introgressed ancestry can increase from an initial frequency of 5% to 25-75% and fix at many loci, even in the absence of beneficial new mutations. Further, deleterious variation and admixture can generate correlations between the frequency of introgressed ancestry and recombination rate or gene density, which are often considered evidence for genomic incompatibility or adaptive introgression. Because the apparent nature of selection on

introgressed ancestry varies widely due to differences in demographic history, the relationship of selection coefficients to dominance coefficients, variation in recombination rates, and the density of functional elements, it is essential that null models consider these factors before invoking processes such as genomic incompatibility or adaptive introgression to explain unusual patterns of genetic variation.

258T Mutation load from slightly deleterious effects at many loci may be countered by beneficial mutations of larger effect, even when linkage disequilibrium restricts adaptation. *J. Matheson*, J. Bertram, J. Masel University of Arizona, Tucson, AZ.

The average human begins life with upwards of a hundred new mutations not found in their parents, of which an average of two are deleterious. Deleterious mutations with selective effects that are not much larger than 1 / population size N can become fixed, causing the fitness of populations to deteriorate. In the absence of other effects, most species of large vertebrates should have "died 100 times over" due to the accumulation of small-effect deleterious mutations (Kondrashov 1995). Synergistic epistasis between small-effect deleterious mutations could theoretically prevent this accumulation, but recent empirical studies suggest that, while epistasis is rampant among deleterious mutations, they exhibit all sorts of epistasis, which average to a net multiplicative (i.e. non-epistatic) effect. Accumulation of small-effect deleterious mutations could also be balanced by beneficial mutations, but this possibility has been comparatively neglected, primarily because symmetric mutation models, where beneficial and deleterious mutations have the same effect sizes, show limited advantage. But while large-effect deleterious mutations are effectively removed by selection, large-effect beneficial mutations can be retained. This suggests that load might be primarily contributed by a deluge of small-effect deleterious mutations and primarily alleviated by a handful of large-effect beneficial mutations (Whitlock 2000). This asymmetric balance is arguably similar to that implied by drift barrier theory. While a previous asymmetric balance model suggested that load accumulation can be averted (Whitlock 2000), it failed to take into account the effects of rampant linkage disequilibrium, which significantly hamper rapid adaptation. We built a forward-time population genetic simulation using innovative algorithms to appropriately include the effects of recombination in adequately-sized populations and explore the degree to which beneficial mutations are able to balance the accumulation of small-effect deleterious mutations.

259T Inversions help maintain sexually antagonistic balanced polymorphism. *Christopher McAllester*, John Pool Laboratory of Genetics, UW Madsion, Madison, WI.

Inversion polymorphisms and fixed differences are well documented across life, despite the unfit generation of unbalanced gametes from heterozygotes. Inversions may fix in linkage with beneficial alleles, but many inversions maintain intermediate, balanced frequencies, potentially by linking alleles that share conditional benefit, such as in ecotypes of *Mimulus guttatus*. In *Drosophila melanogaster*, paracentric inversions are surprisingly common. We hypothesize that balanced sexually antagonistic selection may be a cause, as many inversions maintain stable intermediate frequencies across broad and diverse African lowland habitats, inconsistent with ecological clines. We wrote a forward population simulation with *D. melanogaster* life history to model inversion evolution in a population under sexually antagonistic selection at infinite loci and with male reproductive skew. The model represents female choice on males by a representative quality score with a normal noise parameter. Alleles carry additive reproductive quality to males and multiplicative survival cost to both sexes. We present results demonstrating balancing selection on alleles with a range of antagonistic effects, the persistence of sets of such alleles only under linkage due to competitive effects, and finally the rise in frequency and stable persistence of inversions that establish such linkage associations between sets of sexually antagonistic alleles. These segregating antagonistic and non-antagonistic haplotypes benefit from linkage to the sex determining locus and so can facilitate neo-sex chromosome formation either by chromosomal fusion or novel sex determination loci. In an empirical extension, we are currently tracking inversion frequency changes between a pooled Zambian parental population and their embryo and aged adult offspring to detect correlation between the inversions and viability and mating fitness. Balancing selection, particularly sexual or ecological antagonistic selection, may be prominent in and relevant to contemporary evolution and local adaptation across species. It is a dynamic that needs more consideration in conservation genetics, adaptation, and human disease.

260T Mother's Curse and Father's Curse as Manifestations of Sexual Conflict. *M.A. Munasinghe*[1], J.A. Ågren[2], A.G. Clark[1,2] 1) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Molecular Biology and Genetics, Cornell University, Ithaca, NY.

The asymmetry in the transmission of sex chromosomes and the mitochondrial genome make them hotspots for sexual conflict. For example, Mother's Curse is a form of sexual conflict that arises due to the strictly maternal inheritance of the mitochondrial genome. Mutations within mitochondrial genes that are beneficial or neutral in females can spread in a population via selection or genetic drift even if they are deleterious in males. This is believed to be a key reason for the greater severity of mitochondrial diseases in males relative to females. The male-biased noxious effects of mitochondrial mutations, however, are expected to be mediated by compensatory nuclear modifier loci that alleviate the fitness cost of Mother's Curse variants in males. While sexual conflict caused by either Mother's Curse or sex chromosomes is well-studied, how interactions between mitochondrial genes and genes on sex chromosomes affect such conflicts is unclear. Here, we use

analytical models and computer simulations to show that mito-sex chromosome interactions may both cause and resolve sexual conflicts. We first demonstrate how the strictly paternally inherited Y chromosome provides a safe harbor for Mother's Curse compensators. Next, this analytical framework allows us to discover a novel kind of sexual conflict, by which Y chromosome-autosome epistasis may result in the spread of male beneficial but female deleterious autosomal mutations within a population. We christen this phenomenon the Father's Curse. We describe the mechanics of Father's Curse and use additional analytical models and computer simulations to show that Y chromosome-autosome interactions generate sexual conflict. Finally, as recent evidence in ZW-systems describe the effect of mitochondrial-nuclear interactions on fitness, we compare our results in XY-systems with ZW-systems to characterize expected deviations from traditional mitochondrial-nuclear theory due to this key difference in sex-determination systems and find that effective compensation of Mother's Curse is depressed in ZW-systems. Taken together, our models provide a novel framework to study how genetic transmission asymmetries may both cause and resolve sexual conflicts.

261T Somatic evolution and the inevitability of aging. *Paul Nelson*, Joanna Masel University of Arizona, Tucson, AZ.

Aging is a failure to maintain soma. It is not obvious, however, that sources of somatic degradation, such as somatic mutation, should necessarily result in systemic accumulation of damaged cells in organisms that are capable of growth and regeneration. Just as competition between organisms can purge deleterious alleles, competition between cells within an organism can purge degraded cells, raising the possibility of an immortal soma. However, not all somatic mutations can be purged through intercellular competition. Some mutations allow a cell to proliferate while disrupting normal cell functioning, bestowing higher cell fitness at a cost to the organism. We present a model that incorporates both somatic changes that increase cell fitness, and somatic changes that decrease cell fitness. Using our model of somatic evolution, we find that the combination of senescent and cancer mutation produces a double bind that results in inevitable somatic degeneration over time. If intercellular competition is limited, senescent cells accumulate. If intercellular competition is prevalent, cancerous cells proliferate. Thus, we show that somatic evolution makes aging universal across all branches of life.

262T The expected distribution of fitness effects in populations rescued from extinction by evolution. *M. Osmond*[1], G. Martin[2], O. Ronce[2], S. Otto[1] 1) University of British Columbia, Vancouver, Canada; 2) Université Montpellier II, Montpellier, France.

Populations exposed to harsh environmental changes (such as drug treatment, herbicides, or pollutants) often go extinct. However, resistance sometimes evolves fast enough to rescue populations and promote persistence. There is now over 20 years of theory investigating how the probability of evolutionary rescue depends on the characteristics of available genotypes, but we still don't know how evolutionary rescue affects the characteristics of genotypes that contribute to persistence. Meanwhile, there is a large literature devoted to understanding the characteristics of genotypes that contribute to adaptation in the absence of demography, the distribution of fitness effects (DFE) along an adaptive walk being of central importance. In this talk I will explain how we have added demography to Fisher's geometric model to predict the DFE of rescued populations, i.e., looking at only those adaptive walks that end in persistence. This approach suggests the genetic pathways by which evolutionary rescue will occur, illustrates the importance of sex and recombination in adaptation to changing environments, and begins a line of inquiry into the genetic signatures of evolutionary rescue that may be identified in the genomes of natural populations.

263T Exact and efficient computation of linkage statistics for inference. *Aaron Ragsdale*, Simon Gravel McGill University, Montreal, QC, CA.

Patterns of genetic polymorphism within and across populations encode information about demographic and evolutionary history. Statistical approaches to infer such histories have often relied on simple summaries of the data, such as the distribution of allele frequencies (called the allele frequency spectrum, or AFS). While the AFS has proven to be a powerful tool for inference, it assumes independence between sites and thus does not take advantage of information contained in patterns of correlation of allele frequencies at linked loci. Statistics on linkage disequilibrium (LD) are themselves informative: two-locus haplotype frequencies are commonly used to infer recombination maps and are sensitive to demography. However, computing expected values for two-locus statistics is numerically challenging beyond the simplest situations. Here we present an exact approach to efficiently compute LD statistics with arbitrary recombination rate, a flexible mutation model, and complex multi-population demography with continuous migration. We integrate this method into an inference approach for reconstructing demography from joint AFS and LD statistics.

Technically our method avoids computing the full two-locus sampling probabilities, instead solving a recursion on a drastically reduced set of statistics that includes moments on the covariance of allele frequencies (D). This approach follows a set of equations studied by Hill and Robertson in the late 1960s, who considered the evolution of low order moments of D. We show that this recursion can be extended to efficiently compute arbitrarily high moments of D. We similarly extend the recursion to multi-population demographies with splits and continuous migration. This allows for the direct computation of expected correlation of LD statistics across populations under realistic and complex demographic scenarios. We couple our

computational approach for linkage statistics with our existing moments-based software for computing the AFS in order to perform likelihood-based demographic inferences from the joint distribution of allele frequencies and LD statistics. Finally, we revisit classical models of human demographic history and compare inferences using joint AFS and LD statistics to those using the AFS alone.

264T The hidden contributions of women to theoretical population genetics. S. Dung[1], A. López[1], E. Lopez Barragan[1], *R.-J. Reyes*[1], R. Thu[1], E. Castellanos[1], F. Catalan[1], E. Huerta Sanchez[2], R. Rohlfs[1] 1) San Francisco State University, CA; 2) University of California, Merced, CA.

In this study, we are interested in investigating the previously unknown possible contributions of women programmers to the foundations of theoretical population and genetics. Pointing out these potential women role models is especially important to reduce the persistent gender gap in computational sciences. Between 1970's through 1990's, programming assistants developed computer simulations or performed computational analysis in computational population genetics and evolutionary biology. We documented author's names, number of authors(classifying binary gender when possible), institutional affiliations, acknowledgments text, and number of acknowledged programmers(classifying binary gender when possible) from all articles in *Theoretical Population Biology* between 1970 to 1990. According to authorship norms at the time, we noticed that many programmers were given credit for their programming work in the acknowledgments section, rather than authorship. We observed a total of 34 APs with classifiable binary gender, 18 of whom were women. In comparison to this, there were 1018 authors with classifiable binary gender, of which only 63 were women. Overall, we noticed a decreased proportion in women programmers over time, which could be due to the masculinization of the computer programming field throughout the years. In addition, we observed that several programmers given repeated acknowledgments for their work in computing, suggesting that these individuals were sought out for their skills in programming. While computer programming has become a male dominant field, women have made significant, hidden contributions to science. By conducting this Acknowledged Programmers Project, we hope to shed light on women's contributions to science as well as promote gender equality in the computer programming and scientific field.

265T The spatial allele frequency spectrum. *D.P. Rice*, J. Novembre Human Genetics, University of Chicago, Chicago, IL.

In populations with geographic structure, the frequency of an allele in a sample of individuals will depend on the location of the sample. Understanding this dependence is important for inferring the history of migration and selection, and for explaining the distribution of phenotypic traits in the population. When population structure is discrete, the spatial patterns may be summarized by the joint allele frequency spectrum: the number of alleles at a particular set of frequencies in each sub-population. However, when a population has continuous structure (i.e. exhibits isolation-by-distance), allele frequencies will vary continuously with geographic location. This fine-scale spatial structure is especially important for rare alleles, which tend to be localized near their place of origin. Here, we analyze the spatial allele frequency spectrum for a lattice model. We calculate the joint moments of the power spectrum of the spatial distribution of neutral and deleterious alleles at equilibrium. We show how this information can be used to identify alleles and polygenic traits that are under selection. We also discuss implications for the design and interpretation of genetic association studies of traits under purifying selection.

266T Estimation of genetic relatedness in admixed populations. *A. Sethuraman* Biological Sciences, California State University San Marcos, San Marcos, CA.

The estimation of pairwise genetic relatedness, and inbreeding coefficients is important to the fields of quantitative genetics, conservation, genome-wide association studies (GWAS), and population genetics. Traditional estimators of genetic relatedness assume an underlying model of population structure. Each individual is assigned to a population, depending on a priori assumptions about geographical location of sampling, proximity, genetic similarity, etc. But often, this population assignment is unknown, and assumptions about assignment can lead to erroneous estimates of genetic relatedness. The *RelateAdmix* model of Moltke and Albrechtsen (2014) extends the admixture model of Pritchard et al. (2000) to using estimates of population assignment to obtain genetic relatedness. I develop a generalized extension to this model, to account for (1) multi-allelic genomic data, (2) including all nine IBD states of Jacquard (1972), and implement a maximum likelihood based estimator of pairwise genetic relatedness in structured populations, part of the software, *InRelate*. Bias and mean squared errors in replicated estimation of genetic relatedness between admixed full sib (FS), half sib (HS), first cousin (FC), parent-offspring (PO) and unrelated (UR) dyads in simulated and empirical data from the HGDP-CEPH panel shows considerably low bias and error while using *InRelate*, compared to the methods of Anderson and Weir (2007), Wang (2011), and other methods implemented in the *COANCESTRY* package. I also propose a bootstrap scheme and a series of Wald Tests to assign relatedness categories to pairs of individuals.

267T *Crossover positions influence the total amount of genetic shuffling*. *Carl Veller*[1,2] 1) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Program for Evolutionary Dynamics, Harvard University, Cambridge, MA.

The total amount of genetic shuffling that occurs in gamete production—the genome-wide recombination rate—is a critical quantity for comparative studies in evolutionary genetics, but it has never been measured directly. Traditional measures simply count the average number of crossovers per meiosis, but a crossover in the middle of a chromosome causes more genetic shuffling than a crossover at the tip, and independent assortment of homologs also causes genetic shuffling. We propose a measure of the genome-wide recombination rate that takes into account these features: $\bar{r}$, the probability that a random pair of loci recombine in the production of a gamete. $\bar{r}$ is easily measured using modern cytological or sequencing data. We provide the first direct measurement of the total amount of genetic shuffling by estimating $\bar{r}$ in male and female humans, using both cytological and single-gamete sequencing data. $\bar{r}$ can be decomposed into separate components deriving from crossing over and independent assortment of chromosomes. Performing this decomposition for humans, we find that about 30 times more genetic shuffling derives from independent assortment than from crossovers. We show that $\bar{r}$ is larger when crossovers are more evenly spaced, with the intriguing implication that crossover interference will tend to increase genetic shuffling. $\bar{r}$ should be used in comparative studies when the total amount of genetic shuffling is a variable of interest.

268T Sexual conflict constrains the evolution of reinforcement. *Alison Wardlaw*, Yaniv Brandvain University of Minnesota, Saint Paul, MN.

The evolution of reinforcement can play an important role in speciation and the maintenance of species boundaries. Theoretical reinforcement models typically investigate the evolution of prezygotic isolating barriers in response to selection against hybridization, assuming both males and females are selected to avoid costly interspecific matings. However, sexual conflict over the hybridization rate can arise if females are selected to prevent producing low fitness hybrid offspring, while males that can fertilize heterospecifics have higher fitness than males that cannot overcome heterospecific isolating barriers (e.g., in the case where pollen grains land on heterospecific pistils or sperm adheres to heterospecific eggs). Thus, when there is a postmating prezygotic barrier, female fitness is maximized by mating with conspecifics, while male fitness is maximized by hybridizing. Interspecific sexual conflict over the hybridization rate could lead to the evolution of mechanisms for male gametes to overcome prezygotic barriers in females of closely related species. We built a population genetic model to study the evolution of a postmating prezygotic isolating barrier in a system with sexual conflict. Our model shows that under these circumstances, the evolution of reinforcement is almost always transient. A postmating prezygotic barrier expressed in females increases in frequency in one species until an allele to overcome that barrier spreads into the other species via higher fitness in males. We compare our 'postmating model' to a system without sexual conflict by instead assuming that the isolating barrier is premating (such that male fitness is maximized by mating with conspecifics). We show that, compared to the 'premating model' and classic results, sexual conflict in the postmating model prevents the evolution of reproductive isolation under conditions in which it would otherwise evolve. The transient evolution of reinforcement under sexual conflict may explain the presence of multiple pollen-pistil incompatibilities between closely related species such as *Zea mays mays* and *Zea mays mexicana*, and more generally, contribute to our understanding of the breakdown of isolating barriers between species.

269T Taking the measure of mutation in the light of molecular dynamics. *G.A. Babbitt* T.H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY.

The proper statistical comparison of GPU accelerated molecular dynamic simulations on homologous proteins allows us to visualize the impact of one of the longest time scale processes in the universe, the molecular evolution over hundreds of millions of years, on one of the shortest time scales, the molecular mechanics occurring over nanoseconds. We have recently published user-friendly free software for comparative protein dynamics (DROIDS 1.2) that works as a backend to Amber16 while incorporating UCSF Chimera visualization to quantify statistically significant differences in the molecular motions of homologous PDB protein structures. DROIDS produces both quantitative plots and color-mapped dynamic structures that demonstrate both angstrom shifts in atom fluctuation on polypeptide backbone (dFLUX) and multiple test corrected p-values that highlight significant differences in dynamics due to both amino replacement and/or epigenetic modification. Here, we present a quantitative survey of the molecular dynamic impacts of mutation under different molecular evolutionary processes including functional conservation, adaptive evolution, gene duplication, and disease malfunction. We demonstrate that many congenital disease mutations significantly alter protein dynamics, in many cases towards a general global destabilization characterized by significantly increased dFLUX over the whole protein. Examples include well-known mutations responsible for cystic fibrosis, congenital cataract, prion protein amyloidosis, Raf kinase melanoma, and Huntington's disease. These disease-related mutations have much larger singular impacts on protein dynamics than most other mutations observed during normal molecular evolution. In contrast, the total dFLUX we observed when comparing functionally conserved orthologs and paralogs is exceedingly small, perhaps suggesting that some interesting biophysical limit(s) might constrain protein evolution (e.g. quantum tunneling). We will also discuss our current efforts to incorporate accelerated machine learning methods to detect and classify molecular dynamic signatures of mutations accumulating under various molecular evolutionary regimes.

Code repository at https://github.com/gbabbitt/DROIDS-1.0---free-software-project-for-comparative-protein-dynamics

View examples at https://www.youtube.com/channel/UCJTBqGq01pBCMDQikn566Kw

Relevant Publication
Babbitt, G.A., J.S. Mortensen, E.E. Coppola, L.E. Adams, and J.S. Liao. In press. DROIDS 1.2 – a GUI-based pipeline for GPU-accelerated comparative protein dynamics. Biophysical Journal (CELL press).

270T Beyond the pale ale: insights into temperature tolerance and carbon source evolution through *Saccharomyces eubayanus*. *E. Baker*[1,2], D. Peris[2,3,4], C. Hittinger[2,3] 1) Microbiology Doctoral Training Program, University of Wisconsin–Madison, Madison, WI USA; 2) Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Genome Center of Wisconsin, Wisconsin Energy Institute, University of Wisconsin–Madison, Madison, WI USA; 3) DOE Great Lakes Bioenergy Research Center, University of Wisconsin–Madison, Madison, WI USA; 4) Department of Food Biotechnology, Institute of Agrochemistry and Food Technology (IATA), CSIC, Valencia Spain.

To successfully colonize new environments organisms must evolve to conditions they could not previously cope with and resources they could not to use. Over time, some strains of *Saccharomyces* yeasts have successfully evolved to thrive in industrial fermentations. Here, they regularly contend with conditions they would have rarely encountered in wild settings. The genetic basis for their success can provide insight into the molecular basis of adaptation. Here we present work detailing the genetics of two traits critical for this transition; temperature tolerance and the evolution of novel sugar utilization. Lager-brewing yeasts, hybrids between *S. cerevisiae* and *S. eubayanus* are the most commercially important of all *Saccharomyces* used in brewing. While it has been known for some time that lager yeasts inherited their cryotolerance from *S. eubayanus* the genetic basis of this trait is poorly understood. In dilution assays with artificial hybrids of *S. cerevisiae* and *S. eubayanus* we found that parental mitochondria DNA (mtDNA) played a significant role in temperature tolerance. *S. eubayanus* mtDNA allowed for greater growth at cold temperatures, while *S. cerevisiae* mtDNA was protective at higher temperatures. In an industrial strain of lager yeast, exchanging the native mtDNA of *S. eubayanus* origin for *S. cerevisiae* mtDNA likewise increased heat tolerance and decreased cold growth.
While industrial hybrids of *S. cerevisiae* and *S. eubayanus* can consume maltotriose, one of the most abundant sugars during brewing, *S. eubayanus* alone is completely unable to. Beers made using *S. eubayanus* therefore retain a substantial amount of sugar. While from a business prospective this represents lost product in terms of quality beer; from an evolutionary prospective it is a potential opportunity for any *S. eubayanus* that *can* consume this previously inaccessible food source. To establish how it might evolve to use maltotriose, *S. eubayanus* was experimentally evolved in media with abundant maltotriose and only enough glucose for minimal growth. After 86 passages a maltotriose utilizing strain was isolated. Backcrossing and tetrad dissection implicated a single causative locus. Bulk segregant analysis was used to map the causative gene and has offered clues to the molecular basis of protein gain of function and provided insight into the molecular tools available to cells to evolve to new substrates.

271T The functional consequences of the rapid evolution of Matrimony, a Drosophila female meiosis-specific inhibitor of Polo kinase. *A. Bonner*[1], R.S. Hawley[1,2] 1) Stowers Institute for Medical Research, Kansas City, MO; 2) Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS.

Meiosis is a defining characteristic of eukaryotes, and it is believed to have evolved only once, over one billion years ago. While the general progression of meiotic events is conserved across multiple diverse organisms, the specific pathways and proteins involved can be highly divergent, even within species from the same genus. Indeed, in numerous studies meiotic genes have shown evidence of rapid evolution, sometimes as a result of positive selection. Here we investigate the rapid evolution of Matrimony (Mtrm), a female meiosis-specific regulator of Polo kinase (Polo) in Drosophila. Mtrm physically interacts with Polo and is required to restrict the activity of Polo during meiosis. However, Mtrm must be destroyed after meiosis to allow for maternally deposited Polo to function properly during the early mitotic divisions in the embryo. Despite Mtrm's critical role in oogenesis, sequence conservation within the genus Drosophila is poor, suggesting the *mtrm* gene is rapidly evolving. To explore the functional significance of this rapid divergence, we expressed Mtrm proteins from 12 different Drosophila species in the *D. melanogaster* female germline. Distantly related Mtrm homologs are still able to both physically interact with *D. melanogaster* Polo and rescue the meiotic defects seen in *mtrm* mutants. However, these distant homologs are not properly degraded after the completion of meiosis, and they continue to inhibit Polo function in the early embryo, resulting in dominant maternal-effect lethality. We also show that the ability of Mtrm to be properly degraded, and thus release free Polo, is partially due to residues or motifs found within Mtrm's least-conserved regions, which shows evidence of positive selection. Overall, this work provides a proof of concept of how a rapidly diverging meiotic gene can be studied in an established model system, allowing us to investigate the functional consequences of its rapid evolution. We hypothesize that, while Mtrm regions critical for its meiotic function are under strong negative selection, changes in its unconserved regions may have been advantageous by affecting the timing or duration of the early embryonic divisions in Drosophila species from different environments.

272T Balancing selection is pervasive in *Drosophila* antimicrobial peptides. *J.R. Chapman*, T. Hill, R.L. Unckless Molecular Biosciences, University of Kansas, Lawrence, KS.

Antimicrobial peptides (AMPs) are key components of the innate immune repertoire of insects, providing direct microbicidal activity against pathogens. The selective forces shaping the evolutionary ecology of these genes are unknown. Balancing selection could act to maintain functional diversity in AMPs if the selective advantage of specific alleles is context dependant, thereby promoting maintenance of multiple alleles over evolutionary timescales. Evidence for balancing selection acting on *Drosophila* AMP genes is equivocal. Genome-wide screens do not tend to identify AMPs as balanced. In contrast, studies of individual AMPs have revealed adaptive maintenance of polymorphism. Here, we specifically set out to test whether AMPs show evidence for balancing selection, and whether this is potentially adaptive, using a two-step approach.

First, we estimated population genetic statistics in four *D. melanogaster* and one *D. mauritiana* population(s) to detect selection acting on AMPs and matched control genes. Overall, we found that the evolution of AMPs is consistent with balancing selection, a pattern that is not observed amongst immune genes as a whole. Second, we tested whether AMP alleles segregating in *D. melanogaster* show functional differences. In-vitro assays revealed clear differences in activity amongst naturally occurring AMP alleles. We then tested whether possession of specific AMP alleles influenced infection outcomes in flies. We found that both bacterial load and survival were correlated with the specific AMP repertoire of the fly. Furthermore, these functional tests revealed that differences could be attributed to single amino acid changes in the peptide.

Taken together, these results strongly suggest that AMP alleles segregating within species are functionally divergent, and that the maintenance of multiple AMP alleles in populations is adaptive. Furthermore, single amino acid changes in AMPs can have profound effects on activity. Maintenance of adaptive polymorphism in AMPs may provide hosts flexibility to respond to a rapidly changing and diverse pathogen fauna, as predicted by host-pathogen co-evolutionary dynamics. Alternately, it may allow hosts to adaptively trade off immune requirements with other life-history and fitness traits.

273T Evolution and natural variation of the major heat-shock protein network in humans. *C. Chavez*, N. Nikolaidis Department of Biological Science, Center for Applied Biotechnology Studies, and Center for Computational and Applied Mathematics, California State University, Fullerton, Fullerton, CA.

Elucidating how genetic variation contributes to human evolution, adaptation, and disease predisposition is an overarching goal in modern human genetics. Molecular chaperones as key orchestrators of cellular homeostasis and adaptation are critical for cell health and disease in humans. Therefore, it is of paramount importance to determine how natural mutations alter the function of molecular chaperones and how these changes affect cell survival. However, the impact of the evolutionary process on their function and diversification remains largely unknown. In this study, we determined the natural variation and identified the modes of evolution of an important component of the human chaperone network, composed by Hsp70s, Hsp40s, and BAGs. Specifically, we collected single nucleotide polymorphisms (SNPs) from the 1000 Genome and Exome Aggregation Database projects and analyzed those mutations using descriptive statistics and sequence evolution tools. The results can be summarized as follows: (I) the vast majority of the SNPs are rare having a frequency below 5% within humans. (II) In 90% of the genes, the number of non-synonymous SNPs (nsSNPs) is significantly higher than the number of synonymous SNPs (sSNPs). (III) Sixty percent of the genes had a significantly higher SNP density than the surrounding genes. (IV) Eighty percent of the genes had significantly higher SNP density within their exonic regions as compared to both intronic and un-translated regions. (V) The majority of genes contained more sSNPs than nsSNPs within known functional regions (domains), while the number of sSNPs is similar between domains and non-domain protein regions. (VI) On average only 15% of the genes contained nsSNPs on an amino acid position of known function. (VII) Only 5% of the nsSNPs were predicted to be deleterious and have a functional outcome. (VIII) Calculations of synonymous and non-synonymous distances revealed the action of strong purifying selection, the intensity of which varied dramatically both between and within the gene families with some exceptions in which the analyses supported positive selection. Collectively, these results reveal that these heat-shock genes are very well conserved and suggest that strong purifying selection due to functional constraints shaped their evolution in humans.

274T Allelism and acetyl-CoA carboxylase gene variants in allohexaploid wild oat. *M.J. Christoffers*, R.P. Sabba North Dakota State University, Fargo, ND.

Wild oat (*Avena fatua* L.) is an annual cool-season weedy grass that is primarily autogamous and reproduces by seed. Native to Eurasia, wild oat has been introduced into temperate agricultural areas worldwide and can cause significant crop losses when uncontrolled. Herbicides used to control wild oat include inhibitors of plastidic acetyl-CoA carboxylase (ACCase), and evolution of herbicide-resistant wild oat populations due to repeated use of ACCase-inhibiting herbicides in crop production is well documented. Three *Acc1* genes for plastidic ACCase have been identified in the allohexaploid ($2n = 6x = 42$) wild oat genome. These independently assorting genes are considered homoeologous and have been designated *Acc1;1*, *Acc1;2*, and

Acc1;3. Mechanisms of herbicide resistance in wild oat include altered plastidic ACCase due to nonsynonymous point mutations within one or more *Acc1* genes. We sequenced *Acc1* homoeologs from the herbicide-resistant wild oat VIR35 and identified a known herbicide resistance mutation, Cys2088Arg, within *Acc1;1*. Herbicide resistance also segregated as a single gene and cosegregated with Cys2088Arg among F_2 progeny derived from a cross with herbicide-susceptible USDA96 wild oat. Among other *Acc1* homoeologs obtained from VIR35, an *Acc1;3* sequence was discovered as expected but no sequence clearly corresponding to *Acc1;2* was found. Rather, a sequence temporarily designated *Acc1;X1* was identified and found to be most similar to the *Acc1;1* gene of USDA96. However, *Acc1;X1* was confirmed as allelic with the *Acc1;2* gene of USDA96, despite sequence dissimilarity. This indicates diverse plastidic ACCase homoeolog composition and may suggest multiple polyploidization in the evolution of allohexaploid wild oat.

275T Convergent regressive evolution of the eye and the identification of eye-specific regulatory elements. R. Partha[1], B.K. Chauhan[2], Z. Ferreira[1], J.D. Robinson[3], K. Lathrop[2], K.K. Nischal[2], M. Chikina[1], *N.L. Clark*[1] 1) Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA; 2) Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) Molecular and Cell Biology, University of California, Berkeley, CA.

Several lineages of subterranean mammals have independently adopted an exclusively underground lifestyle. Extreme examples include the naked mole-rat, blind mole-rat, star-nosed mole, and cape golden mole. In response to their dark environments, these species evolved greatly reduced eyesight in a process known as regressive evolution. Using our new comparative genomic methods we found that hundreds of genes show parallel increases in evolutionary rate specifically in these subterranean species. Most genes were accelerated due to loss of functional constraint, chiefly those involved in eye physiology, such as lens crystallins and photoreceptors, while other genes were accelerated in an apparent adaptation for tunneling and excavating through hard substrate. Regressive evolution proceeded very differently between eye tissues; while lens and retinal genes are highly accelerated, corneal genes remain under constraint, perhaps because they continue to provide a protective outer barrier for the vestigial eye. Moreover, genes important for embryological eye development remain highly conserved, likely because they are important for the development of other tissues. For example, the coding portion of *PAX6*, a key transcription factor in eye, forebrain and pancreas, showed no signs of regression. In contrast, we found that *PAX6's* eye-specific enhancers were evolving at a much faster rate in subterranean species, likely due to relaxed constraint. We then performed a genome-wide screen and identified hundreds of new candidate eye-specific cis-regulatory sequences, which preferentially clustered near confirmed eye development genes. Thus, the results of convergent, regressive evolution provide a powerful means to assign functions to uncharacterized elements in the genome. Upon further examination, mole-accelerated regulatory sequences preferentially lost sequence motifs recognized by transcription factors active in late development and adult tissues, while motifs involved in early development remain relatively conserved. This pattern supports the hypothesis that initial stages of eye development are important cues for proper development of neighboring tissues, and hence early eye stages remain conserved, despite devastating regression at later stages in subterranean mammals. We conclude that eye-related genes and regulatory elements show convergent patterns of loss and retention during repeated instances of regressive evolution, and that regression occurs differentially across tissues, physiological processes and developmental stages. Broadly, this strategy of studying phenotypic convergence in a comparative genomic context is emerging as a powerful approach to characterize functional elements underlying evolutionarily important phenotypes.

276T Evolution of the recombination pathway in mammals. *A.L. Dapper*[1,2], B.A. Payseur[1] 1) Laboratory of Genetics, University of Wisconsin - Madison, Madison, WI; 2) Genomic Sciences Training Program.

Meiotic recombination, the exchange of genetic material between homologous chromosomes during meiosis, is required for successful gametogenesis in most sexually reproducing species. Recombination is also a fundamental evolutionary force, influencing the fate of new mutations and determining the genomic scale over which selection shapes genetic variation. Despite the central importance of recombination, basic questions about its evolution have yet to be addressed. Although many genes that play roles in recombination have been identified, the molecular evolution of most of these genes remains uncharacterized. Using a phylogenetic comparative approach, we measure rates of evolution in 26 recombination pathway genes across 18 mammalian species, from human to platypus. By analyzing a carefully-selected panel of genes involved in key components of recombination – spanning double strand break formation, strand invasion, the crossover/non-crossover decision, and resolution – we generate a comprehensive picture of the evolution of the recombination pathway in mammals. Recombination genes exhibit marked heterogeneity in the rate of protein evolution, both across and within genes. We report signatures of rapid evolution and positive selection that could underlie species differences in recombination rate. These cases include *IHO1*, a gene recently shown to function in double strand break formation, and *RNF212*, a gene involved in the crossover/non-crossover decision associated with recombination rate variation within species. Patterns of co-evolution across genes and domain-specific evolution provide clues into functional divergence of the recombination pathway.

277T A Fitness Carol: The evolutionary trajectory of a protein's past, present, and future. *Drew T. Doering*[1], Chris Todd Hittinger[1,2] 1) Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI USA; 2) DOE Great Lakes Bioenergy Research Center, University of Wisconsin-

Madison, Madison, WI USA.

Predicting the relationship between protein sequence and function remains a major challenge in biology. Homologs that are divergent in sequence often have conserved functions, while rare single nucleotide polymorphisms (SNPs) cause many human diseases. With the emergence of precision medicine enabled by affordable genome sequencing, physicians could make use of data on the fitness effects of specific amino acid changes generated in model organisms when diagnosing their patients. Modern techniques such as deep mutational scanning enable researchers to map the fitness landscape of all possible mutations on a protein's function. Here I present a method that uses amplicon sequencing to characterize the evolutionary fitness landscape of Atx1p (a small, widely-conserved copper chaperone involved in reductive iron uptake) across multiple timescales using the model system *Saccharomyces cerevisiae*. Fitness values of *ATX1* variants are derived from competing a library of variants expressed in *S. cerevisiae* under selective conditions and monitoring changes in variant frequencies by amplicon sequencing. Frequency changes reflect the effect of genetic variants on protein function and ultimately illuminate the constraints present in the fitness landscape. This method enables quantification of the fitness of extant *ATX1* orthologs across the tree of life, reconstructed ancestral *ATX1* sequences, and (via random mutagenesis) potential "future" *ATX1* sequences to determine the factors influencing *ATX1* evolution in the past, present, and future. Growing cells expressing diverse *ATX1* orthologs in media with varying concentrations of copper and iron reveals a general trend of greater sequence divergence resulting in lower fitness. Yeast strains harboring variants of the human *ATX1* (called *ATOX1*) display measurable fitness defects, suggesting that there is functional *ATOX1* variation in people today. Additionally, this work reveals higher fitness costs to mutations in critical domains of the Atx1 protein, such as the copper-binding domain. Studying fitness patterns for past, present, and potential future *ATX1* sequences represents a comprehensive look at the forces underlying molecular evolution of this protein and can easily be adapted to study the evolution of other proteins, such as the roughly 60% of yeast genes that are conserved between yeast and humans.

278T "mtDNA fossils" suggest that our hominine ancestors were engaged in distant interspecies hybridizations. *Z. Fleischmann*[1], K. Popadin[4], K. Gunbin[3], L. Peshkin[2], G. Kraytsberg[2], N. Markuzon[5], S. Annis[1], R. Ackermann[6], K. Khrapko[1] 1) Northeastern University, Boston, MA; 2) Harvard Medical School, Boston MA; 3) Novosibirsk State University, Russia; 4) University of Lausanne, Switzerland; 5) Draper Laboratory, Boston, MA; 6) University of Cape Town, South Africa.

Evidence accumulates to support relatively recent genetic exchanges between homo lineages (e.g. Humans, Neanderthals and Denisovans). Less is known about earlier exchanges involving more divergent lineages. Phylogenic analysis of nuclear pseudogenes of mtDNA (NUMTs) provides insight into such events. Specifically, homologous NUMTs on chromosome 5 in humans, chimps and gorillas share a very long common stem, which by synonymity analysis appear to consist mostly of mitochondrial (*not* pseudogenic) mutations. This allows to estimate the length of this common stem at ~4.5My. We interpret this unusual phylogeny as evidence for the former existence of a hominine species that separated from the mainstream ancestral population before the separation of the Gorilla lineage. This species should have evolved in isolation for 4.5My at which point its mitochondrial genome should have given rise to the NUMT on chromosome 5. Then individuals from this species must have delivered this NUMT to Human, Cimp and Gorilla by re-hybridizing into the mainstream HCG ancestral population. The hypothetical hominine species should have gone extinct thereafter. We show that phylogeny of this NUMT is unlikely a result of large effective population size at the time of integration, incomplete lineage sorting, or branch length artifacts due to the difference in NUMT and mtDNA branch lengths. Reassuringly, interspecies hybridization events of comparable magnitude have been shown in other primates, including baboons and colobines, so such a large genetic distance does not preclude such an event. This observation has important implications for human evolution. Namely, it could provide explanation for extra-phenotypic variation and mosaicism in the fossil record of this epoch. It could also be a possible driver for speciation. They could also serve as possible drivers for speciation.

Reference: Popadin et al. 2016 BioRxiv https://doi.org/10.1101/134502.

279T Diamond in the garbage: A hypothesis on human evolution based on castoffs of the human genome. *Z. Fleischmann*[1], K. Gunbin[2], L. Peshkin[3], K. Popadin[4], S. Annis[1], R. Ackermann[5], K. Khrapko[1] 1) Northeastern University, Boston, MA; 2) Novosibirsk State University; 3) Harvard Medical School; 4) University of Lausanne, Switzerland; 5) University of Cape Town.

Nuclear pseudogenes of mitochondrial DNA (NUMTs) form when fragments of the mitochondrial genome become inserted into the nuclear genome. Hundreds of these (essentially, mutational) events have been catalogued within the human genome alone. In estimating the timing of specific mutation events, characteristics of NUMTs confer substantial advantages over point mutations. With conventional point mutations, the exact time of the event cannot be determined beyond the interval spanning the two nodes of that branch. For instance, a point mutation that falls on the branch after the Human-Chimp separation could have happened anywhere in the past 6 million years. NUMTs, in contrast, are represented by branches on the phylogenetic tree and their branching points can be determined with some precision. The branching point of a pseudogene is not exactly the point of NUMT integration into the nuclear genome. NUMTs contain both mitochondrial and pseudogenic mutations. Mitochondrial mutations conform to the well-established mtDNA clock under the constraints of selection, while the pseudogenic mutations occur in the protective environment of the nucleus but under no selective

pressure. The two mutation types can be distinguished by their synonymity, and the ratio of synonymous, mitochondrial mutations and the nonsynonymous pseudogenic mutations can be used to pinpoint the transition between mitochondrial and nuclear DNA and thus to determine the time of NUMT insertion. Further, the rate at which a collection NUMTs are inserted gives an indication of changes in mutation rate throughout evolutionary time, particularly with respect to important geological or speciation events. We analyzed 18 NUMTs spanning 6 million years of human evolution to ask whether mutation rate is uniform, consistent with gradual adaptation, or concentrated and nonrandom, as in punctuated equilibria. We found a nonrandom distribution of NUMTs clustering at 2.8 Mya, which coincides with the emergence of the genus Homo and major climatic change ca. 2.9-2.5 Mya. We hypothesize this clustering of NUMT insertions suggests a connection with this speciation event and implies an increased mutation rate consistent with the punctuated equilibria model of evolution. NUMTs may be either 'drivers' of evolution if the point of insertion confers a selective advantage or 'riders' if insertion passively coincides with a period of increase genomic flexibility.

References:

Gunbin et al Mitochondrion (2017) https://doi.org/10.1016/j.mito.2016.12.001
Gunbin et al. Data in Brief (2017) https://doi.org/10.1016/j.dib.2017.05.024

280T Evidence for a substantial contribution of damage-induced mutations among human germline mutations. *Ziyue Gao*[1], Priya Moorjani[2], Guy Amster[3], Molly Przeworski[3,4] 1) Department of Genetics & Howard Hughes Medical Institute, Stanford University, Stanford, CA; 2) Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA; 3) Department of Biological Sciences, Columbia University, New York, NY; 4) Department of Systems Biology, Columbia University, New York, NY.

Mutations can arise from errors in DNA replication or from unrepaired DNA lesions that occur prior to DNA replication. We had previously shown that along with replication errors, damage-induced mutations should track cell divisions and accumulate at a higher pace in fast dividing cells, so long as there is DNA repair machinery in operation. With the exception of CpG transitions, however, little is known about the importance of non-replication driven mutations in the germline. In mammals, the distinct development trajectories and epigenetic dynamics of the male and female germlines provide a unique setting to investigate this problem. Here, we exploit human *de novo* mutation data to characterize and contrast the effects of paternal and maternal ages at conception on the rates of different mutation types. Surprisingly, our results point to substantial contributions of mutations of non-replicative origins in maternal and possibly paternal mutations. First, the paternal to maternal mutation ratio is already high at puberty and does not increase much with reproduction age. Second, for many mutation types, fewer than one third of maternal mutations seem to have occurred at birth of the future mother, despite the absence of cell divisions in the female germline after that point. Third, the rate of mutations shared by siblings is consistent with our estimated fraction of maternal mutation at birth, and further supports a small contribution of mutations accumulated during the cell divisions of embryonic development. Lastly, the contrasting time-dependencies of maternal and paternal CpG transitions accord with the distinct methylation profiles of female and male germlines, consistent with chemical modifications of methylcytosines being a major source of CpG transition mutations. Together, these findings provide empirical evidence for a substantial contribution of DNA lesion-induced mutations in human germline, challenging the widespread belief that replication errors are the predominant source of mutations.

281T Recombinant apomixis and genotypic diversity in asexual ferns. *A. Grusz*[1], M. Windham[2], K. Pryer[2] 1) Biology, University of Minnesota Duluth, Duluth, MN; 2) Biology, Duke University, Durham, NC.

Asexual organisms reproduce clonally, without the benefits of segregation, recombination, and gene conversion that result from sex. As such, they are limited in their ability to repair DNA lesions, purge deleterious mutations, or experiment with novel, potentially advantageous, gene combinations. In this study, we show that one asexual reproductive pathway—shared among distantly-related eukaryotic lineages—facilitates sex-like recombinant processes in ferns. Specifically, we demonstrate that apomixis by premeiotic endomitosis yields genetically variable offspring in the apomictic triploid fern *Myriopteris lindheimeri* via exchange between non-sister homologues during meiosis. Using microsatellite DNA markers, we characterize parent-offspring genotypes from across the species' range to reveal that this phenomenon is widespread within populations and across the geographic distribution of this apomictic lineage. Our discovery of recombinant apomixis in *M. lindheimeri* suggests that genetic mixing in apomictic ferns may yield evolutionary consequences resembling long-term selfing in angiosperms. Previous studies of organisms that reproduce via premeiotic endomitosis have primarily focused on parthenogenetic hybrids. In these groups, the prevailing wisdom is that chromosome pairing at metaphase I occurs strictly between sister-homologues, enforcing the maintenance of fixed heterozygosity in the corresponding offspring. The discovery of recombinant asexuality in ferns expands the spectrum of reproductive mode in eukaryotes, representing a paradigm shift with the potential to overturn the assumption, at least in some cases, that asexual organisms are strictly evolutionary dead ends.

282T Non-B DNA affects polymerization speed and error rate in sequencers and living cells. *W. Guiblet*[1], M. Cremona[1], M. Cechova[1], R. Harris[1], I Kejnovska[2], E. Kejnovsky[2], K. Eckert[3], F. Chiaromonte[1,4], K. Makova[1] 1) Penn State

University, University Park, PA; 2) Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic; 3) Penn State University, College of Medicine, Hershey, PA; 4) School of Advanced Studies, Pisa, Italy.

DNA conformation may deviate from the classical B-form in ~13% of the human genome. Non-B DNA regulates many cellular processes; however, its effects on DNA polymerization speed and accuracy — as well as on mutations, many of which arise from polymerase errors — have not been investigated genome-wide. Such an inquiry is critical for understanding neurological diseases and cancer genome instability, which are frequently associated with mutations in non-B DNA. Here we present a comprehensive study of DNA polymerization kinetics in the human genome sequenced with Single-Molecule Real Time (SMRT) technology. We also investigate the genome-wide effects of non-B DNA on germline and somatic mutation rates. We show that polymerization speed differs markedly between non-B and B-DNA: e.g., it strongly decelerates at G-quadruplexes and fluctuates periodically at tandem repeats, including disease-causing ones. We demonstrate that non-B DNA affects sequencing errors and human germline (1,000 Genomes Project, human-orangutan divergence, and re-sequenced trios) and somatic (The Cancer Genome Atlas) mutation rates. Among the most significant effects we observe are an elevation in germline point mutation rates at G-quadruplexes and Z-DNA — both are common non-B DNA types in the genome. Thus, non-B DNA is an important factor contributing to localized (at the scale of tens of nucleotides) variation in mutation rates across the genome and has a large impact on genome evolution.

283T Convergent evolution of phosphatase gene families allows for specialization in phosphate and thiamine starvation in multiple yeast species. John Nahas[2], Christine Iosue[2], Noor Shaik[2], Kathleen Selhorst[2], *Bin He*[1], Dennis Wykoff[2] 1) Department of Biology, University of Iowa, Iowa City, IA; 2) Department of Biology, Villanova University, Villanova, PA.

Gene duplication provides the raw material for evolution, while subsequent changes in gene regulation and function may allow species to adapt to the new environment. Here we report the observation of repeated duplications (and losses) of a histidine phosphatase gene family in a spectrum of *Saccharomycetaceae* yeasts, resulting in multiple phosphatases regulated by different nutrient conditions — thiamine and phosphate starvation. This specialization occurs both at the level of phosphatase substrate specificity as well as transcriptional regulation. Surprisingly, in one of the species, *Candida glabrata*, the complete loss of this phosphatase family was compensated by the co-option of a different histidine phosphatase family. We functionally demonstrate that members of this family evolved novel enzymatic functions for phosphate and thiamine starvation, and specialized their transcriptional regulation to either nutrient condition. The convergent evolution involving two different families of histidine phosphatases, and the resulting specialization for different nutrients, provides an example of how gene duplication, co-option, and transcriptional and functional specialization together allow species to adapt to their environment with existing genetic materials.

284T Pseudogenization of PON1 in Marine Mammals Implies Sensitivity to Organophosphate Pesticides. *J.M. Jamison*[1], W. Meyer[1], R. Richter[2], C. Furlong[2], N. Clark[1] 1) Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA; 2) Departments of Medicine- Division of Medical Genetics and Genome Sciences, University of Washington, Seattle, WA.

Paraoxonase 1 (PON1) is an enzyme important for the oxidation of lipids in the bloodstream, however the flexible nature of the protein allows it to break down other types of molecules, as well. The most notorious of these substrates are organophosphate pesticides, which are neurotoxins related to sarin gas. While all terrestrial mammals studied thus far have a functional copy of PON1, many marine mammals have genetic lesions (such as stop codons or frameshifts) in their *PON1* coding sequence, that are predicted to make the protein nonfunctional. The sequencing of the dugong *PON1* coding sequence via PCR, revealed a shared lesion between it and its closest living relative, the manatee, showing that *PON1* was lost in the common ancestor of the two. By examining available sequences, a similar pattern has been found in cetaceans (dolphins and whales). Through compiling DNA sequences and performing enzyme studies on the blood plasma of pinnipeds (seals, sea lions, and walruses), a more complex pattern of loss has been identified, which suggests the independent loss of *PON1* at least twice within the clade. Additionally, indicators of PON1 nonfunction have been found in semiaquatic species, including a frameshift deletion in sea otters, and a complete lack of enzyme activity in beavers, predicted to be caused by a single amino acid change. The probable lack of a functional PON1 enzyme in these marine species may be disastrous for the animals in question, as they are defenseless against these dangerous and commonly used pesticides.

285T Widespread expression divergence of young duplicate genes in grasses. *X. Jiang*[1], R. Assis[1,2] 1) Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA; 2) Department of Biology, Pennsylvania State University, University Park, PA.

Gene duplication has played a key role in plant domestication and evolution. Yet the mechanisms driving the functional divergence and long-term retention of plant duplicate genes remain unclear. Here we address this question in the Poaceae (grass) family by applying a phylogenetic approach to spatial gene expression data from nine tissues in *Brachypodium*

distachyon, *Oryza sativa japonica* (rice), and *Sorghum bicolor* (sorghum). We discover that approximately 39% of grass genes undergo expression divergence after duplication. Further analysis reveals that expression divergence preferentially occurs in younger copies created by RNA-mediated duplication, and that it often generates novel tissue-specific expression profiles. Last, we find that sequence evolutionary rates are elevated in duplicates with diverged expression profiles, and uncover evidence of recent positive selection in those expressed specifically in reproductive tissues. Together, our findings illustrate that expression divergence is a common outcome of gene duplication in grasses, associated with RNA-mediated duplication, and possibly driven by natural selection on reproductive tissue-specific functions.

286T Current challenges to Neo-Darwinism, molecular phylogenetics, and DNA barcoding: a solution arising from modern molecular data. *Yuri Kartavtsev*[1,2] 1) Lab of Molecular Systematics, National Scientific Center of Marine Biolology, Far Eastern Branch, Russian Academy of Sciences, Vladivostok 690041, Russia; 2) Far Eastern Federal University, Vladivostok 690095, Russia.

Data on a possible impact of genetic introgression on species' genetic divergence, evolutionary fate of taxa, reticulations in phylogenetic trees, and the consistency of the latest molecular sequence data with the main modern paradigm, Neo-Darwinism, are considered in many of papers. In the report, the author will focus on animals, although many ideas suit other phyla too.

The main issues of the report include 4 items: (1) A combination of nDNA and mtDNA markers best suits the hybrid identification and estimation of the genetic introgression (gene flow). (2) The available facts for both nDNA and mtDNA diversity seemingly make the hybrid occurrence among many taxa of animals and plants obvious. Although even in the wide hybrid *Mytilus* spp. zones, for example, the genetic introgression may be quite restricted or asymmetric, thus holding at least the "source" taxon intact. (3) If we admit that a sexually reproducing species in marine and terrestrial realms is introgressed, as it is still evident for many cases, then we should recognize that the orthodox Biological Species Concept (BSC), in terms of complete lack of gene flow among species, is inadequate due to the fact that many wild nature species are not biological species yet. However, sooner or later they definitely become biological species, thus returning BSC validity. This conclusion is supported by the genetic distances, that are minimal intraspecies and increasing with taxa ranks, as for bulk single mtDNA genes, as well as for complete mitogenomes, and for abundant representative nDNA observations (Kartavtsev, 2013; Hedges et al., 2015; Kartavtsev et al., 2016). (4) The recent investigation of fish taxa divergence by the author (Kartavtsev, 2013), using vast BOLD (www.boldsystems.org) data, shows that gene trees investigated for taxa up to the family level are basically monophyletic, and interspecies reticulations are rare.

Main outcomes from 4 above listed points have general impact on the paradigms of evolutionary genetics, molecular systematics & phylogenetics, iBOL (www.barcoding.org) science policy and on the practice of species delimitation in particular. Evidently, the most common successful species delimiting, based on a widely applied barcoding technique, and reliable phylogeny reconstructions based on molecular sequences are possible due to the vast predominance of species origin throughout the geographic speciation mode.

Research is supported by Russian Science Foundation, #14-50-00034, and Russian Foundation for Basic Research, #15-29-02456.

287T Alpha-amanitin resistance in *Drosophila melanogaster*: physiology and gene expression characterizing the trait. Chelsea Mitchell[1], Michael Saul[2], Liang Lei[3], Hairong Wei[3], Roger Yeager[1], Zachary Johnson[1], Stephanie D'Annuzio[1], Kara Vogel[4], *Prajakta Kokate*[1], Thomas Werner[1] 1) Biological Sciences, Michigan Technological University, Houghton, MI1 Department of Biological Sciences, Michigan Technological University, Houghton, Michigan, United States of America; 2) Department of Zoology, University of Wisconsin-Madison, Madison, Wisconsin, United States of America; 3) School of Forest Resources and Environment Science, Michigan Technological University, Houghton, Michigan, United States of America; 4) Department of Experimental and Systems Pharmacology, Washington State University, Spokane, Washington, United States of America.

Insect resistance to toxins greatly affects the ecology of many species as well as our economy. Intriguingly, certain insect species develop resistance to toxins that they have apparently never encountered in their environment. In this study, we investigated the resistance of the fruit fly *Drosophila melanogaster* to a potent mushroom toxin, α-amanitin. Because this fruit fly species does not feed on mushrooms in nature, it may have evolved α-amanitin resistance as cross-resistance to other toxins, e.g., pesticides. The three Asian *D. melanogaster* stocks used in this study, Ama-KTT, Ama-MI, and Ama-KLM, acquired α-amanitin resistance at least five decades ago in their natural habitats in Taiwan, India, and Malaysia, respectively. Here we show that these three stocks have retained their resistance phenotype despite the absence of selective pressure over five decades in the stock center. Furthermore, signs of developmental retardation appeared in response to α-amanitin exposure. In a concentration-dependent manner, pre-adult mortality and larva-to-adult developmental time increased, while adult body size and longevity decreased, the more toxin was added to the larval food. Surprisingly, female fecundity doubled at the second highest sub-lethal α-amanitin concentration in all three stocks. We further performed a whole-genome microarray analysis, using an isochromosome stock derived from the original Ama-KTT line to investigate the mechanisms underlying α-amanitin resistance. We identified three genes of the phase I detoxification gene family *Cyp* (Cytochrome P450), which were

several hundred-fold constitutively up-regulated in the α-amanitin-resistant fly stock, as compared to the susceptible wild type stock Canton S. Because these genes are known to detoxify DDT and other pesticides, we postulate that the resistance to α-amanitin in *D. melanogaster* has evolved as cross-resistance to pesticides, thus underscoring the long-lasting and undesirable effects of pesticides on ecosystems.

288T The molecular evolution of spermatogenesis. *E. Kopania*[1], E. Larson[2], J. Good[1] 1) University of Montana, Missoula, MT; 2) University of Denver, Denver, CO.

Understanding the relative roles of protein coding and gene expression changes underlying adaptation is a fundamental goal in evolutionary biology. A major challenge in this field has been the accurate quantification of protein and gene expression evolution in the context of specific developmental processes. Here, we test how developmental constraint and sexual selection interact to shape evolutionary rate at different molecular levels across spermatogenesis in mice. Mammalian spermatogenesis is well understood, providing a powerful context to understand how developmental constraints interact with other evolutionary forces, such as mutation and positive selection, to shape molecular evolution. Genes are predicted to experience strong pleiotropic constraint during early and sensitive stages of developmental processes. In parallel, many reproductive traits related to sperm competition are predicted to develop during specific spermatogenesis stages. For example, relative rates of sperm production are established early, while sperm head morphology and motility are determined late. The middle meiotic phase is predicted to be highly constrained and less prone to intense sexual selection. Under these conflicting forces, molecular evolution is hypothesized to be somewhat conserved during early spermatogenesis, highly conserved during the intermediate stages, and rapid during late spermatogenesis. Two taxa comparisons have found more rapid divergence in both gene expression and protein coding late in spermatogenesis, and that the strength and direction of correlation between these two levels of evolution depend on chromosome and spermatogenesis stage. We expand on these results to test these predictions in a powerful phylogenetic framework across four species of mice, using cell specific expression data collected across spermatogenesis stages. We use fluorescence activated cell sorting, a powerful technology that allows us to isolate cells at different spermatogenesis stages and study molecular evolution across a developmental process at a much finer scale than previously possible. Preliminary results suggest that protein coding and gene expression are evolving more rapidly in late spermatogenesis, consistent with earlier findings. This broader dataset allows a powerful phylogenetic approach to address questions in molecular evolution. Further work will investigate the underlying causes shaping the relationship between these levels of evolution.

289T Identifying genes associated with longevity in mammals using patterns of convergent evolution. *A. Kowalczyk*, M. Chikina, N. Clark Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA.

The molecular mechanisms behind aging are of considerable interest to both the scientific community and the public at large because of the prevalence of age-related ailments such as cancer, cardiovascular disease, and type II diabetes in human populations. As human life expectancy increases, understanding the molecular origins of senescence is increasingly relevant to inform medical advances that may increase lifespan and improve quality of later life. The wide range of lifespans that exist in mammal species provides an excellent dataset in which to examine the genetic machinery behind aging. By studying how various species have evolved different life expectancies, we can reveal the mechanisms that underlie the aging process.

We used a novel method to scan the genomes of 61 mammal species to find genes whose rates of evolution are associated with longevity phenotypes. Since longevity and body size are strongly correlated in mammal species, we evaluated the genetic mechanisms associated with both traits. We used principal component analysis to identify the first two principal components of longevity and body size and used them as our trait vectors. The first principal component represents the linear relationship between body size and longevity, the “long-lived large-bodied” phenotype, and the second principal component represents the residuals of the linear relationship, the “independently long-lived” phenotype. We then used a linear regression-based approach to measure the correlations between the trait values and the relative evolutionary rates of the genes. Gene evolutionary rates were corrected for the genome-wide evolutionary rate of each species and branches were transformed and weighted using a specialized function to correct for heteroscedasticity.

We found a variety of genes whose evolutionary rates negatively correlated with species longevity, most notably cell cycle control genes, meaning that these genes are under increased evolutionary constraint in long-lived species. We also found enrichments for increased constraint in pathways involving cell cycle control and DNA repair in species with long lifespans. Increased evolutionary constraint of these genes may indicate that control of DNA damage, both at the level of normal DNA function and at the level of DNA replication during cell division, plays an important role in defining species longevity. These results suggest that DNA maintenance capabilities could be major determinants of lifespan.

290T Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts. *D.J. Krause*[1,2], J. Kominek[1,2], D.A. Opulente[1,2], X.X. Shen[3], X. Zhou[3], J. DeVirgilio[4], A.B. Hulfachor[1], C.P. Kurtzman[4], A. Rokas[3], C.T. Hittinger[1,2] 1) Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute, J.F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706; 2) DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706; 3) Department of Biological Sciences, Vanderbilt University, Nashville,

TN 37235; 4) Mycotoxin Prevention and Applied Microbiology Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, US Department of Agriculture, Peoria, IL.

Secondary metabolites are produced by a vast range of diverse species encompassing all domains of life. They play a major role in how organisms interact with their environment and each other. The iron-binding molecule pulcherrimin was described in budding yeasts 100 years ago and chemically characterized in the decades thereafter, but the genes responsible for its production have not been identified. We identified the gene cluster associated with pulcherrimin production using phylogenomic footprinting on more than 100 genomes from across the budding yeast subphylum Saccharomycotina. Using targeted gene replacements in the pulcherrimin producer *Kluyveromyces lactis*, we characterized the four genes of the aptly named *PUL*cherrimin gene cluster: *PUL1* and *PUL2*, required for pulcherrimin biosynthesis; *PUL3*, a putative transporter required for the utilization of pulcherrimin-bound iron; and *PUL4*, a putative transcription factor involved in both processes. This ability of yeasts with a functional *PUL3* gene to utilize extracellular pulcherrimin-iron complexes implied that pulcherrimin is a yeast siderophore, a chelator that binds iron outside the cell for subsequent uptake. Surprisingly, we identified homologs of the *PUL3* and *PUL4* genes in multiple yeast genera that lacked *PUL1* and *PUL2* genes and could not make pulcherrimin. We confirmed the roles of *PUL3* and *PUL4* in pulcherrimin utilization by deleting them in the model yeast *Saccharomyces cerevisiae*. *PUL3 (YNR062C)* and *PUL4 (YNR063W)* were previously uncharacterized and unnamed in *S. cerevisiae*, raising the possibility that other uncharacterized *S. cerevisiae* genes may be linked to secondary metabolism without having clear roles during standard laboratory growth. The phylogenetic distribution of this gene cluster suggested that pulcherrimin biosynthesis and utilization were ancestral to all budding yeasts, but the genes were lost in many lineages, leading to non-producers that lost the entire cluster and cheaters that lost only *PUL1* and *PUL2*. Ongoing work is currently focused on pulcherrimin trait variation within species, including quantitative variation in pulcherrimin production and the frequency of *PUL* gene loss.

291T Different evolutionary dynamics revealed by functional SNP classes in global chicken groups. *D.K. Malomane*[1], C. Reimer[1], S. Weigend[2], A. Weigend[2], H. Simianer[1] 1) Center for Integrated Breeding Research, Department of Animal Sciences, University of Goettingen, Goettingen, Germany; 2) Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, Neustadt, Germany.

Evolutionary forces such as selection and genetic drift have played a huge role in the chicken diversification. Such forces led to genomic alterations (i.e. fixation of favorable alleles) accountable for the many features present in chickens and differentiation from their ancestral state. Consequently, functional single nucleotide polymorphism (SNP) classes in various breeds might have been subjected to different evolutionary forces which resulted in the various phenotypes. The differences in allele frequencies (AF) between synonymous and missense SNPs can be used to study if amino acid changes were neutral or reflected selection. Due to the functional implication of missense SNPs, a higher rate of fixation or loss can indicate possible selection, be it favorable or not. We used data consisting of 18 breed types (3231 individuals) from the SYNBREED chicken diversity panel, genotyped with the Affymetrix 600K SNP array. One group represented the wild type. The SNPs were classified into 6 functional classes. We estimated and compared AF spectra and observed heterozygosity (H_o) across the various SNP classes. We found that some breed types exhibited deviating patterns from the wild type. There were 3 patterns for the differences in AF distribution between missense and synonymous SNPs, i) for the wild type and many of the breed types, there was an over-representation of missense SNPs compared to synonymous ones towards the rare and common AF bins, indicating an erosion of some (presumably the undesired) variants and fixation of favorable ones. ii) A more rapid fixation of missense SNPs than synonymous SNPs which was observed in some breeds predominately the crested types, showing subjection to directional selection. iii) A lack of difference between the 2 SNP classes which was only observed in the egg laying types, such similarities in the 2 classes may be explained by linkage drag. There was an excessively higher rate of fixation and loss of missense SNPs in the white layer type than in the rest of studied types, indicating an intensive selection pressure. The difference in the level of H_o in the considered SNP classes were systematic across the studied breed types whereby the missense sites exhibited significantly lower levels of H_o than the rest of the SNP classes, except for the white layer type which had no difference in H_o between missense and synonymous sites. The results demonstrate that chicken breed types have experienced different evolutionary forces at the various SNP classes.

292T Traversing the fitness landscape of lung adenocarcinoma *in vivo* using tumor barcoding and CRISPR/Cas9-mediated genome editing. *Christopher McFarland*[1], Zoe Rogers[2], Ian Winters[2], Dmitri Petrov[1], Monte Winslow[2] 1) Department of Biology, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA.

Cancers exhibit a highly diverse landscape of somatic alterations. However, the effects of the vast majority of these alterations on tumor evolution, both alone and in combination, remains unknown. We developed a method that integrates DNA barcoding with CRISPR/Cas9-mediated genome editing to interrogate multiple tumor genotypes simultaneously in autochthonous mouse models of human cancer. First, via ultra-deep sequencing of barcoded tumors, we track the size of hundreds of early tumors within a single mouse. Isogenic tumors within the same mouse quickly diverge in size. Different archetypal cancer genotypes exhibiting categorically-different tumor size distributions. Thus, we proposed alternative Markov

models of tumor evolution to explain these different size distributions, which we then tested by tracking the growth of hundreds of tumor trajectories over time using our quantitative barcoding approach. Next, we paired our barcoding technology with CRISPR/Cas9-mediated genome editing to measure the fitness of thirty-one of the most frequent human lung adenocarcinoma genotypes. We observe that most gene losses confer context-dependent growth, i.e. they are adaptive *only* in the presence (or absence) of other genetic events. While our fitness measurements correlated well with mutational co-occurrence frequencies in human lung adenocarcinoma, most genomic co-occurrence patterns were still deemed statistically insignificant by existing statistical algorithms, suggesting that statistical approaches for detecting genetic interactions in human cancers are still underpowered and understate the degree of context-dependency in tumor evolution. Our new tumor-barcoding and CRISPR/Cas9-mediated genome-editing approach measures tumor growth with unprecedented parallelism and precision, reveals tremendous variability in tumor growth that informs the mode of cancer evolution, and identifies a rugged fitness landscape of tumor evolution that should help direct and personalize cancer therapeutics.

293T Single locus experimental evolution of ancestral steroid receptor proteins. *B. PH. Metzger*, J. Thornton Ecology and Evolution, University of Chicago, Chicago, IL.

The genetic paths and biochemical mechanisms underlying historical changes in molecular function have begun to be identified. But why were these the paths and mechanisms used during evolution? Were they the only paths and mechanisms accessible to evolution, or were they the ones requiring the fewest number of changes? Perhaps many alternative paths and mechanisms were equally likely and evolution simply choose one at random. Could a different function have evolved if evolution were repeated? Answering these questions requires identifying not only what happened historically, but also what else was evolutionarily possible. Here we describe recent work combining ancestral protein reconstruction and single locus experimental evolution to address these questions. We focus on the steroid hormone receptors, a family of DNA binding proteins that have undergone a functional shift in DNA specificity following a historical duplication. We use the phage assisted continuous evolution (PACE) system to rapidly evolve ancestral steroid receptors towards both their historically derived DNA specificity as well as novel DNA specificities not observed in nature. By performing replicate evolution experiments, we can identify alternative paths and mechanism that were historically available to evolution, estimate the repeatability of the evolutionary process, and determine the probability that evolution would have followed the historical path. Combining single locus experimental evolution with ancestral protein reconstruction thus allows us to 'replay the tape of life' and place historical changes in function within the broader context in which they evolved.

294T The tempo and mode of genome evolution across the budding yeast subphylum. *D. Opulente*[1,2], X. Shen[3], J. Kominek[1,2], X. Zhou[3,4,5], J. Steenwyk[3], K. Buh[1], M. Haase[1], J. Wisecaver[3,6], M. Wang[3], Q. Langdon[1], J. DeVirgilio[7], A. Hulfachor[1], M. Groenewald[8], C. Kurtzman[7], A. Rokas[3], C. Hittinger[1,2] 1) Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706, USA; 2) DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA; 3) Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA; 4) Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou, P.R. China; 5) Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Department of Plant Pathology, South China Agricultural University, Guangzhou, P.R. China; 6) Department of Biochemistry, Purdue University, West Lafayette, IN 47907-1153, USA; 7) Mycotoxin Prevention and Applied Microbiology Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, Peoria, Illinois 61604; 8) Westerdijk Fungal Biodiversity Institute, 3584 Utrecht, The Netherlands.

The subphylum Saccharomycotina, the budding yeasts, comprises most known yeast species. This diverse group includes the eukaryotic model system *Saccharomyces cerevisiae* (baker's yeast), the common human commensal and opportunistic pathogen *Candida albicans*, and over 1000 other known species, with more continuing to be discovered. Yeasts exhibit remarkably diverse heterotrophic metabolisms, allowing them to inhabit every continent and every major aquatic and terrestrial biome. Furthermore, they display remarkable genetic diversity. Comparative genomics has revealed genetic diversity that is on par with the green plant and bilaterian animal lineages. However, many of these analyses are broad and focus on taxa that are relevant to industry and biomedical research, and therefore cannot identify major evolutionary trends across the subphylum. To address this, we sampled the genomes of 332 yeast species, including 196 newly-sequenced genomes and 78 of 92 known genera from the 12 major clades within the subphylum. We analyzed genome-wide variation to construct a robust phylogeny, establish a timeline of budding yeast diversification, examine the influence of horizontal gene transfers (HGTs) on yeast genome evolution, and reconstruct the evolution of 45 metabolic traits. Our analyses reveal multiple lineage-variable HGT events and an overall trend of metabolic specialization that is driven by repeated loss of metabolic genes and traits. These results provide a clearer picture of the evolution of this eukaryotic group.

295T Deep mutational scanning along an evolutionary trajectory. *Y. Park*, B. Metzger, J Thornton Ecology & Evolution, University of Chicago, Chicago, IL.

Recent research has traced the genotype-phenotype space around individual proteins, elucidating the biophysical basis of epistasis and revealing the interplay of neutral and function-switching substitutions in shaping evolutionary trajectories.

However, it remains unknown how the local phenotype space changes as a protein evolves along an evolutionary trajectory. How does the distribution of substitution effects differ between an ancestral protein and an extant protein with the same function? Does the neutral network change in size and topology as the protein accumulates substitutions? Are there epistatic interactions that persist through long evolutionary trajectories, and what role do they play in protein structure and function? To address these questions, we developed a deep mutational scanning method to functionally characterize all single and double mutants of the DNA-binding domain of steroid hormone receptors. We applied this method to two ancestral proteins flanking an evolutionary interval where the DNA-specificity changed. The substitutions in this interval can be divided into neutral and specificity-switching substitutions. To separately examine how these neutral and function-switching substitutions altered the phenotype space, we are also characterizing intermediates within this interval.

296T Developing robust evolutionary-rates-based methods for detecting convergent genomic changes underlying phenotypic adaptations. *Raghavendran Partha*, Nathan Clark, Maria Chikina Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA.

Identifying genomic elements underlying phenotypic adaptations is one of the outstanding problems in evolutionary biology. Comparative analyses learning from patterns of convergent evolution of traits are gaining momentum in accurately detecting genomic elements associated with the trait. We previously developed an evolutionary-rates-based method that successfully detected convergent rate acceleration in genomic elements associated with two phenotypic adaptations, namely the independent subterranean and marine transitions of terrestrial mammalian lineages. We hereby describe key improvements to our method that make it more robust and generally applicable.

Our method uses linear regression to compute the relative evolutionary rates of genomic elements on the branches of a phylogenetic tree. We remove any genome-wide effects on branch lengths of individual trees by correcting for the average branch lengths, leaving only gene-specific relative rates. In doing so, we are faced with one of the classical problems in regression analysis – heteroscedasticity. We observe that relative rates on branches that are longer on average show higher variance - a problem that adversely affects the confidence with which we can make inferences about rate shifts. We correct this heteroscedasticity in the branch lengths using a two-step approach of data transformation followed by weighted regression. We test multiple variants of the method, each using a different transformation scheme, to reduce the degree of heteroscedasticity. Subsequent to data transformation, we do an initial round of naive regression to empirically fit a function describing the relationship between the variance of residuals and average branch lengths. We then compute weights for every branch based on this function to perform a weighted regression that produces residuals showing near constant variance.

To identify the best performing variant, we additionally compare their power to detect convergent rate shifts in three independent datasets – i. in eye-specific genes showing convergent regression in subterranean mammals, ii. in meiosis genes of pathogenic yeast species showing relaxation of constraint on sequence evolution, and iii. in simulated gene trees where we can explicitly control foreground branches showing rate shifts.

Overall, we present an important extension to our evolutionary-rates-based method that performs more robustly and consistently at detecting convergent shifts in rates. We plan to release our methods as an R package, to facilitate investigation of candidate genomic regions associated with a phenotypic trait of interest. We have additionally built an interactive visualization tool using Rshiny that will be hosted online, with the goal of making our methods more accessible to the broader scientific community.

297T Natural Selection Has a Broad Impact on Codon Usage Biases across Eukaryota. *Z. Peng*, H. Zaher, Y. Ben-Shahar Department of Biology, Washington University in St. Louis, Saint Louis, MO.

Genomic DNA data from diverse taxa indicate that frequencies of synonymous codon usage often deviate from expected random usage patterns. While the specific patterns of codon usage bias can vary across species, this phenomenon seems to be a universal feature of protein-coding DNA sequences. However, the molecular forces that shape codon usage bias at the evolutionary timescale remain poorly understood. Nevertheless, since synonymous mutations do not affect the primary protein sequence, they are generally considered neutral at the evolutionary timescale. Yet, anecdotal reports indicate that synonymous mutations can have phenotypic consequences. By combining novel theoretical and experimental approaches, we show that in contrast to the current dogma, natural selection on gene-specific codon usage bias is common across Eukaryota. Furthermore, by using bioinformatic and experimental approaches, we demonstrate that in *Drosophila melanogaster*, a specific combination of rare codons contributes to spatial and sex-related regulation of protein expression. Together, our data indicate that patterns of gene-specific codon usage represent a broadly functional genomic feature that is shaped by natural selection, and plays an important role in determining the molecular landscape across time and space. Therefore, the widely accepted dogma that synonymous mutations are generally functionally and evolutionarily neutral should be reevaluated.

298T Evolutionary origin of multimeric assembly in Hemoglobin. *Arvind Pillai*[1], Georg Hochberg[1], Yang Liu[2], Arthur Laganowsky[2], Jay Storz[3], Joseph Thornton[1], Antony Dean[3] 1) Ecology and Evolution, University of Chicago, Chicago; 2) Department of Chemistry, Texas A&M University, College Station; 3) School of Biological Sciences, University of Nebraska-Lincoln, Lincoln.

Heteromeric protein complexes perform a wide range of biological functions within the cell, from transcription initiation to ion transport. In addition to their biological importance, such "molecular machines" exemplify questions about the evolution of biological complexity. A major goal for evolutionary biochemistry is therefore to explain the evolutionary mechanisms by which multisubunit protein complexes and the structural interfaces that mediate their assembly came to be.

We use vertebrate Hemoglobin (Hb) as a system to study the origin and evolution of multimerization in a gene family. Hb is a heterotetramer of two α and two β subunits, which are paralogs. It binds four oxygen molecules cooperatively and transports them in blood to peripheral tissues. Two large interfaces, both between α and β subunits, primarily mediate formation of the complex. Most other members of the globin family are nonallosteric monomers, suggesting that Hb's tetramerization, cooperativity, and allostery are derived. The evolutionary trajectory from presumably ancient monomer to the more complex Hb architecture and function is unknown.

We dissected the origin of the Hb complex using ancestral protein reconstruction in combination with biophysical and biochemical techniques, including SEC-MALS, Native Mass Spectrometry and oxygen-binding affinity assays. We show that Hb evolved from a non-cooperative, homodimeric ancestor over 450 million years ago. We identify a subset of essential historical mutations that were required for the evolutionary transition to a more complex subunit stoichiometry. Through mutagenesis, we uncover the biochemical and structural mechanisms by which a new protein-protein interface was created de novo post-duplication, and also demonstrate that one interface existed in homomeric form prior to gene-duplication.

This work reveals in clear mechanistic detail how a biologically crucial heteromeric complex evolved from a simpler ancestral form.

299T Immune genes are hotspots of convergent positive selection in birds. *A.J. Shultz*[1,2,3], T.B. Sackton[1] 1) Informatics Group, Faculty of Arts and Sciences, Harvard University, Cambridge, MA; 2) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 3) Museum of Comparative Zoology, Harvard University, Cambridge, MA.

Evolutionary theory predicts that genes encoding proteins involved in host-pathogen interactions will be subject to arms race dynamics, in which repeated adaptations and counter-adaptations will fix in both hosts and pathogens. In genomic sequence comparisons, this dynamic leads to a pattern of diversifying selection, detectable by an excess of amino acid substitutions fixed by natural selection (positive selection). Previous work in mammals and other groups has provided evidence that this process is important in the evolution of immune system proteins and proteins that interact with viruses. In this talk, we infer the history of selection on genes using comparative genomic methods, present new work on the evolution of the innate immune system in birds, and contrast patterns of selection seen in these species with mammals. Using alignments of orthologous protein-coding genes from 39 species of birds, we estimated parameters related to models of positive selection for 11,000 genes conserved across birds. We show that diversifying selection is particularly important in the evolution of pathways involved in the response to Influenza A and other pathogens, particularly viruses. By comparing these results to previous work in mammals, we then show that genes under selection in birds are enriched for genes that are also under selection in mammals. Finally, we demonstrate that sets of genes known to interact with viruses, bacteria, or Plasmodium are more likely to be under selection in both birds and mammals. These findings suggest that pathogens consistently target the same genes, and that these genes are hotspots of host-pathogen conflict over deep evolutionary time.

300T Increased rates of molecular adaptation, but not divergence, in proteins of dimorphic sperm from Monarch butterflies (*Danaus plexippus*). Andrew Mongue[1], Emma Whittington[2], Desiree Forsythe[1], Tim Karr[1], Steve Dorus[2], *James Walters*[1] 1) Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS; 2) Department of Biology, Syracuse University, Syracuse, NY.

Reproductive traits are often functionally enigmatic and evolutionarily distinctive. One striking example of this is dimorphic sperm in Lepidoptera (moths and butterflies), where males produce fertilizing-competent "eusperm" as well as a vastly greater quantity of "parasperm" which lack a nucleus and nuclear DNA. Neither the function, molecular composition, nor selective pressures are known for lepidopteran parasperm. We employed mass-spectrometry proteomics to separately characterize the proteomes of eusperm and parasperm in Monarch butterfly (*Danaus plexippus*). A total of 662 sperm proteins were identified, with about half shared between the two sperm morphs, and the eusperm containing a large majority of the remainder. Subsequently, we employed genome-wide divergence and polymorphism data to assess the molecular evolution of these sperm proteins. No significant difference in divergence (Ka/Ks) was detected between sperm proteins and the genomic background. In contrast, a significant reduction in non-synonymous relative to silent polymorphism was observed for sperm proteins. Combining divergence and polymorphism data to assess adaptation indicates a significantly increased rate of adaptive substitutions among sperm proteins compared to the genome. However, no

significant differences in divergence, polymorphism, or adaptation were detected among the shared or morph-specific subsets of the sperm proteome.

301T α-Amanitin resistance in *Drosophila melanogaster*: a genome-wide association approach. Chelsea Mitchell[1], Catrina Latuszek[1], Kara Vogel[2], Ian Greenlund[1], Rebecca Hobmeier[1], Olivia Ingram[1], Shannon Dufek[1], Jared Pecore[1], Felicia Nip[3], Zachary Johnson[4], Xiaohui Ji[5], Hairong Wei[5], Oliver Gailing[5], *Thomas Werner*[1] 1) Department of Biological Sciences, Michigan Technological University, 1400 Townsend Dr., Houghton, MI 49931, United States of America; 2) Department of Neurology, University of Wisconsin School of Medicine and Public Health, 1300 University Ave., Madison, WI 53706, United States of America; 3) College of Human Medicine, Michigan State University, 804 Service Road, Clinical Center, East Lansing, MI 48824, United States of America; 4) U.S. Forest Service, Salt Lake Ranger District 6944 S, 3000 E, Salt Lake City, UT 84121, United States of America; 5) School of Forest Resources and Environmental Sciences, Michigan Technological University, 1400 Townsend Dr., Houghton, MI 49931, United States of America.

We investigated the mechanisms of mushroom toxin resistance in the *Drosophila* Genetic Reference Panel (DGRP) fly lines, using genome-wide association studies (GWAS). While *Drosophila melanogaster* avoids mushrooms in nature, some lines are surprisingly resistant to α-amanitin - a toxin found solely in mushrooms. This resistance may represent a pre-adaptation, which might enable this species to invade the mushroom niche in the future. Although our previous microarray study had strongly suggested that pesticide-metabolizing detoxification genes confer α-amanitin resistance in a Taiwanese *D. melanogaster* line Ama-KTT, none of the traditional detoxification genes were among the top candidate genes resulting from the GWAS in the current study. Instead, we identified *Megalin*, *Tequila*, and *widerborst* as candidate genes underlying the α-amanitin resistance phenotype in the North American DGRP lines, all three of which are connected to the Target of Rapamycin (TOR) pathway. Both *widerborst* and *Tequila* are upstream regulators of TOR, and TOR is a key regulator of autophagy and *Megalin*-mediated endocytosis. We suggest that endocytosis and autophagy of α-amanitin, followed by lysosomal degradation of the toxin, is one of the mechanisms that confer α-amanitin resistance in the DGRP lines.

302T Comprehensive identification of *cis*-regulatory variants in yeast promoters. *F.W. Albert*[1,6], R. Cheung[2,6], L. Day[3,4,5], S. Kosuri[2], L. Kruglyak[3,4,5] 1) Department of Genetics, Cell Biology, & Development, University of Minnesota, Minneapolis, MN; 2) Department of Chemistry & Biochemistry, University of California, Los Angeles, CA; 3) Department of Human Genetics, University of California, Los Angeles, CA; 4) Department of Biological Chemistry, University of California, Los Angeles, CA; 5) Howard Hughes Medical Institute, University of California, Los Angeles, CA; 6) joint first authors.

Regulatory genetic variation in a species influences gene expression levels and is a key source of variation in complex traits. Considerable progress has been made in identifying genomic regions that harbor regulatory DNA variants (expression quantitative trait loci; "eQTL"). However, resolving these regions to the underlying causal variants remains challenging due to linkage between neighboring variants. Our limited ability to identify causal variants at scale precludes a systematic understanding of how natural regulatory variation shapes gene expression and complex traits.

To tackle this challenge, we have developed a massively parallel reporter assay to comprehensively identify *cis*-regulatory variants in yeast (*Saccharomyces cerevisiae*) promoters. Our assay probes DNA variants in promoters in two ecologically and genetically different strains: a laboratory strain and a vineyard isolate. Each variant is represented by pairs of synthetic promoter fragments that differ only in the given variant allele. We synthesized a library of 27,000 synthetic DNA oligonucleotides to assay 9,605 natural variants in 3,848 promoters. The library covers two thirds of all putative *cis*-acting variants in these strains, including all variants within 200 bp immediately upstream of each gene. We also synthesized pairwise allele combinations for closely linked variants, allowing us to measure non-additive interactions. We cloned our libraries into reporter plasmids *en masse*, such that each promoter fragment is coupled to hundreds of unique expressed barcodes. We quantify gene expression driven by each promoter fragment via high-throughput sequencing of these barcodes. We determine the effect of each variant by comparing the expression driven by alleles that differ only at the given variant. Our design assays *cis*-acting variants comprehensively and with single variant resolution.

Using a subset of our library, we have so far identified 315 *cis*-acting alleles. These variants include several that destroy or create transcription factor binding sites between the two strains. We are currently finalizing data collection for the entire library and anticipate discovery of hundreds of additional causal variants. By relating these causal variants to the excellent annotations of regulatory elements in yeast, we will ask to what extent the effects of natural promoter variants on gene expression can be predicted from genome sequence alone.

303T High throughput assessment of natural variation in the resistance to starvation stress in C. elegans using microfluidics. *H. Archer*, B. Blue, S. Banse, P. C. Phillips IEE, University of Oregon, Eugene, OR.

Caenorhabditis elegans typically feeds on rotting fruit and plant material in a fluctuating natural habitat, a boom-and-bust lifestyle. Moreover, stage specific developmental responses to low food concentration suggest that starvation-like conditions are a regular occurrence. In order to assess variation in the *C. elegans* starvation response under precisely controlled conditions and simultaneously phenotype a large number of individuals with high precision, we have developed a microfluidic device that, when combined with image scanning technology, allows for high-throughput assessment at a temporal

resolution not previously feasible and applied this to a large mapping panel of fully sequenced intercross lines. Under these conditions worms exhibit a markedly reduced adult lifespan with strain-dependent variation in starvation resistance, ranging from

304T Network Architecture and Mutational Sensitivity of the *C. elegans* Metabolome. Lindsay Johnson[1], Luke Chandler[1,2], Sarah Davies[3], *Charles Baer*[1,2] 1) Dept Biol, University of Florida, Gainesville, FL; 2) University of Florida Genetics Institute; 3) Faculty of Medicine, Department of Surgery & Cancer, Imperial College, London.

A fundamental issue in evolutionary systems biology is understanding the relationship between the topological architecture of a biological network, such as a metabolic network, and the evolution of the network. The rate at which an element in a metabolic network accumulates genetic variation via new mutations depends on both the size of the mutational target it presents and its robustness to mutational perturbation. Quantifying the relationship between topological properties of network elements and the mutability of those elements will facilitate understanding the variation in and evolution of networks at the level of populations and higher taxa.

We report an investigation into the relationship between two topological properties of 29 metabolites in the *C. elegans* metabolic network and the sensitivity of those metabolites to the cumulative effects of spontaneous mutation. The relationship between several measures of network centrality and sensitivity to mutation is weak, but point estimates of the correlation between network centrality and mutational variance are positive, with only one exception. There is a marginally significant correlation between core number and mutational heritability. There is a small but significant negative correlation between the shortest path length between a pair of metabolites and the mutational correlation between those metabolites.

Positive association between the centrality of a metabolite and its mutational heritability is consistent with centrally-positioned metabolites presenting a larger mutational target than peripheral ones, and is inconsistent with centrality conferring mutational robustness, at least *in toto*. The weakness of the correlation between shortest path length and the mutational correlation between pairs of metabolites suggests that network locality is an important but not overwhelming factor governing mutational pleiotropy. These findings provide necessary background against which the effects of other evolutionary forces, most importantly natural selection, can be interpreted.

305T A tissue specific system of chromatin regulatory genes. *C.L. Baker*, C Byers, M Walker, N.R. Powers, G Ananda, S Arat, P.M. Petkov, G.W. Carter, K Paigen The Jackson Laboratory, Bar Harbor, ME.

Most disease-associated variants in the human population fall in regulatory elements. Determining how genetic variation shapes regulatory function is a critical step in understanding complex traits. Although a variety of protein writers, readers, and erasers help to shape the chromatin landscape, we have little information about the organizing principles and underlying regulatory systems controlling its establishment and maintenance. To address this lack, we have explored how natural variation impacts chromatin function by leveraging the strength of genetically divergent inbred mice, finding considerable differences in activity at individual regulatory elements. We compared levels of the histone modification H3K4me3 and open chromatin (ATAC-seq) in male germ cells, embryonic stem cells (ESCs), hepatocytes and cardiomyocytes finding evidence of cell-type-specific *trans*-acting chromatin regulatory systems. To identify the underlying control systems, we mapped quantitative trait loci (QTL) controlling these chromatin features in male germ cells and ESCs using the BXD recombinant inbred genetic reference population. While many QTL map proximally, indicating *cis* control, commonly caused by variants in regulatory motifs, we also identified major cell-type specific *trans*-acting QTL that collectively modulate hundreds of distal regulatory elements. In germ cells, these *trans*-QTL affect both the frequency of meiotic recombination initiation, and activity of regulatory elements, suggesting a general mechanism controlling chromatin accessibility. Analysis in F1 hybrids showed that most *trans* QTL act dominantly to reduce H3K4me3 levels and close chromatin. QTL locations do not correspond with known enzyme systems metabolizing chromatin features. Collectively, these data describe an extensive, tissue-specific, chromatin regulatory system in which each regulatory factor controls a particular cohort of functionally related chromatin sites, suggesting different QTL may play key roles in establishing or maintaining differentiation. Together these QTL constitute an integrated system controlling the chromatin landscape and raises questions about the identity of the genes, their mechanism of action, and their roles in regulating differentiation and chromatin function.

306T How much of the response to selection can be attributed to discrete genes ? *N.H. Barton*[1], S. Belohlavy[1], H. Sachdeva[1], F. Chan[2] 1) Inst Sci & Technology Austria, Klosterneuburg, NOE, AT; 2) Friedrich Miescher Laboratory of the Max Planck Society, Tubingen, DE.

When analysing a selection experiment, a natural null model is the infinitesimal with linkage: genetic variance is spread evenly along the genome. Under this model, even strong selection has only a modest effect on allele frequencies - though it does increase the fraction of extreme allele frequency shifts. Such extreme shifts are difficult to distinguish from sweeps due to discrete loci, making it hard to attribute much of the selection response to specific genes. These arguments will be illustrated by an experiment that selected for longer-legged mice.

307T Malathion resistance in the *Drosophila* genetic reference panel is associated with selective sweeps and structural variation. *P. Battlay*[1], J. Schmidt[1,2], P. LeBlanc[1], C. Robin[1] 1) School of BioSciences, University of Melbourne, Parkville, VIC, AU; 2) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

There is substantial evidence that insecticides have played a strong selective role in the population ancestral to the *Drosophila melanogaster* genetic reference panel (DGRP). Here, resistance to malathion, one of the oldest and most widely used insecticides, was investigated in the DGRP in hopes of studying the quantitative nature of a trait under strong, recent selection. Genome-wide association studies identify extended haplotypes indicative of the previously described selective sweeps surrounding *Cyp6g1* and *Ace* as major malathion resistance loci. Analysis of structural variation within the *Cyp6g1* sweep reveal additional novel gene amplification events acting as eQTLs for *Cyp6g1* and its tandem paralog, *Cyp6g2.* Malathion phenotype associations with the DGRP transcriptomes include an enrichment for transcripts regulated by the mammalian Nrf2 homolog CncC, suggesting that natural variation in regulation of this pathway contributes to resistance. Results from these analyses support malathion resistance as a strong candidate selective agent in DGRP ancestors, and interrogation of sequences from outbred samples show that many of these results are applicable to *D. melanogaster* populations worldwide.

308T Genomics of larval locomotor evolution across *Drosophila mojavensis* populations. *Kyle Benowitz*[1], Joshua Coleman[1,2], Luciano Matzkin[1,3,4] 1) Department of Entomology, The University of Arizona, Tucson, AZ; 2) Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL; 3) Department of Ecology and Evolutionary Biology, The University of Arizona, Tucson, AZ; 4) BIO5 Institute, The University of Arizona, Tucson, AZ.

The ability of behavior to rapidly evolve is often thought of as a key component of animal adaptation to novel environments. Despite a breadth of model organism data elucidating the genetic basis of within-population behavioral variation, limited information is available describing the genetic architecture and specific modules responsible for producing evolutionary change in behavior. The cactophilic fruit fly *Drosophila mojavensis* has adapted to utilize four different cactus species as a host plant in four geographically and genetically distinct populations throughout the southwest US and northern Mexico. We identified between-population variability in the locomotor activity of larvae, which we hypothesize is related to foraging activity; populations inhabiting larger cacti demonstrate higher levels of activity. To investigate the genetic basis of these locomotor differences, we focused on two populations (Sonoran Desert, AZ and Santa Catalina Island, CA) with the highest and lowest activity levels. We first performed bulk segregant analysis on the F8 generation of a cross between these two populations. We next analyzed RNA-seq of larval brains from both populations to elucidate potential transcriptional contributions to behavioral variability. Additionally, we analyzed allele-specific expression from RNA-seq of F1 hybrid larval brains to assess patterns of cis-regulatory evolution. Bulk segregant analysis revealed at least two major effect QTL underlying locomotor variation within a complex 16 mb region of chromosome 3, perhaps suggesting epistatic effects on behavior. Transcriptional analyses of genes within these QTL regions identified two genes with known behavioral functions, transcriptional differences between populations, and evidence for cis-regulatory evolution between populations: *quick-to-court* and *hormone receptor 39*. Interestingly, these two genes are best known for their courtship functions in *D. melanogaster*, indicating that molecules may have species-specific behavioral roles. These results present key data regarding the mechanistic basis of subtle but ecologically relevant quantitative behavioral evolution.

309T Distinguishing close linkage from pleiotropy in multiparental populations. *F. Boehm*[1], M. Keller[2], A. Attie[2], B. Yandell[1], K. Broman[3] 1) Statistics, University of Wisconsin-Madison, Madison, WI; 2) Biochemistry, University of Wisconsin-Madison, Madison, WI; 3) Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI.

Multiparental populations, such as the Diversity Outbred (DO) mouse population, are a new resource for systems genetics studies. Distinguishing close linkage of distinct quantitative trait loci from one pleiotropic locus that associates with multiple traits has implications in biomedical research, plant and animal breeding, and other genetics applications. We develop a likelihood ratio test for the alternative hypothesis of close linkage of two loci against the null hypothesis of pleiotropy for a pair of traits that map to a single genomic region. Unlike previous tests of these competing hypotheses, our test incorporates polygenic random effects to account for complex patterns of relatedness among subjects. Additionally, our test accommodates more than two founder alleles. We use a parametric bootstrap-based method to determine statistical significance. To demonstrate its practical utility, we apply our test to a study of 400 DO mice. We share our methods in a freely available software package (https://github.com/fboehm/qtl2pleio) for the R statistical computing environment.

310T Screen for enhancers and suppressors of Parkinson's disease with the *Drosophila*-rotenone model. *A. Bracci*, A. Clark Molecular Biology and Genetics, Cornell University, Ithaca, NY.

Parkinson's disease is a neuromuscular disorder characterized by loss of dopaminergic neurons and accumulation of alpha-synuclein in the brain. There are familial forms of Parkinson's caused by rare mutations in a few genes, however most cases are sporadic (non-familial). There has been some effort to identify genes that play a role in sporadic Parkinson's, including meta-analysis of genome-wide association studies, but the cause of sporadic Parkinson's remains largely unknown. Previous

studies have established chronic exposure to rotenone as a model of Parkinson's disease in *Drosophila melanogaster*. This pesticide exposure causes severe locomotor defects, reduction in longevity, and loss of dopaminergic neurons. We have used this method to model Parkinson's disease in approximately 150 of the *Drosophila* Genetic Reference Panel (DGRP) lines. After chronic rotenone exposure, we assessed variation in locomotor phenotype among the lines by a rapid iterative negative geotaxis assay. We have utilized this phenotypic variation and the range of genetic variation in the DGRP lines to perform a genome-wide association study to identify enhancers and suppressors of the sporadic Parkinson's phenotype.

311T Systems genetics, fine mapping, and validation of candidate genes involved in opioid and psychostimulant addiction traits in a reduced complexity cross. *C.D. Bryant*[1], L.R. Goldberg[1], J.C. Kelliher[1], J.A. Beierle[1], K.P. Luttik[1], J.L. Scotellaro[1], A.M. Luong[1], J. Wu[1], E.R. Reed[1], D.F. Jenkins[1], Q.T. Ruan[1], A Al Abdullatif[1], S.L. Kirkpatrick[1], N. Yazdani[1], R.W. Williams[2], G.E. Homanics[3], W.E. Johnson[1], B. Wolozin[1], M.K. Mulligan[2] 1) Boston University School of Medicine, 72 .E. Concord, St., Boston, MA USA ; 2) University of Tennessee Health and Science Center, Memphis, TN USA; 3) University of Pittsburgh, Pittsburgh, PA USA.

Murine forward genetic studies of addiction traits can identify genetic factors and biological pathways relevant to humans. Reduced Complexity Crosses (**RCC**) between closely related inbred strains facilitate gene mapping and identification by decreasing the number of candidate variants by up to 300-fold. We used an RCC between C57BL/6J (**B6J**) and C57BL/6NJ (**B6NJ**) substrains to map opioid (oxycodone; **OXY**) and psychostimulant (methamphetamine; **MA**) behaviors in our Multi-Stage Addiction Assessment Protocol (**MSAAP**). We identified a major QTL on distal chr. 1 underlying OXY-induced locomotor activity and anxiety-like withdrawal. Due to the extreme reduced genetic complexity of the cross, complex traits frequently demonstrate a Mendelian-like mode of inheritance. This property provides the unique opportunity to immediately begin fine mapping a single locus and essentially ignoring the rest of the "genometype". We fine mapped the chr. 1 locus to 167-174 Mb by backcrossing select recombinant F2 mice. Striatal *cis*-expression (eQTL) combined with transcript/behavior covariance and regression analysis identified *Pcp4l1*, *Atp1a2*, and *Cadm3* as three high priority positional/functional candidate genes. Transcriptome analysis of the chr. 1 locus in OXY-treated mice identified a dual-hub network of *trans*-downregulated genes comprising neurodegenerative proteins APP and TAU. *Mapt* (TAU) knockout mice show preliminary evidence of enhanced OXY tolerance and withdrawal, suggesting endogenous TAU protects against chronic opioid-induced neurobehavioral plasticity. We also mapped a QTL on medial chr. 5 for MA-induced locomotor activity that contains the *cis*-modulated α2 subunit of the GABA-A receptor (*Gabra2*). The B6J allele of *Gabra2* harbors an intronic deletion resulting in decreased mRNA and protein levels. The causal role of *Gabra2* on MA-induced locomotor activity was confirmed using gene editing to repair the B6J deletion with the B6N allele. Lower *Gabra2* levels were associated with the B6J allele and conferred a greater MA behavioral response compared to the repair B6N allele. Studies are underway to validate *Gabra2* in conditioned aversive behaviors induced by the opioid receptor antagonist naloxone that also mapped to precisely the same region of chr. 5. To summarize, systems genetic and fine mapping analysis in the RCC permitted the rapid identification and validation of candidate QTL genes and network genes underlying complex behavioral addiction traits.

312T Modeling gene-by-treatment effects in a Replicable Genetic Reference Population: DrugX-induced liver injury in the Collaborative Cross as an example. *Y. Cai*[1,2], R. Corty[1,2], M. Mosedale[3], W. Valdar[1,2] 1) Department of Genetics, University of North Carolina Chapel Hill; 2) Curriculum of Bioinformatics and Computational Biology, University of North Carolina Chapel Hill; 3) Eschelman School of Pharmacy, University of North Carolina Chapel Hill.

The Collaborative Cross (CC) is a mouse multiparent recombinant inbred lines with high genetic diversity, balanced allele frequencies and dense, evenly distributed recombination sites. Using CC can increase the statistical power to detect genetic effects of a treatment control panel. Our study focuses on building an analysis framework for gene-by-treatment study in CC strains. Topics to consider include transformation of phenotypic data, QTL mapping process in CC lines, estimating confidence interval of candidate mapping peaks and finding candidate SNP in high association regions.

Taking the DrugX-included liver injury study as an example, 50 CC strains with 8 mice each were treated with DrugX or vehicle in a paired study design. For each liver injury related phenotype, we transformed the raw data to a delta that would properly represent phenotypic difference between drug-treated and vehicle mice and thereby estimate the causal effect of treatment (GxT). Then based on genetic haplotype diversity of CC lines, a linear mixed model mapping process was applied to locate candidate peaks for GxT, with a multiple-imputation process to avoid false positive peaks during mapping. Peaks were then further validated by checking founder haplotype effects and applying non-parametric bootstraps to get a location confidence interval. Within the confidence interval, we did a SNP-level genotype association test to detect candidate SNPs that provides stronger significance compared to founder haplotype association. In this way, we located QTLs associated with plasma drug concentration and plasma bile acid concentration, several candidate genes with high biological relevance were selected for further analysis.

The future direction of the study includes refining this process, integrating additional information, such as transcript

expression data, and examining other dimensions of current data, such as genetic correlations between difference phenotypes, things are that are only possible using a replicable genetic reference population-based design.

313T Identifying functional variants influencing human height variation. *T.D. Capellini*[1,2], M. Guo[2,3,4], Z. Liu[1], J. Willen[1], J Hirschhorn[2,3,4], D. Richard[1,2,5], C.P. Shaw[1], E. Jagoda[1], A.C. Doxey[5] 1) Human Evolutionary Biology, Harvard University, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Division of Endocrinology, Boston Children's Hospital, Harvard Medical School, Boston, MA; 4) Department of Genetics, Harvard Medical School, Boston, MA; 5) Department of Biology, University of Waterloo, Waterloo, Ontario, Canada.

GWAS have identified hundreds of loci associated with height, a complex trait that varies in human populations. Yet, identifying functional variants at these loci remains extremely difficult, since lead GWAS variants are often in strong linkage disequilibrium with hundreds of other common and rare variants, and are strongly enriched in the relatively unexplored non-coding portions of the genome. While height associated variants are enriched near genes involved in the cartilage growth plate – i.e., the action site of height – determining causal mechanisms is additionally challenging because growth plate chondrocytes are difficult to isolate and study. These issues have limited our ability to find and interrogate putative functional variants influencing height. To discover mechanisms by which height GWAS variants function, we performed epigenetic profiling of growth plate chondrocytes of the developing femur using the Assay for Transposase-Accessible Chromatin using sequencing (ATACseq). We uncovered tens of thousands of open chromatin regions that are shared as well as unique to the proximal and distal femoral growth plates, and which recapitulate known chondrocyte biology in humans and mice. These open chromatin regions are enriched for height GWAS loci, and height ATACseq regions are enriched near genes differentially expressed in the growth plate. Furthermore, height ATACseq regions are enriched in and disrupt binding motifs for transcription factors with known chondrocyte roles. At individual loci, our analyses identify compelling mechanisms for GWAS variants and serve to whittle-down haplotypes to fewer functional variants. For example, at the *CHSY1* locus, we identified a putatively causal variant (rs9920291) present within a repressor element, which alters the binding of the HOXD13 transcription factor, important for skeletal development. This variant significantly alters repressor activity, and its deletion modulates *CHSY1* expression in human chondrocytes. At another well-known height locus, *GDF5*, our ATACseq method allowed us to re-identify a height variant (rs4911178) within the *GROW1* enhancer, thus revealing the importance of integrating functional genomics with GWAS to more rapidly identify causal variants. Overall, integrating biologically relevant epigenetic information with genetic association results can identify biological mechanisms important for human growth.

314T Antagonistic pleiotropy at a human *GDF5 cis*-regulatory variant. *T.D. Capellini*[1,2], A.M. Kiapour[3], J. Cao[1,2,4], H. Chen[5,6], A.C. Doxey[7], Z. Liu[1,2], S. Yarlagadda[1] 1) Human Evolutionary Biology, Harvard University, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Orthopedic Surgery, Boston Children's Hospital, Harvard Medical School, Boston, MA; 4) Farm Animal Genetic Resources Exploration and Innovation Key Laboratory, Sichuan Agricultural University, Chengdu; 5) Department of Developmental Biology, Stanford University, Stanford, CA; 6) Genentech, South San Francisco, CA; 7) Department of Biology, University of Waterloo, Waterloo, Ontario.

Natural selection can at times lead to antagonistic pleiotropy, whereby beneficial consequences of a genes function in one tissue are mirrored by deleterious effects in other tissues. Many previous studies have considered antagonistic pleiotropic loci with regards to coding sequences. However, with the wealth of haplotype and phenotype data investigated via GWAS, the potential to identify regulatory variants acting as such may be possible. Human variants in the *GDF5* locus exhibit GWAS associations with height, as well as with developmental dysplasia of the hip (DDH), knee osteoarthritis (OA), and other skeletal diseases. The associated SNPs span a 130kb "short height, high risk haplotype", containing many variants, and recent studies have pointed to the importance of non-coding variants underlying each association. Using a battery of transgenic and functional genomic assays we revealed the regulatory landscape of *GDF5*, highlighting distinct enhancers controlling its expression in joints, growth plates, and other anatomical sites. For the genes' height association, we additionally discovered a short height "A" allele at rs4911178 residing within a growth plate-specific enhancer (*GROW1*) that controls bone length and has been under selection in human populations. Strikingly, we can now report that this "A" allele is also a putative causal variant underlying *GDF5* DDH association, as *GROW1* specifically influences femoral head offset, an alteration in DDH. Thus, regulatory variant rs4911178 may constitute an example of human antagonistic pleiotropy. Importantly, selection on rs4911178 likely drove the "short height, high risk haplotype" to high frequency in Eurasians, and this also has major consequences for knee OA risk. Indeed, OA risk variants reside in nearby joint-specific enhancers and several have activity in human articular chondrocytes. Within the *R4* knee joint enhancer, an OA risk "T" allele, markedly reduces enhancer activity in human chondrocytes, and deletion of the element *in vitro* or *in vivo* reduces *GDF5* expression and results in mice with dysmorphic femora with flatter condyles, shallower grooves, and narrower notches. These alterations are reminiscent of those found in the aging human OA knee. Overall, our findings reveal an evolutionary phenomenon whereby antagonistic pleiotropy at a *cis*-regulatory variant has substantial fitness trade-offs for a number of skeletal traits and diseases in humans.

315T Unraveling fatty acid variation in oat (*Avena sativa* L.) with multivariate genome-wide association analyses. *M. Carlson*[1,2], G. Montilla-Bascon[3], O. Hoekenga[4], M. Sorrells[2], J.L. Jannink[2,5], M.A. Gore[2] 1) Current: Committee on

Genetics, Genomics, & Systems Biology, University of Chicago; 2) Plant Breeding & Genetics Section, School of Integrative Plant Science, Cornell University; 3) CSIC, Institute for Sustainable Agriculture, Co?rdoba, Spain; 4) Cayuga Genetics Consulting Group LLC; 5) R.W. Holley Center for Agriculture and Health, US Dept. Agric., Agricultural Research Service.

Among a host of nutritional compounds beneficial to human health, oat (*Avena sativa* L.) contains a high concentration of oils, comprised primarily of unsaturated fatty acids, relative to other cereal crops. This favorable nutritional profile, and pre-existing adaptation to a cool, humid climate, have long rendered oats a target for crop improvement. However, without information about the genetic basis of phenotypic variation, breeders rely solely on phenotypic selection for crop improvement. Thus, to accelerate oat plant breeding efforts, we sought to identify loci associated with variation in fatty acid content. We measured the concentrations of ten fatty acids in a diversity panel of 500 oat cultivars. In parallel, we genotyped the panel with genotyping-by-sequencing, anchoring sequencing reads to the published consensus genetic map to score single-nucleotide polymorphisms (SNPs). Measurements of individual fatty acids were highly correlated, consistent with fatty acids participating in common biosynthetic pathways. We leveraged these correlations in two multivariate genome-wide association study (GWAS) approaches, multi-trait GWAS and principal component GWAS. In the first analysis, we fitted a multivariate mixed model, accounting for population structure and kinship, with all ten fatty acids. In the second, we performed a univariate association test for each principal component derived from a singular value decomposition of the phenotypic data matrix. To interpret results from the multivariate analyses, we conducted univariate association tests for total concentration (the sum of all fatty acids) and each of the ten traits. We identified five primary, putative genomic regions correlated with variation in fatty acid concentrations. By decomposing the allelic effects in these regions, we found that three regions were associated only with increases in concentration across fatty acids, while two regions were associated with divergent effects, i.e. increases and decreases of specific compounds. Multivariate GWAS exhibited the highest power to detect correlated loci (148 SNPs detected at a false discovery rate of 10%, compared to 129 and 73 in principal component and univariate analyses, respectively). By explicitly modeling the correlation structure between fatty acids in a multivariate framework, we uncovered additional associations relative to the univariate analysis. Ultimately, breeders can translate these findings to improve oat nutritional quality.

316T Estimation of phenotypic and additive genetic covariance functions for function-valued traits in the presence of amplitude and phase variation. *P.A. Carter*[1], D. Pigoli[2], K.K. Irwin[3], J.S. Marron[4], W. Mio[5], J.A.D. Aston[2] 1) School of Biological Sciences, Washington State University, Pullman, WA; 2) Statistical Laboratory, University of Cambridge, Cambridge England; 3) School of Life Sciences, École Polytechnique Fédérale de Lausanne,Lausanne, Switzerland; 4) Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC ; 5) Department of Mathematics, Florida State University, Tallahassee, FL.

Some biological traits, such as growth trajectories, vary along a continuous independent axis and are known as function-valued (FV) traits. These traits can be assessed for continuous genetic variation using a quantitative genetic approach. However, when the statistical unit is a function (or curve) both amplitude (or vertical) and phase (or horizontal) variation in the trait can arise across individuals. For example, the growth process may follow a specific time dynamic for each subject, even if the shape of the curve is similar throughout the population. This makes cross-individual comparison problematic and necessitates decoupling amplitude and phase variation prior to the analysis. In this work, we assess the impact of time warping on the estimation of phenotypic and additive genetic covariances and propose a comprehensive approach to both align the curves and account for variation in time dynamics. We apply these methods to growth trajectories in larval *Tribolium castaneum*. We find that temporal registration dramatically affects the estimates of phenotypic and additive genetic covariance functions and this analysis reveals interesting biological details that were obscured in the original study. In particular, utilizing temporal registration reveals otherwise undetected genetic variance late in the larval phase for both phase and amplitude, and uncovers genetic variance for slow early growers spending less time in the wandering phase and achieving a larger pupal mass and a younger pupation age. Implications for interpreting potentially interesting biological information about genetic variances in growth curves from analyses using and not using temporal registration are discussed.

317T A genome-wide association study to identify genetic factors affecting resistance allele formation in CRISPR gene drives. *Jackson Champer*[1,2], Joan Chung[1,2], Chen Liu[1,2], Anisha Luthra[1,2], Riona Reeves[1,2], Yoo Lim Lee[1,2], Jingxian Liu[1,2], Zhaoxin Wen[1,2], Emily Yang[1,2], Philipp W. Messer[1], Andrew G. Clark[1,2] 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.

Gene drives could allow for control of vector-borne diseases by directly suppressing vector populations or spreading genetic payloads designed to reduce pathogen transmission. CRISPR homing gene drives work by converting cells heterozygous for the drive allele into homozygotes, increasing the frequency of the drive allele in a population. However, all current CRISPR gene drives in insects form resistance alleles at high rates. Such alleles cannot be converted to drive alleles, and would halt the spread of a drive through a population. Furthermore, it seems likely based on previous results that genetic variation in natural populations would include variation in the tendency to produce resistance alleles. We developed a CRISPR homing gene drive in *Drosophila melanogaster* and crossed it into the genetically diverse *Drosophila* Genetics Reference Panel (DGRP)

lines, measuring several gene drive parameters. Successful drive conversion in the germline averaged 56±8%, with germline resistance allele formation at 39±10% among the lines. Most strikingly, resistance allele formation post-fertilization in the early embryo averaged 40±19%, with variation in resistance rates among lines ranging from 6% to 81%. To uncover the potential genetic explanation for this variation, we performed a Genome-Wide Association Study (GWAS) using our results in the DGRP lines. We found genetic polymorphisms in 25 genes that were significantly correlated with differences in the embryo resistance allele formation rate. These include *Camta*, a calmodulin-binding transcription activator, *CG7220*, involved in ubiquitin ligation, and *CG6225*, an aminopeptidase of unknown function. Such genes may increase understanding of how natural variation is involved in resistance allele formation and be good target candidates for manipulation to develop gene drives with reduced rates of resistance allele formation.

318M The genomic and molecular basis of rapid and polygenic response to selection for long leg length in mice. J.P.L. Castro[1], M.N. Yancoskie[1], M. Marchini[2], S. Belohlavy[3], M. Kucka[1], W.H. Beluch[1], R. Naumann[4], I. Skuplik[2], J. Cobb[2], N.H. Barton[3], C. Rolian[2], *Y.F. Chan*[1] 1) Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany; 2) University of Calgary, Calgary, AB, Canada; 3) IST Austria, Klosterneuburg, Austria; 4) Max Planck Institute for Cell Biology and Genetics, Dresden, Germany.

A major goal in evolutionary genetics is to understand how genomes evolve in response to selection. Here we present a genomic dissection of the Longshanks selection experiment, in which mice were selectively bred for longer tibiae relative to body mass over 20 generations, resulting in 13% increase. We combined whole genome sequencing, modeling and population genetics to show that >100 loci contributing to polygenic selection response (See abstract by N. Barton for modelling the Longshanks experiment). Here we will focus on the evolutionary processes underlying loci responding independently and in parallel between the two Longshanks replicate lines. Out of 329 putatively selected outlier loci, more than half of the loci (56%) were specific to one of two Longshanks lines, vs. 27% showing parallel response and 17% other patterns. However, we found that as the selection responses increases, parallelism became increasingly likely. The selected loci showed significant enrichment for limb developmental genes and enhancers, with the impact of *cis*-acting but not coding changes positively correlating with the strength of selection signature in both Longshanks selection replicate lines. Through functional testing of enhancers at *Gli3* and *Nkx3-2*, we show that both gain- and loss-of-function enhancer variants contributed to selection response. Using the Longshanks experiment we were able to obtain a detailed picture of the polygenic selection response and demonstrate the critical role of regulatory changes for specific loci in rapid intraspecific evolution of a major morphological trait.

319T Dissecting the genetic basis of head morphology and evolution using haplodiploid wasps. *L. Cohen*[1], J. Werren[2], J. Lynch[1] 1) Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL; 2) Department of Biology, University of Rochester, Rochester, NY.

Complex traits such as shape and size are known to be multigenic in nature. Gene interactions can be exposed by the phenomenon of epistasis, where two alleles produce a phenotype significantly different from the expected sum of their individual effects. Detecting interactions between loci becomes rapidly intractable as the number of genes increases. This has caused a serious deficit in our understanding of the complex genetic basis of morphological evolution. Use of a naturally haploid system, however, completely eliminates dominance effects and thus considerably reduces the difficulty in detecting intricate webs of genetic interactions. The *Nasonia* clade of parasitic wasps is fitting for studies in developmental genetics and molecular evolution due to their haplodiploid genetics and the ability to make interspecies crosses that result in fertile hybrid offspring. Crosses between two closely related species reveal novel hybrid phenotypes visible in the recombinant F2 males, many of which appear to be due to negative epistatic interactions among several loci. In order to identify the loci causing disrupted development, we have generated targeted introgression lines containing intervals affecting head development, and are using Multiplexed Shotgun Genotyping of hybrids to identify additional contributing loci. We are also combining our powerful genetic tools with a method of 3D imaging that allows for high-throughput phenotyping and quantitative analysis of their geometric morphometrics. Combined, these techniques allow us to identify loci responsible for complex traits, developmental incompatibilities and the nature of their epistatic interactions. Our preliminary results confirm that *Nasonia* is a uniquely powerful system with which to probe the role of complex gene interactions in the evolution of form.

320T QTL Mapping of Epicuticular Waxes to Understand the *gl* Locus of Onion. *Eduardo D. Munaiz*[1], Michael J. Havey[1,2] 1) University of Wisconsin - Madison, Madison, WI; 2) USDA.

Glossy onions have significantly less epicuticular wax compared to wild-type “waxy’ onions, and a single recessive locus (*gl*) has been proposed to condition the glossy trait. We carried out a genetic study on epicuticular wax profiles using a segregating family from a cross between a glossy inbred (B9885) and the “semi-glossy” (intermediate amounts of waxes) inbred B5351. Epicuticular wax profiles of segregating progenies were determined using Gas Chromatography Mass Spectrometry (GCMS), and were affected by genetic and environmental factors. Transgression segregation for foliage waxiness was revealed across environments and did not fit a single recessive model. Composite interval mapping revealed significant QTLs on chromosomes (chrom) 5 and 8 controlling amounts of hentriocontanone-16 (H16) the primarily wax on

onion leaves. These two QTLs accounted for 48% of the phenotypic variation with a LOD score of 17. Interestingly, epistasis was revealed between QTL on chrom 1 and 8. Over 100-fold increase of H16 was conditioned by the homozygous genotype from B5351 on chrom 1 together with the homozygous region from B9885 on chrom 8. This study demonstrates that chrom 8 carries the *gl* locus and at least three loci control accumulation of H16 on onion foliage.

321T Phenotypic clustering reveals novel genetic associations unique to distinct subtypes of polycystic ovary syndrome. *Matthew Dapas*[1], Frederick T. J. Lin[1], Richard S. Legro[2], Margrit Urbanek[1], M. Geoffrey Hayes[1,3,4], Andrea Dunaif[5] 1) Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Department of Obstetrics and Gynecology, Penn State College of Medicine, Hershey, PA; 3) Department of Anthropology, Northwestern University, Evanston, IL; 4) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) Division of Endocrinology, Diabetes and Bone Disease, Icahn School of Medicine at Mount Sinai, New York, NY.

Polycystic ovary syndrome (PCOS) is a common, complex genetic disorder. It is diagnosed by the presence of at least two of its three cardinal reproductive features: hyperandrogenism, chronic anovulation and polycystic ovarian morphology. Accordingly, there are four affected phenotypes resulting in an overall prevalence of 15% of reproductive age women globally. PCOS is frequently associated with insulin resistance, obesity, and dysglycemia, which further diversifies its symptomology. A number of susceptibility loci have been reproducibly mapped in genomewide association studies (GWAS). However, even in GWAS limited to one phenotype, the loci identified confer only modest increases in disease risk. We performed this study to test the hypothesis that there are subtypes of PCOS with distinct genetic architectures.

Using data from 932 cases from our previously published GWAS[5], we performed unsupervised, agglomerative, hierarchical clustering using Ward's minimum variance method[6] on Euclidean distances between eight normalized, age-adjusted quantitative traits: testosterone, dehydroepiandrosterone sulfate (DHEAS), insulin, glucose, BMI, sex hormone binding globulin (SHBG), luteinizing hormone (LH), and follicle-stimulating hormone (FSH). The clustering revealed two distinct phenotypic subtypes: 1) a low-BMI group (23.4%) characterized by relatively low insulin levels and high LH and SHBG levels, and 2) a high-BMI group (37.4%) characterized by higher glucose and insulin levels with relatively low SHBG and LH levels. We then performed a GWAS, as previously described[5], limiting the cases to either the low or high BMI subtypes.

We identified alleles in two novel loci (chr 4q22.3 in *BMPR1B,* $P_{meta}=1.5\times10^{-8}$; chr 7p21.1 near *FERD3L/TWIST1,* $P_{meta}=3.4\times10^{-8}$) that associate with the lean PCOS subtype at genome-wide significance, as well as several other novel and previously identified PCOS loci that associated with each subtype approaching genome-wide significance. These findings support the notion of distinct subtypes with unique etiologies in PCOS and demonstrate how precise phenotype delineation may be more important than increases in sample size in genetic association studies.

322T Association analysis reveals the importance of the *Or* gene in carrot (*Daucus carota* L.) carotenoid accumulation and domestication. S.L. Ellison[1,2], C.H. Luby[1], K. Corak[1], K. Coe[1], D. Senalik[2], M. Iorrizo[3], I.L. Goldman[1], P.W. Simon[1,2], *J.C. Dawson*[1] 1) University of Wisconsin-Madison, Madison, WI; 2) Agricultural Research Service, United States Department of Agriculture; 3) Department of Horticultural Science, North Carolina State University.

Carrots are among the richest sources of provitamin A carotenes in the human diet. Genetic variation in the carotenoid pathway does not fully explain the accumulation of such high levels of carotenoids in carrot roots. Using a diverse collection of modern and historic domesticated varieties and wild carrot accessions, an association analysis revealed a significant genomic region that contains the *Or* gene, advancing this gene as a candidate for carotenoid accumulation in carrot. *Or* has been found to control carotenoid accumulation in other crops but has not previously been described in carrot. Our analysis also allowed us to more completely characterize the genetic structure of carrot, showing that the Western domesticated carrot largely forms one genetic group, despite dramatic phenotypic differences among market classes. Eastern domesticated and wild accessions form a second group, which reflects the recent cultivation history of carrots in Central Asia. Other wild accessions form distinct geographic groups, with well-defined groups on the Iberian peninsula and in Northern Africa. Domesticated and wild carrot contain similar levels of genetic diversity, and linkage disequilibrium decays very rapidly in both groups. Using genome-wide F_{ST}, nucleotide diversity and XP-CLR, we analyzed the genome for regions putatively under selection during domestication, and identified twelve regions that were significant for all three methods of detection, one of which includes the *Or* gene. This provides further evidence that this gene was important in the early stages of carrot domestication and improvement and may explain why it has not been found with less genetically diverse mapping populations.

323T Impact of early summer hair shedding on susceptibility to fescue toxicosis and heat stress in taurine cattle. *H. Durbin*, H. Yampara-Iquise, R. Schnabel, J. Decker Animal Sciences, University of Missouri, Columbia, MO.

Beef cattle are subject to a diverse range of environmental stressors compared to other livestock production systems, creating greater opportunity for economic loss. For example, losses associated with heat stress total over $87 million a year and will likely rise in the face of future climate change. Fescue toxicosis, another major abiotic stressor, is caused by the

consumption of fescue forage infected with the endophytic fungus *Neotyphodium coenophialum* and costs the U.S. beef industry an estimated $1 billion a year.
Though no cattle are completely resistant to fescue toxicosis, some are more susceptible than others. At the beginning of the summer when well-adapted cattle are shedding their thick winter coats in preparation for warmer months, severely susceptible cattle tend not to shed their hair coat. Adaptive hair shedding is also an effective measure of heat tolerance and is closely related to cow performance, particularly day 205 weaning weight of progeny.
In cooperation with 48 producers across the United States, over 7,000 cattle were scored on a scale of 1-5 for the early-summer hair shedding phenotype. Participating cattle were genotyped using the GGP-F250 SNP panel developed by the University of Missouri, which contains ~170,000 candidate functional variants and ~25,000 variants in common with beef cattle industry standard genotyping assays. We fit several genome-wide association models in order to identify variants associated with the hair shedding phenotype while also taking breed and environmental variation into account. We fit a model with only samples grazed on fescue pasture and a separate model with samples not grazed on fescue pasture. In combination with more sophisticated modeling of genotype-by-environment interaction, these models will allow us to compare the genetic architecture of hair shedding between environments. The outcomes of this project will enable the improvement of selection tools for use in beef cattle production.

324T Reconstructing the history of polygenic adaptation using local coalescent trees. *M. Edge*, G. Coop Evolution and Ecology, UC Davis, Davis, CA.

Genome-wide association studies (GWAS) are revealing that many important human traits are polygenic, meaning that they are influenced by small-effect variants at many genetic loci. Standard population-genetic methods for inferring evolutionary history are underpowered for polygenic traits—when there are many variants of small effect, signatures of natural selection are spread across the genome and subtle at any one locus. In the last five years, several methods have emerged for detecting the action of natural selection on polygenic scores, which predict trait values from genotypes using GWAS effect sizes. However, existing methods are limited in that they do not necessarily reveal the timing or strength of selection. Here, we present a set of methods for estimating the historical trajectory of a polygenic score using a sample of contemporary genomes, taking reconstructed ancestral recombination graphs (ARGs) as input. The resulting trajectories, which are estimated on the basis of fundamental results in coalescent theory, can be tested for evidence of natural selection. We present theory and simulations supporting our procedures, as well as estimated trajectories of polygenic scores for selected human traits. Because of its grounding in coalescent theory, the framework presented here can be extended to a variety of demographic scenarios, and its usefulness will increase as both GWAS and ARG inference continue to progress.

325T The genetic basis of exploration tendency in a multiparent population of *Drosophila melanogaster*. *Z.F. Elkins*, L. Storks, A. Rahman, P.F. Petrowski, E.G. King Division of Biological Sciences, University of Missouri, Columbia, MO.

The ability of animals to move throughout their environment to find food, mates, and suitable habitat is critical to their survival and reproduction. However, this behavior can be energetically expensive and potentially costly. As a result, individuals often vary widely in their overall motility, exploration, and dispersal tendency. We used a pool-seq approach using a multiparent population to uncover the genetic basis of exploration tendency in Drosophila melanogaster. Our measurement of exploration tendency was the tendency of female fruit flies to move from a starting chamber to a novel fly chamber. We first demonstrated our measure of exploration tendency has a genetic basis by assaying 20 recombinant inbred lines to estimate the broad-sense heritability of exploration tendency (H2=0.4). To identify the source of this genetic variability, we generated 17 pairs of “high exploration” and “low exploration” pools consisting of 40 - 100 female flies and performed whole genome sequencing. We compared allele frequency and haplotype frequency differences between these pools to identify regions of the genome implicated in exploration tendency.

326T Genetic variation underlies differential responses to docetaxel and zinc treatments in *Caenorhabditis elegans*. *K. Evans*[1,2], J. van der Zwaag[3], S. Zdraljevic[1,2], D Cook[1,2], S. Brady[1,2], E. Andersen[1,2,4] 1) Molecular Biosciences, Northwestern University, Evanston, IL; 2) Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL; 3) Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands; 4) Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Chicago, IL.

Zinc is an essential element for growth and development, acting as a co-factor for more than 300 enzymes and transcription factors in the cell. For this reason, altering intracellular zinc levels can produce dramatic phenotypes ranging from cell proliferation to cell death. Furthermore, zinc concentrations create a unique cellular environment that may also have an effect on how an organism responds to other stimuli, including chemical compounds. Docetaxel is a semisynthetic compound derived from the European yew tree, *Taxus baccata*, with effective antitumor activity against many cancers. Variation in response to docetaxel is the major limiting factor of this drug, leading to the continued investigation of its metabolism and drug interactions. Previous studies show that the intracellular zinc concentration might influence, both positively and negatively, the effectiveness of docetaxel. Furthermore, it is suggested that this interaction may be genetically regulated,

making it important to study the effect of docetaxel in combination with zinc. Leveraging the power of *Caenorhabditis elegans* as a tractable metazoan model for quantitative genetics, we can identify genetic factors that underlie responses to both agents individually and in combination. Using a panel of recombinant inbred lines constructed from two genetically and phenotypically divergent strains, we mapped several quantitative trait loci (QTL) in response to each agent independently. Efforts to narrow these QTL have begun with the construction of near-isogenic lines (NILs) and the subsequent generation of CRISPR/Cas9-mediated deletions of prioritized candidate genes. Additionally, to assess the effects of both conditions in combination, we generated a dual-dose response curve and analyzed a panel of wild isolates for interactions between zinc and docetaxel. Here, we show that zinc supplementation has a range of effects on docetaxel-induced toxicity in *C. elegans*. Future work will explore this variation in response and will identify QTL for this interaction to elucidate the genetic mechanisms specific to this unique combination of treatments.

327T Evolution of age-specific decline in stress phenotypes is driven by both antagonistic pleiotropy and mutation accumulation. *E. Everman*[1,2], T. Morgan[2] 1) Molecular Biosciences, University of Kansas, Lawrence, KS; 2) Division of Biology, Kansas State University, Manhattan, KS.

As organisms age, the effectiveness of natural selection weakens, leading to age-related decline in fitness-related traits. Decreased effectiveness of natural selection at old age allows the accumulation of deleterious alleles and is likely influenced by two non-mutually exclusive genetic mechanisms: mutation accumulation (MA) and antagonistic pleiotropy (AP). MA predicts that age-related decline in fitness components is driven by age-specific sets of alleles, nonnegative genetic correlations within trait across age, and an increase in the coefficient of genetic variance. AP predicts that age-related decline in a trait is driven by alleles with positive effects on fitness in young individuals and negative effects in old individuals, and is expected to lead to negative genetic correlations within traits across age. We build on these predictions using an association mapping approach in the *Drosophila melanogaster* Genetic Reference Panel (DGRP) to investigate the change in additive effects of SNPs across age and among traits for multiple stress-response fitness-related traits, including survival following cold stress with and without acclimation and starvation resistance. We found support for both MA and AP theories of aging in the age-related decline in stress tolerance and describe specific changes in the additive effects of SNPs associated with each trait that support MA and AP. Our study demonstrates that the evolution of age-related decline in stress tolerance is driven by a combination of alleles that have age-specific additive effects, consistent with MA, as well as nonindependent and antagonistic genetic architectures characteristic of AP.

328T Genetic dissection of variation in copper resistance in *Drosophila melanogaster*. *E. Everman*, S. Macdonald Molecular Biosciences, University of Kansas, Lawrence, KS.

Heavy metals are required in trace amounts for normal development, homeostasis, and cell function; however, overexposure due to environmental pollution and contamination is a serious concern for the health of both human and non-human organisms. Copper is one such heavy metal that is essential for normal physiological function and development at low concentrations, but excessive copper exposure can ultimately lead to severe health effects and disease. Copper ion concentration is maintained by a specialized family of proteins (Metallothioneins; Mtns) that have high affinity for copper and other metals; however, little is known about how naturally-segregating variation for these and other copper-responsive proteins influences copper toxicity. *Drosophila melanogaster* is an excellent model to investigate the role of genetic variation in response to copper toxicity given high functional conservation of *Mtn* genes. We are using the *Drosophila* Synthetic Population Resource (DSPR) to identify and characterize naturally segregating alleles that contribute to copper resistance. QTL mapping will provide important insight into the genetic architecture of copper resistance using the DSPR as these lines are variable at several known genes that have been shown to influence copper resistance. Genetic variation in copper resistance will also be explored through RNA-seq of midgut tissue following exposure of DSPR lines to copper. Ultimately, this expression data will allow examination of differences in the response to copper between high- and low-resistance genotypes. Future study will focus on functionally validating candidate genes identified through mapping and gene expression analyses using RNAi- and CRISPR-Cas9-induced knockouts.

329T Complex genetic interactions drive peanut allergy models. Kelly Orgel[1], Johanna Smeekens[1], Andrew Hinton[1], Ginger D. Shaw[2], Darla R. Miller[2], Timothy A. Bell[2], Fernando Pardo-Manuel de Villena[2], A. Wesley Burks[1], Michael D. Kulis[1], *Martin T. Ferris*[2] 1) Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Peanut allergy has a prevalence of roughly 1% of the population. Severe, sometime fatal, anaphylaxis can occur in allergic individuals with even incidental peanut contact. This disease burden highlights a critical need for the development of therapeutics and prophylactics for peanut allergy; however, no such broad and effective treatments have yet been identified. One hurdle in this progress has been the lack of good small animal models of peanut allergy. Anaphylaxis can be induced by a physiologically irrelevant intra-peritoneal (IP) exposure in the C3H/HeJ (C3H) mouse strain. In contrast, C57BL/6J (B6) mice do not physiologically respond to any peanut exposure. We have recently identified the Collaborative Cross mouse strain CC027/GeniUnc as a mouse strain which exhibits severe anaphylaxis following oral exposure to peanut. We have identified a

variety of immunological, physiological and microbiota differences segregating between these three strains. Utilizing sets of crosses between these strains, we have identified interacting sets of loci driving the spectrum of refractory to permissive responses between these strains. Together these data allow us to develop more accurate models of human disease, while also dissecting the genetic architecture of allergic disease.

330T Deciphering sex-specific genetic architectures using Bayesian methods. *S. Funkhouser*[1], G. de Los Campos[2] 1) Genetics Program, Michigan State University, East Lansing, MI; 2) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.

Understanding the genetic architectures of complex human traits and diseases is of fundamental importance to science and for the development of precision medicine. Abundant evidence suggests that sex differences in the genetic architecture of traits is pervasive. For instance, many anthropometric and body composition traits (e.g., waist-to-hip ratio (WHR), a measurement of obesity and indicator of serious health conditions) are estimated to have low (< 0.8) cross-sex genetic correlations. In GWAS studies, sex differences in SNP effects have been predominantly examined using sex-stratified single marker regressions. However, these studies may fail to detect small but non-trivial additive effects among many common variants, especially after multiple test correction. Alternative Bayesian multiple regression methods have been proposed for SNP effect estimation. These methods do not suffer the burden of multiple test correction, can estimate the effects of many SNPs concurrently, and can provide better resolution in locating non-zero effect variants. We extended this approach for the study of sex differences by including in the model SNP-sex interactions and applied the resulting model to data from distantly related Caucasian individuals from the UK Biobank (n = 316,000) for several anthropometric traits including human height (a trait for which there is little evidence of sex differences in SNP effects), weight, and body composition traits. The autosomal genome was partitioned into 3,038 overlapping 1Mb windows, with male and female SNP effects estimated simultaneously within each window. Model assumptions facilitated interpretable probability estimations, such as the posterior probability that a SNP possesses a nonzero sex difference (PP-diff) in effect, and the posterior probability that a SNP has concordant or opposite effect direction between men and women (CED or OED, respectively). For WHR, 37% of all SNPs that have a non-zero effect in either men or women possess high-confidence (PP-diff >= 0.95) sex differences, while that proportion reduces to 3% and 5% for height and weight, respectively. Consistent with previous sex-stratified efforts, we observe no convincing evidence of OED-type interactions for any of the traits studied. Bayesian methods are an alternative but valuable approach to human GWAS, and we show that such methods may be crucial in unraveling sex-specific genetic architectures.

331T Pleiotropic genetic effects percolate through underlying networks of traits. *K.A. Geiler-Samerotte*[1,2], Annalise Paaby[3], Austin Taylor[2], Shuang Li[2], Charalampos Lazaris[2], Chelsea Ramjeawan[2], Naomi Ziv[2], Mark Siegal[2] 1) Stanford University, Palo Alto, CA; 2) New York University, New York, NY; 3) Georgia Institute of Technology, Atlanta, Georgia.

Understanding the mapping from genotype to phenotype is a major goal of biology. When one gene contributes to multiple traits – a phenomenon called pleiotropy – this mapping becomes more complex. We recently identified ~50 genes contributing to single cell morphology in a model eukaryote (budding yeast) and found that the majority of these genes each contribute to more than one morphological trait; some contribute to as many as 70 single cell features! To understand the mechanism by which pleiotropic genes influence so many traits, we study thousands of clonal cells from within each yeast strain. We find that the relationships between morphological features are multifaceted. For example, morphological features are related through geometric constraints (*e.g.* nuclear density decreases as nucleus area increases) as well as through cell division (*e.g.* nuclear area and density both increase during mitosis). These multifaceted relationships can obscure one and other, making it seem like pleiotropic genes influence independent traits when in fact the traits in question are related in many ways. We show that pleiotropy in this system does not often result from each gene having multiple independent functions. Instead, pleiotropic genetic effects percolate through underlying networks of traits. This conclusion is promising in terms of mapping genotype to phenotype. It suggests that a more comprehensive understanding of cell and organismal biology can lead to predictions about when a genetic change will have pleiotropic effects.

332T Discriminating primary from metastatic tumors in melanoma using Whole-Genome Gene Expression Profiles. *A. Gonzalez-Reymundez*[1,2], G. de los Campos[1,2,3], A.I. Vazquez[1,2] 1) IQ-Institute of Quantitative Health, Michigan State University, East Lansing, MI; 2) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 3) Department of Statistics and Probability, Michigan State University, East Lansing, MI.

Metastasis is one of cancer's most lethal characteristics, and this is particularly true for skin melanomas where a metastatic event drastically reduces the patient's chance of survival. Differentiating metastasis from primary tumors is essential to assess cancer aggressiveness. The metastatic process has distinct gene expression profiles that can be captured by RNA sequencing technologies. Here we investigate whether pre-selected genes or Whole-Genome Gene Expression profiles (WGGEP) can discriminate melanoma samples into primary and metastasis, as well as to compare modeling approaches to do so. Data (n=448) was from primary and metastatic skin melanoma samples provided by The Cancer Genome Atlas. The set of genes informing the predictive models included (*i*) subsets of genes internally selected by their marginal association to the outcome (primary/metastasis) in independent training sets, (*ii*) sets of an increasing number of genes, selected at random,

and (*iii*) all genes available, WGGEP. All the models were adjusted in training sets and their predictive ability was measured with their AUC in a separated testing set –in a Cross-Validation design (AUC-CV). We trained Bayes-B and Bayesian Ridge Regressions (Bayes-RR). The prediction ability was highly dependent on the number of genes in the model. When the genes were selected in the training set based on association to outcome, both regressions reached their maximum of AUC-CV ≈ 0.7 at 50 genes adjusted via Bayes-B. When these models had more genes incorporated, prediction accuracy was reduced reaching a plateau at approximately AUC-CV ≈ 0.65. The WGGEP regression had a total prediction accuracy of ~0.65 AUC-CV. We conclude that, although multiple gene-enable models provide a gain in prediction accuracy, selecting genes a *priori* gives substantially better results. The internal selection was also more effective than Bayes-B with WGGEP. For the same reason, Bayes B, which combines features of variable selection and shrinkage, performed slightly better than Bayesian-RR, which does not perform any variable selection. Overall, we speculate that in reduced sample size settings, the models cannot accurately be trained for the demands of models with more than 50 features -in this research problem-, and marginal association becomes more informative.

333T The effects of IIS/TOR signaling on sex-differential gene expression in Drosophila. *R.M. Graze*[1], R. Tzeng[2], T.S. Howard[1], M.N. Arbeitman[2] 1) Biological Sciences, Auburn University, Auburn, AL; 2) Biomedical Sciences Department and Center for Genomics and Personalized Medicine, Florida State University, Tallahassee, Florida.

The core functions of the insulin signaling (IIS/TOR) pathway are in nutrient sensing, energy homeostasis and growth. Insulin signaling is known to interact directly and indirectly with sex determination, and often plays a role in directing sexually dimorphic complex traits. For example, in Drosophila the IIS/TOR pathway is required during development for body size dimorphism and in adults for activity level dimorphism and modulation of mating behaviors. To understand the degree to which the insulin signaling pathway contributes to sexually dimorphic gene expression in adult animals, we examined the effect of perturbation of the pathway on gene expression in male and female Drosophila. A dominant negative *insulin-like receptor* transgene (InR^{DN}) was expressed in adults using the drug inducible, ubiquitously expressed, GeneSwitch GAL4 system. Expression was assessed by RNA-seq of head tissues, in replicate for each sex expressing the dominant negative allele, and for age matched controls. Our data reveal that males and females have a shared regulatory response to reduction of insulin signaling by GeneSwitch. This shared response is heavily enriched for genes and pathways involved in metabolism. However, a large number of genes also show striking sex differences only under the perturbation conditions. Interestingly, this includes sex-differences in expression of immune, defense and stress response genes primarily driven by male-specific effects of the perturbation. Although, female specific effects, predicted to be associated with lifespan extension, are also observed. Finally, a subset of genes are dimorphically expressed only when insulin signaling functions normally. For instance, energy homeostasis genes regulated by insulin signaling- including those already known to be sex-differentially expressed in Drosophila (e.g. *sxe2)*. Collectively our results suggest that insulin signaling is important for sex differences related to energy homeostasis and may also mediate differences between males and females in stress responses, including defense responses and potentially the response to infection. Collectively, these results have broad implications for the role of insulin signaling in the physiological underpinnings of trade-offs, sexual conflict and sex differences in expression variability.

334T Uncovering the genetic basis of leaf shape, a potentially adaptive trait, in *Ipomoea trifida*. *Sonal Gupta*, Dr. Regina Baucom Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI.

Identifying the genes underlying functionally important traits is a major goal in evolutionary biology. Leaf shape, for example, is thought to be an adaptive trait that enables plants to grow in varied environmental conditions due to its physiological role: lobed leaves are a potential adaptation to stressful environments (especially heat and water stress environment) because a larger perimeter/area ratio and thinner boundary layers can reduce heat and water stress. Additionally, leaf shape has been hypothesized as a defense adaptation against disease and herbivores. Despite the physiological and functional significance of leaves, we understand very little about the genetic basis of leaf shape in most wild plant species. *Ipomoea trifida*, native to Northern and Southern America, is highly diverse for leaf shape. Individuals exhibit either entire (heart-shaped) or lobed leaves (dissected) with varied aspect ratio (the extent to which they are round vs. long). There is an association between leaf shape and the environment in this species: leaf circularity decreases with increasing temperature, whereas aspect ratio increases. Thus, it is likely that leaf shape is an adaptive trait in this species. As a first step towards identification of genes underlying leaf shape, we performed a transcriptomic study on six families of *I. trifida*. The data obtained was used to assemble and annotate a comprehensive de-novo leaf transcriptome of *I. trifida*. The assembled transcriptome was used as a reference for the identification of differentially expressed transcripts (DETs) associated leaf circularity in all six families. Among the differentially expressed transcripts, there were proteins involved in plant development and growth, which may represent possible mechanisms responsible for cell proliferation leading to leaf shape differences. We also found proteins involved in resistance against blight and TMV, which warrant further investigation to validate the role of leaf shape in defense mechanisms. Overall, the identified DETs form a preliminary set of genes exhibiting expression difference in variants of leaf shape in *Ipomoea trifida*. Ongoing work involves using the transcriptomic data to identify SNPs that might potentially be associated with leaf shape and might be under selection. Further work will involve QTL

mapping of leaf shape and a reciprocal transplant experiment to assess fitness of leaf shape variants in different environments.

335T Single Plant GWA study validated by Bulk Segregant Analysis for plant height in the maize *Shoepeg* population. *A. Gyawali*[1], M. Emery[1], V. Shrestha[1], R. Angelovici[1], S. Garcia[2], T. Beissinger[1] 1) Division of Biological Sciences, University of Missouri Columbia, Columbia, MO; 2) USDA,ARS/University of Missouri, Columbia, MO.

Genome wide association (GWA) studies have been a powerful tool for identifying quantitative trait loci (QTL) and casual SNPs/genes associated with various important traits in crop species. Typically, GWA Studies in crops have been done using inbred lines, where multiple replicates of the same inbred are measured and the average phenotype is taken as the response variable. However, GWA studies have rarely been performed using phenotypes and genotypes from individual plants sampled from an open-pollinated population. Here we have used the *Shoepeg* maize landrace, collected from a farm in Southern Missouri in the 1960's, to evaluate if such a 'single-plant' GWA study can efficiently and powerfully be used to detect significant associations SNPs for plant height (PH). A total of 306k markers were identified and FarmCPU was used to perform the association analysis which detected 25 significant SNPs ($P \leq 0.00001$) for PH. The results from our single-plant GWA study were validated by comparing to Bulk Segregant Analysis (BSA) for PH.

336T Demography obscures signatures of selection in island populations of *Drosophila santomea* and *D. yakuba*. *C. Han*[1], P. Reilly[1], M. Rebeiz[2], P. Andolfatto[3] 1) Quantitative and Computational Biology, Princeton University, Princeton, NJ; 2) Biological Sciences, University of Pittsburgh, Pittsburgh, PA; 3) Ecology and Evolutionary Biology, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ.

Traits controlled by multiple genes under strong selection are often seen as key opportunities for studying the genomic signatures and phenotypic variations that result from selection. One such example of a well-studied trait is abdominal color pigmentation in the *Drosophila yakuba* subgroup. The loss of abdominal pigmentation in *D. santomea*, an island species endemic to the African volcanic island of São Tomé, has been suggested to be the result of soft selective sweeps acting on the loci underlying the trait since *D. santomea*'s divergence from its mainland sister species *D. yakuba*. Inspection of the flanking sequence of the candidate loci in *D. santomea* and *D. yakuba* shows large clusters of long identical haplotypes in *D. santomea*, while *D. yakuba* haplotypes cluster more in accordance with neutrality. Extended haplotypes are also observed in putatively neutral non-coding regions selected as controls in *D. santomea*. These extended haplotypes as well as long tracts of identity-by-descent across the genome suggest that the *D. santomea* population likely have experienced a recent, strong bottleneck on São Tomé that may also have affected the *D. yakuba* population. Such demographic events severely confound selection signatures and should be taken into proper consideration when making inferences based on population genetic data.

337T Using metabolomics to study a complex trait in *Drosophila melanogaster*. *B.R. Harrison*[1], L Wang[2], E.V. Hoffman[1], D.E. Promislow[1] 1) Department of Pathology, University of Washington, Seattle, WA; 2) Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA.

Variation in complex traits is influenced by genotype, environment and the multidimensional endophenotype, composed of the epigenome, transcriptome, proteome, microbiome and metabolome. Here we describe our work using untargeted metabolomics to study the variation in lifespan of *Drosophila melanogaster* under stress. Our goal is to evaluate the use of the metabolome in association studies with several potential benefits in mind. We ask if the metabolome can inform our understanding of variation in complex traits, provide useful biomarkers, and what power the metabolome has to explain missing heritability. We measured the lifespan of 180 inbred and sequenced lines from the Drosophila Genetic Reference Panel under stress and chose lines at the phenotypic extremes for metabolomic analysis. Using liquid chromatography mass spectroscopy, we measured ~6000 features in soluble fractions of whole flies under both stress and control conditions. Multivariate analysis reveals that long-lived lines have a more robust metabolome compared to that of short-lived lines, and univariate analysis revealed specific metabolites that associate with stress sensitivity. We also used partial least-squares discriminant analysis to build models that accurately distinguished long from short-lived lines based on their metabolomic profiles, even in the absence of stress. Thus, the metabolome provides adequate biomarker data for trait prediction. We perform pathway analysis to discover metabolic pathways that associate with stress sensitivity and report the effects of glycogen and folate metabolism on stress sensitivity. Lastly, we try to leverage the metabolome profiles along with the fully sequenced genomes of each line to explain the heritability of stress sensitivity in our study.

338T The role of detoxification gene *Cyp12d1* in resistance to caffeine in *Drosophila*. *D. Hernandez-Aquino*, S. Macdonald Molecular Biosciences, University of Kansas, Lawrence, KS.

Xenobiotic compounds, such as those produced by plants as a defense against herbivory, as well as a wide range of agricultural pesticides, are a constant challenge for animals. By finding loci underlying variation in the toxicity of such compounds we seek to understand the mechanisms behind xenobiotic response. Previously, we mapped QTL contributing to resistance to caffeine, a model xenobiotic compound. At one locus we identified the cytochrome P450 *Cyp12d1,* a known

detoxification gene, and supported its involvement in resistance via RNAseq and midgut-specific RNAi. *Cyp12d1* is known to exhibit copy number variation (CNV) within populations of *D. melanogaster*. CNVs have been linked to disease susceptibility and are associated with insecticide resistance, confirming that structural alterations caused by CNVs are an important contributor to phenotypic variation. Thus, we hypothesize that there is a causative, positive relationship between *Cyp12d1* copy number and caffeine resistance. To directly test this we used CRISPR-Cas9 editing to generate loss-of-function mutations in *Cyp12d1*. We crossed the mutant strains (with 0 functional copies of *Cyp12d1*), and the un-edited wildtype progenitor strain to inbred lines varying in copy number. This generated multiple pairs of F1 genotypes with different copy number (1 or 2, 2 or 3) in otherwise identical genetic backgrounds. In each case resistance increased with copy number, with the effect being significant in 3/4 cases ($P < 0.0005$). This experiment directly validates the hypothesis that more copies of *Cyp12d1* increases resistance to caffeine. To confirm this observation we are carrying out two additional experiments. First, we are editing a line that is homozygous for a *Cyp12d1* duplication, creating lines that contain mutations in one or both copies of the gene while maintaining the same background. We have already generated animals with premature stop codon mutations in the gene, and are making the animals homozygous in preparation for phenotyping. Second, by intercrossing many inbred strains we have created a pair of outbred populations that are fixed for either the single or double copy allele, allowing us to test effects of *Cyp12d1* copy number in a variable genetic background. Our research will provide further insight into the functions of copy number variants, and the genetic basis and mechanisms of xenobiotic resistance in general.

339T Sex-ratio heritability in the common snapping turtle: a reptile with temperature-dependent sex determination. *K. Hilliard*, T. Rhen, S. Singh University of North Dakota, Grand Forks, ND.

Temperature-dependent sex determination (TSD) is observed in many lizards, turtles and crocodilians. Among-family sex ratio differences are well-documented in reptiles with TSD. However, the underlying causes of phenotypic variance in TSD are unclear. To determine the relative contribution of additive genetic effects and non-genetic maternal effects to sex-ratio variability, we used a paternal half-sib, maternal full-sib breeding design in a captive population of common snapping turtles. Eggs were collected and incubated at constant temperatures that produce mixed sex ratios in 2016 (29 clutches, 1377 eggs) and 2017 (31 clutches, 1462 eggs). Preliminary analysis of data from 2017 produced a narrow-sense heritability estimate of 63.764% of the total phenotypic variance. Variances among sires and among multiple dams mated to a single sire are approximately equal, indicating that dominance variation and non-genetic maternal effects have negligible roles in inter-family sex-ratio differences. In comparison, the same analysis conducted for egg mass at oviposition and hatchling body size yielded substantial variance due to non-genetic maternal effects and heritability near zero. In conclusion, sex-ratio variation in this system was found to be highly heritable. This may allow adaptive evolution of TSD in response to climate change in the common snapping turtle.

340T Detecting quantitative trait loci in mice bred for high levels of voluntary wheel running. *D. Hillis*[1], L. Yadgary[2], G. Weinstock[3], F. Pardo-Manuel de Villena[2], D. Pomp[2], S. Xu[1], T. Garland[1] 1) University of California: Riverside, Riverside, CA; 2) University of North Carolina, Chapel Hill, NC; 3) Jackson Laboratory, Bar Harbor, ME.

Understanding the biological basis of exercise behavior has become increasingly relevant for maintaining healthy lifestyles in humans and other mammals. Various quantitative genetic studies and artificial selection experiments have conclusively demonstrated substantial narrow-sense heritability for exercise behavior in both humans and laboratory rodents. One selection experiment with laboratory mice incorporates 4 replicate lines bred for high voluntary wheel running and 4 non-selected control lines. After 61 generations, the genomes of 80 mice (10 from each line) were fully sequenced and single nucleotide polymorphisms (SNPs) were identified. Using the Minimum Quadratic Variance Unbiased Estimation (MIVQUE) method (Xu and Garland 2017), we determined which SNPs are differentiated between the 4 High Runner (HR) and 4 control lines. Results are compared with data from the MegaMUGA SNP chip published previously, and demonstrate similar regions of high differentiation, present on most chromosomes. Additionally, as expected, the higher resolution of the sequence data has produced additional promising SNPs outside these regions.

341T Tracking signatures of response over 20 generations of selection for long leg length in mice. *L. Hiramatsu*[1], S. Belohlavy[2], M. Ku?ka[1], N. Barton[2], C. Rolian[3], Y. F. Chan[1] 1) Friedrich Miescher Laboratory, Max Planck Institute, Tübingen, Germany; 2) Institute of Science and Technology Austria, Klosterneuburg, Austria; 3) University of Calgary, Calgary AB, Canada.

Selection experiments offer a powerful approach to study how genomes evolve in response to selection pressures. Especially useful are pedigreed selection experiments, which offer time-series reconstruction of genomic change. We used mice from the Longshanks selection experiment for longer relative tibia length in mice. Two replicate lines were selectively bred over 20 generations using up to 16 breeding pairs per line per generation, resulting in a 13-15% increase in tibiae length (see abstract by Y.F. Chan). We tracked haplotype segregation through each generation by pedigree-assisted imputation. We contrasted locus-by-locus allele trajectories with theoretical predictions of selection response (see abstract by N. Barton). We also traced variation of local recombination rate and events through tracking the breakdown of haplotypes. We show here that despite small population size of only $N_e = 46$, selection response remains rapid and robust

within the first 20 generations. The Longshanks experiment provides a comprehensively detailed system to study how the mammalian genome responds to selection and allowed the testing of theoretic models of varying genetic architectures.

342T Empirical distributions of mutational effects define gene-specific neutral models of regulatory evolution. *A. Hodgins-Davis*, F. Duveau, E. Walker, P. Wittkopp Dept of EEB, University of Michigan, Ann Arbor, MI.

Mutations occurring randomly throughout the genome have the potential to have non-random effects on phenotypes based on the regulatory steps required to translate genotypes into phenotypes, *i.e.* the genotype-phenotype map. Differences in the structure of the genotype-phenotype map among phenotypes may bias the phenotypic effects of new mutations, shaping the variation available for evolutionary change. For the phenotype of gene expression, an example of this is the observation that a canonical TATA box motif in a gene's promoter predicts higher mutational variance as well as faster rates of change in experimental evolution and greater polymorphism and divergence. Distinguishing the origin of patterns in phenotypic variation in natural populations thus requires separating the biases introduced by mutation alone from the effects of processes like selection and drift. The Wittkopp lab has previously described the spectrum of new mutations influencing expression of the gene TDH3 and shown that, even within a single gene, local (*cis*) and distal (*trans*) mutations exhibit opposite biases in the magnitude of expression changes they induce: *cis* TDH3 mutations show larger magnitude decreases while trans mutations show larger magnitude increases in expression. However, the extent to which these patterns are generalizable across genes with different promoter architectures and evolutionary histories remains an open question. We will present results from mutagenesis of yeast promoters differing in expression noise, canonical TATA box status, number of predicted regulators, and nucleosome positioning to test hypotheses about the consequences of promoter architecture for the mutational variation that provides the raw material for evolution.

343T Genome-wide Association Study identifies genetic variants associated with facial attractiveness. *Bowen Hu*, James Li, Jinkuk Hong, Jason Fletcher, Jan Greenberg, Marsha Mailick, Qiongshi Lu University of Wisconsin Madison, Madison, WI.

Genome-wide association studies (GWAS) have been successful in identifying associations between single-nucleotide polymorphisms (SNPs) and a variety of human complex traits. In this project, we studied the genetic basis of facial attractiveness, which is of great interests to sociologists and psychologists due to its impact on mate quality, fertility, and job-related outcomes. Twin-based and family-based studies have suggested substantial heritability for facial attractiveness. However, few genetic components for facial attractiveness has been identified and our understanding of its genetic architecture is far from complete. We conducted a GWAS for facial attractiveness using data from 3,928 individuals with self-reported European-ancestry in the Wisconsin Longitudinal Study (WLS). Facial attractiveness was quantified by 12 referees' ratings on high-school yearbook photos. Each photo was rated by six female coders and six male coders recruited from the same cohort. We stratified our analyses by sex (of sample and coder) to better understand the shared and distinct genetic architecture of male and female attractiveness. We identified two genome-wide significant loci. SNP rs2999422 is associated with non-sex-stratified attractiveness rated by female coders (p=9.2E-10). The closest gene *ANTXRL* is specifically expressed in the testes. SNP rs10165224 is significantly associated with female attractiveness rated by male coders (p=3.4E-08). This locus has previously been suggested to associate with height. In addition, nine loci showed suggestively significant associations in our analyses (p*CTSS* (p=9.63E-07), *RPL22* (p=1.52E-07), and *SYT15* (p=1.04E-06). Finally, we estimated genetic correlations between facial attractiveness and 50 complex traits. We identified a significant and negative genetic correlation between female BMI and female attractiveness rated by male coders (p=4.71E-05). This correlation was not observed in males. Instead, total cholesterol level was negatively correlated with male attractiveness rated by female coders (p=5.11E-04). In summary, our analysis identified novel genetic loci for facial attractiveness and provided insights into its genetic architecture.

344T Systems biology of age-related phenotypes in response to diet restriction in *Drosophila*. *K. Jin*[1], K. Wilson[2], T. Hilsabeck[2], J. Beck[2], C. Nelson[2], R. Brem[2], P. Kapahi[2], D.E. Promislow[1,3] 1) Pathology, University of Washington, Seattle, WA; 2) The Buck Institute, Novato, CA; 3) Biology, University of Washington, Seattle, WA.

Statement of purpose: Understanding how changes in environment influence how organisms age is a critical yet complex challenge in the field of aging. Many studies have shown that most organisms live longer and are generally healthier when they undergo dietary restriction (DR). However, several studies have found that within the same species, different genotypes respond differently to DR than others. Furthermore, it is yet unclear whether lifespan and healthspan are always correlated within a population, or if the two phenotypes can be uncoupled. Here, we use a systems biology and experimental population approach to explore this complex set of age-dependent traits.

Methods: To explore the mechanisms underlying variation in DR-response, we observed a set of phenotypes in 178 inbred fly lines from the *Drosophila* Genome Reference Panel (DGRP) that were exposed to a restricted (low yeast; DR) or *ad libitum* (high yeast; AL) diet. These phenotypes included lifespan, climbing ability, and other fitness-related traits. We use metabolic and genetic markers to interrogate mechanisms underlying DR response, with a focus on lifespan response. We also use these data to examine relationships that exist between various quantitative traits.

Summary of results: Univariate statistical analysis reveals metabolites that significantly correlate with lifespan response, including most of the proteinogenic amino acids, metabolites involved in purine metabolism, and a-ketoglutarate metabolism. Multivariate statistical analysis indicates the metabolome can accurately predict lifespan response across the DGRP. Metabolic and genetic associations reveal relationships between lifespan and healthspan traits that suggest common metabolic and genetic pathways that underlie these age-related phenotypes.

345T Haplotype Based Analysis Established *SorCS1* as a Positional Candidate Gene for Tame Behavior in a Fox Model of Animal Domestication. *J. Johnson*[1], H. Rando[1], G. Zhang[2,3,4], R. Gulevich[5], A. Vladimirova[5], A. Kharlamova[5], L. Trut[5], A. Kukekova[1] 1) Animal Sciences Department, College of ACES, University of Illinois at Urbana-Champaign, IL 61801, USA; 2) China National Genebank, BGI -Shenzhen, 518083, Shenzhen, Guangdong, China; 3) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223, Kunming, China; 4) Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark; 5) Institute of Cytology and Genetics of the Russian Academy of Sciences, Novosibirsk, 630090, Russia.

Statement of Purpose

Outbred animal models developed by selection for specific traits and wild animal populations which underwent adaptations to specific environments hold great potential for the identification of genomic targets of selection utilizing next generation sequencing data. However, while some genomic regions under selection are characterized by a significant reduction in heterozygosity, others do not reach fixation, making them difficult to characterize. Preservation of heterozygosity in these regions complicates the identification of haplotypes which have an effect on a trait and the discovery of causative mutations. Here we used strains of red fox (*Vulpes vulpes*) with markedly different behavioral phenotypes to identify genomic regions associated with the response to selection for behavior. We re-sequenced a subset of foxes from tame, aggressive, and conventional farm-bred strains and identified 103 regions with possible signs of selection. We found that a subset of these regions overlap with previously identified behavioral quantitative trait loci (QTL) and analyzed one such region in detail. This region contained a single gene, *SorCS1*, which encodes the main trafficking protein for AMPA glutamate receptors and neurexins.

Methods used

The sequencing data was too shallow to identify individual haplotypes; we developed 25 short insertion/deletion markers distributed relatively equally across a 5-Mb interval that spans the *SorCS1* region and extends to either side. The markers were genotyped in the original samples, additional samples from tame and aggressive foxes, and in 537 F2 animals that demonstrate a wide spectrum of behaviors.

Summary of Results

Haplotype analysis of the tame strain identified a single linkage disequilibrium block located over the SorCS1 gene. Within this block we found three haplotypes that differed in frequency between the tame and aggressive strains, one haplotype was observed with a frequency of 60.6% in the tame strain that was not observed in the aggressive strain, two haplotypes were rare in tame but frequent in the aggressive strain. Differences in behavior of F2s homozygous for any of the these three haplotypes were statistically significant (p=0.03), supporting SorCS1 as a strong positional candidate for fox domestication. Our study underscores the value of outbred populations for finding potential targets of selection combined with cross-bred pedigrees for identifying critical haplotypes with an effect on a trait.

346T Automated measurement of ecologically-relevant social behavior in a laboratory environment. *Z.V. Johnson*, M.C. Lowder, T. Lee, J.T. Streelman, P.T. McGrath School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA.

Advances in next-generation sequencing and gene editing have enabled the use of new species to study the genomic basis of evolutionary change. Lake Malawi cichlid fishes are promising in this regard, having undergone explosive speciation in the past two million years resulting in ~1,000 species that can be intercrossed to produce hybrids. We study the genetic basis of species differences in bower building behavior, an extended phenotype used by males to signal reproductive fitness. Depending on the species, males construct “pit” or “castle” bower structures by manipulating sand with their mouths for days or weeks at a time. In order to quantify bower building over long time periods in standard aquatics facilities, we have designed a low-cost, automated, chronically recording system to measure behavior in ~12 tanks simultaneously. This system combines use of a Raspberry Pi computer to record HD video and a Microsoft Xbox Kinect depth sensor to measure the height of the bower structure. To measure spatial patterns of sand manipulation, we have developed a novel Hidden Markov Model based approach to analyze video data. Using this system, we demonstrate species-level and individual-level differences in spatial decision-making about where to collect and deposit sand over time, including examples of nocturnal bower building. We also demonstrate the sequential and distinct expression of both parental phenotypes in pit x castle F_1 hybrid males, motivating neurogenetic dissection of how one brain can express two species-divergent behaviors. Combined with quantitative mapping and resequencing approaches, this system will enable powerful experimental strategies for investigating the genetic basis of bower building.

347T Putting up with parasites: a developmental regulator confers tolerance of transposition in the *Drosophila* female germline. *Erin Kelleher*[1], Uche Akoma[1,2], Jaweria Jaweria[1], Lily Ortega[1], Wenpei Tan[1] 1) Biology and Biochemistry, University of Houston, Houston, TX; 2) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

Transposable elements (TEs) are obligate genetic parasites that propagate in host genomes by replicating in germline nuclei, thereby ensuring transmission to offspring. This selfish replication not only produces deleterious mutations---in extreme cases, TE mobilization induces genotoxic stress that prohibits the production of viable gametes. Host genomes could reduce these fitness effects in two ways: resistance and tolerance. Resistance to TE propagation is enacted by germline specific small-RNA-mediated silencing pathways, such as the piRNA pathway, and is studied extensively. However, it remains entirely unknown whether host genomes may also evolve tolerance, by desensitizing gametogenesis to TE-induced genotoxic stress. In part, the absence of research on tolerance reflects a lack of opportunity, as small-RNA-mediated silencing evolves rapidly after a new TE invades, thereby masking existing variation in tolerance.

We have exploited the recent the historical invasion of the *Drosophila melanogaster* genome by *P*-element DNA transposons in order to study tolerance of TE activity. In the absence of piRNA mediated silencing, the genotoxic stress imposed by *P*-elements disrupts oogenesis, and in extreme cases leads to atrophied ovaries that completely lack germline cells. By performing QTL-mapping on a panel of 1600 recombinant inbred lines (RILs) that lack piRNA-mediated silencing of *P*-elements, we uncovered multiple QTL that are associated with differences in tolerance of oogenesis to *P*-element transposition. We observed continuous phenotypic variation in germline loss, which we associated with multiple QTL on chromosomes 2 and 3. The most significant QTL explains 14% of phenotypic variation, and is localized to a small 230 Kb region in the euchromatic portion of chromosome 2L. The LOD peak resides in the *bruno* locus, which codes for a critical and well-studied developmental regulator of oogenesis. Surprisingly, we have discovered that multiple *bruno* loss of function alleles are strong dominant suppressors of ovarian atrophy, allowing for the development of mature egg-chambers in the face of *P*-element activity. Further genetic and cytological analyses reveal that *bruno* tolerance is explained by enhanced retention of germline stem cells in dysgenic ovaries, which are typically lost due to DNA damage. Our observations reveal segregating variation in TE tolerance for the first time. They further suggest that developmental regulation of oogenesis is an important tolerance mechanism, which could minimize the fitness consequences of newly invading TEs.

348T Repeated cis-regulatory evolution of a metabolic bottleneck gene implies evolution might be predictable. *Meihua Christina Kuang*[1,2], Jacek Kominek[1], William G. Alexander[1], Jan-Fang Cheng[3], Russell L. Wrobel[1], Chris Todd Hittinger[1] 1) Laboratory of Genetics, Genome Center of Wisconsin, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53706, USA; 2) Current address: Department of Neurobiology, University of California San diego; 3) DOE Joint Genome Institute, Walnut Creek, CA 94598, USA.

Repeated evolutionary events imply there are underlying genetic constraints that likely make evolution predictable. The frequent cis-regulatory changes in morphological traits are thought to resolve constraints in pleiotropic developmental genes that are reused during development. However, the constraints acting on metabolic traits during evolution are less well studied. Here we show a metabolic bottleneck gene has repeatedly adopted similar cis-regulatory solutions during evolution, likely due to its roles as a metabolic bottleneck and a pleiotropic gene integrating flux from multiple metabolic pathways. Specifically, the genes connecting *GAL*actose catabolism to glycolysis, which encode phosphoglucomutase activity (*PGM1*/*PGM2*), have likely gained and lost binding-sites for the galactose-specific transcription factor Gal4 repeatedly during yeast evolution. There is a strong association between the number of predicted Gal4 binding-sites of the bottleneck genes *PGM1/2* and galactose growth rate in 17 species across the yeast family Saccharomycetaceae. This galactose-mediated regulation of *PGM1/2* is necessary to support vigorous growth on galactose in multiple yeast species, including *Saccharomyces uvarum* and *Lachancea kluyveri*. Moreover, the addition of galactose-inducible *PGM1* alone is sufficient to improve galactose growth of multiple species, including *Saccharomyces cerevisiae*. The strong association even enables remarkably accurate predictions of galactose growth phenotypes between closely related species in the genus *Lachancea*. Since metabolic pathways are highly interconnected, we argue that cis-regulatory evolution might be widespread at pleiotropic genes that control metabolic bottlenecks and intersections.

349T Transcriptome analysis of anterior pituitary and adrenal gland in the red fox strains with different stress-responsiveness. *Anna Kukekova*[1], Jessica Hekman[1], Jennifer Johnson[1], Whitney Edwards[2], Anastasiya Vladimirova[3], Rimma Gulevich[3], Anastasiya Kharlamova[3], Gregory Acland[4], Lori Raetzman[2], Lyudmila Trut[3] 1) Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL 61801, USA; 2) Department of Molecular and Integrative Physiology, University of Illinois at Urbana-Champaign, IL 61801, USA; 3) Institute of Cytology and Genetics of the Russian Academy of Sciences, Novosibirsk, 630090, Russia; 4) Baker Institute for Animal Health, Cornell University, Ithaca, NY 14853, USA.

Statement of Purpose: Domesticated species are characterized by reduced fearfulness, increased social tolerance, and increased resistance to stress. These behaviors are closely linked to reduced reactivity of the hypothalamic-pituitary-adrenal

(HPA) axis, the hormonal cascade associated with the stress response in mammals. Specifically, reductions in circulating levels of adrenocorticotrophic hormone (ACTH), released by the anterior pituitary, and glucocorticoids, released by the adrenals, have been demonstrated in several domesticated species. Here we took advantage of the tame and aggressive strains of the red fox developed by long term-selection for behavior to explore mechanisms which could be associated with changes in the HPA axis reactivity in the course of animal domestication.
Methods used: RNA extracted from the anterior pituitaries of six tame and six aggressive foxes and from the right adrenal glands of 11 tame and 11 aggressive foxes was sequenced on an Illumina HiSeq2500. After quality filtering, reads were aligned to the dog genome and read counts were used to assess gene expression levels. For both tissues, enrichment analysis for Gene Ontology terms in differentially express genes and weighted gene co-expression network analysis were performed.
Summary of Results: The gene expression and network analyses looking at the differences in anterior pituitaries of tame and aggressive foxes indicated the importance of genes related to the regulation of exocytosis, specifically mediated by cAMP, the organization of pseudopodia, and cell motility. In adrenals, differential gene expression analysis suggested differences in ectodermal cell differentiation and interleukin-8 production, while weighted gene co-expression network analysis found differences in cholesterol biosynthesis and cell migration, suggesting that the biosynthesis of cholesterol (a steroid hormone precursor), rather than the biosynthesis of adrenal steroid hormones may differ between the two strains. Finding differentially expressed genes which are important for cell signaling and migration in both pituitary and adrenals suggests that tame and aggressive foxes may have differences in regulation of the hormone release. Although our findings are in the experimentally domesticated fox, the similarity of phenotypes across domesticated species suggest the possibility of selection on a shared set of gene pathways.

350T Distinct Genetic Architectures for Phenotype Means and Plasticities in *Zea mays*. *A. Kusmec*[1], S. Srinivasan[2,3], D. Nettleton[4], P. Schnable[1,2] 1) Department of Agronomy, Iowa State University, Ames, IA; 2) Plant Sciences Institute, Iowa State University, Ames, IA; 3) School of Computing and Electrical Engineering, IIT Mandi, Mandi, Himachal Pradesh, India; 4) Department of Statistics, Iowa State University, Ames, IA.

Phenotypic plasticity describes the phenotypic variation of a trait when a genotype is exposed to different environments. Understanding the genetic control of phenotypic plasticity in crops such as maize is of paramount importance for maintaining and increasing yields in a world experiencing climate change. Here, we report the results of genome-wide association analyses of multiple phenotypes and two measures of phenotypic plasticity in the maize nested association mapping (US-NAM) population grown in multiple environments and genotyped with ~2.5 million single nucleotide polymorphisms (SNPs). We show that across all traits the candidate genes for mean phenotype values and plasticity measures form structurally and functionally distinct groups. Such independent genetic control suggests that breeders will be able to select semi-independently for mean phenotype values and plasticity, thereby generating varieties with both high mean phenotype values and levels of plasticity that are appropriate for the target performance environments.

351T A novel sex-ratio distorter segregates at low frequency in a *D. miranda* population. *E. Landeen*, P. Chen, D. Bachtrog Integrative Biology, University California Berkeley, Berkeley, CA.

Though sex ratio distortion has been reported in numerous species, few sex ratio distorter systems are characterized at the molecular genetic level. We screened a population of *Drosophila miranda* from Gerle Creek, California (hereafter Gerle) for sex ratio distortion by crossing Gerle females with geographically distant males from Coquitlam, Canada. We find that ~8% of crosses show significantly female-biased sex ratio, with a median proportion female progeny of 76% in driving lines. These data suggest that an X-linked meiotic driver is segregating at low frequencies in *D. miranda*. We used next-generation Illumina sequencing to identify structural variants specific to the X chromosome of a driving line that are candidate regions for meiotic drive.

352T The genetic basis of natural variation in chemical communication. *D. Lee*[1], O. Panada[2], P. Reis-Rodrigues[2], M. Helf[2], F. Schroeder[2], E. Andersen[1] 1) Department of Molecular Biosciences, Northwestern University, Evanston, IL; 2) Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York.

From bacterial quorum sensing to human language, communication is essential for social interactions. Nematodes communicate through pheromones called ascarosides, which consist of a sugar ascarylose linked to diverse fatty acid-like side chains as well as derivatives of amino acids, folate, and other metabolites. These modular structures produce a unique and diverse chemical language. Nematodes release distinct combinations of ascarosides at specific concentrations, which can signal other nematodes to modulate a variety of biological processes, including developmental arrest, social and sexual behavior, olfactory learning, stress response, and longevity.
Intraspecific natural variation of ascaroside production has been reported previously, implying that a diversity in nematode languages exist. However, the genetic basis and molecular mechanisms underlying the observed diversity have been insufficiently explored. To study natural variation in the nematode chemical language, we profiled 16 excreted ascarosides from 110 divergent *Caenorhabditis elegans* wild strains using mass-spectrometric analysis. We

discovered significant natural variation in the production of ascarosides and investigated the genetic architectures of these traits through genome-wide association analyses. We identified seven unique quantitative trait loci (QTL) that explain variation in production of ten ascarosides, including three loci underlying multiple ascaroside production traits. Additionally, because the ratios of various ascarosides serve as chemical cues, we used the ratios of pairwise combinations of ascarosides and identified 66 additional QTL. Fine mapping of these QTL suggests genetic variants in lipid and amino acid metabolism, as well as other regulatory pathways, all of which could contribute to the diversity of a nematode language. In summary, our study will facilitate (1) exploring the landscape of pheromone diversity in *C. elegans*, (2) discovering novel regulatory pathways involved in the modular production of ascarosides, and (3) elucidating the genetic underpinnings and molecular mechanisms for how evolution has shaped regulatory pathways to produce diversity in this pheromone language.

353T Genetics of chromosome-X nondisjunction rate variation in *C. elegans*. *Tzitziki Lemus Vergara*[1,2], Leonid Kruglyak[1,2] 1) Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Howard Hughes Medical Institute.

When cells divide each daughter cell gets a chromosome set. However, the process is imperfect and occasionally chromosomes fail to separate, generating a nondisjunction event. Nondisjunction events can lead to aneuploidy and genetic instability, the former a hallmark of cancer in somatic cells. If nondisjunction occurs during gametogenesis, and the aneuploid gametes become part of a zygote, the resulting embryo has a high probability to abort, or to develop physical or mental abnormalities such as human chromosome-21 trisomy, and Turner syndrome, in which females have only one X chromosome. Studying nondisjunction can give us a better understanding of how nondisjunction events occur, and could allow us to predict and alleviate some of its consequences.

Our work aims to understand how chromosome nondisjunction is regulated, specifically, how genetic variation and stress influences the rates of chromosome-X nondisjunction. We leverage *C. elegans* sex determination system to study nondisjunction. *C. elegans* has two sexes, hermaphrodites and males. Hermaphrodites have two X chromosomes, whereas males have only one. Hermaphrodites are physically females; they produce eggs, sperm and can self-fertilize but cannot mate with each other. Males arise in the population by mating or by nondisjunction events during hermaphrodite gametogenesis. Nondisjunction can produce chromosome X-depleted gametes, which upon fertilization form single-X embryos that will give raise to males. Thus, starting with a hermaphrodite population, the percentage of males over one generation is a proxy of chromosome-X nondisjunction rates. We can also increase the proportion of male progeny by subjecting hermaphrodites to stress. Although this approach is commonly used to generate males in the laboratory, the specific mechanism of how this is achieved is unknown.

To map genetic variants affecting the rate of chromosome-X nondisjunction, we are using a panel of 359 Recombinant Inbred Lines (RILs). This RIL panel was built by mating the reference strain N2 with the CB4856 strain. N2 and CB4856 have different rates of chromosome-X nondisjunction. Thus far, we have phenotyped 55 strains. Preliminary analyses show that the trait is heritable, with genetic variation explaining 46% of the trait variance. An initial QTL mapping revealed that variants on chromosome III might influence chromosome-X nondisjunction.

354T Exploiting natural variation in yeast stress responses to connect gene expression variation to organismal traits. Tara Stuecker, Amanda Scholes, *Jeffrey Lewis* Biological Sciences, University of Arkansas, Fayetteville, AR.

Gene expression variation is extensive in nature, and is hypothesized to play a major role in shaping phenotypic diversity. However, connecting differences in gene expression across individuals to higher-order organismal traits is not trivial. In many cases, gene expression variation may be evolutionarily neutral, and in other cases expression variation may only affect phenotype under specific conditions. To understand connections between gene expression variation and stress defense phenotypes, we have been leveraging extensive natural variation in the gene expression response to acute ethanol in laboratory and wild *Saccharomyces cerevisiae* strains. Previous work found that the genetic architecture underlying these expression differences included dozens of "hotspot" loci that affected many transcripts in *trans*. We now provide new evidence that one of these expression QTL hotspot loci is responsible for natural variation in one particular stress defense phenotype—ethanol-induced cross protection against severe doses of H_2O_2. The causative polymorphism is in the heme-activated transcription factor Hap1p, which we find directly impacts cross protection, but not the basal H_2O_2 resistance of unstressed cells. This further underscores the idea that basal and acquired stress resistance are mechanstically distinct. The Hap1p-effect relies on novel regulation of cytosolic catalase T (Ctt1p) during ethanol stress in wild strains. Because ethanol accumulation precedes aerobic respiration and accompanying reactive oxygen species formation, wild strains with the ability to anticipate impending oxidative stress would likely be at an advantage. This study highlights how strategically chosen traits that better correlate with gene expression changes can improve our power to identify novel connections between gene expression variation and higher-order organismal phenotypes.

355T Losing sweetness: insights into reduced nectar in the selfing syndrome. *I. Liao*, M. Rausher Department of Biology, Duke University, Durham, NC.

A composite trait comprises several traits that appear to evolve together as a unit. One such composite trait is the selfing syndrome, which is a suite of traits associated with the transition from outcrossing to highly selfing. In flowering plants, these traits include reductions in flower size, pollen/ovule ratios, scent, and nectar. Of these, nectar has been overlooked, and how reduced nectar evolved remains unexamined. If no genetic correlations exist between nectar and other traits, reduced nectar may have evolved independently. At the other extreme, if genetic correlations exist between nectar and other traits, then reduced nectar may have evolved as a correlated response to selection on those traits. Additionally, nectar is also a composite trait composed of nectar volume, nectar sugar concentration, and nectary (nectar-producing tissue) size. We are examining how reduced nectar evolved in two sister morning glory species – *Ipomoea cordatotriloba* and *I. lacunosa*. *Ipomoea lacunosa* is highly selfing and displays typical selfing-syndrome traits compared to the mixed-mating *I. cordatotriloba*. We have conducted a pilot quantitative trait locus (QTL) study examining only nectar as a composite trait. Using 130 F3s generated from a *I. cordatotriloba* X *I. lacunosa* cross, we found that each nectar trait is polygenic. Overlapping QTLs were found between nectary size and nectar volume and between nectary size and nectar sugar concentration. This implies that reduction in nectary size can indirectly lead to reductions in volume and sugar concentration. We are currently expanding the study to 300 F5s and including other syndrome traits (e.g. flower size, pollen amount). Preliminary results indicate that nectar traits are correlated with flower size in F5s. This suggests that traits with similar functions are integrated as a unit and implies that selection on one trait may facilitate reductions in other traits in the selfing syndrome.

356T The effects of quantitative trait architecture on detection power in artificial selection experiments. *R. Nicolas Lou*[1], Nina Therkildsen[1], Philipp Messer[2] 1) Department of Natural Resources, Cornell University, Ithaca, NY; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Evolve and resequence (E&R) experiments, in which artificial selection is performed on organisms in a controlled environment, are becoming an increasingly accessible tool for understanding the genetic basis of adaptation. Previous work has assessed how different experimental design parameters affect the power of E&R studies, but so far there has been little exploration of how the underlying architecture of polygenic traits influences our ability to detect selection in these types of experiments. In this study, we use forward simulation to build a realistic model of an E&R experiment in which a quantitative, polygenic trait is under truncation selection. We test how the power for quantitative trait loci (QTL) detection can be influenced by different aspects of trait architecture, including the number of QTLs, their starting frequencies, effect sizes, clustering along a chromosome, dominance, and epistasis. We show that all of the above parameters affect how polygenic traits respond to selection and, consequently, our power to detect the underlying QTLs. In addition, we demonstrate that existing detection methods based on a model of independent selective sweeps at the individual QTLs typically have equal or less power than a simple measurement of allele frequency differences at these loci before and after selection. Our findings highlight the importance of taking trait architecture into consideration in designing studies of molecular adaptation with temporal data. Furthermore, the customizable modeling framework we provide will enable researchers to easily simulate experimental or natural populations under selection, with different trait architectures and parameters tuned to their specific setups, allowing them to assess detection power and optimize experimental or sampling designs.

357T Sex-specific mechanisms of paraquat susceptibility in *D. melanogaster*. *P. Lovejoy*, A. Fiumera Department of Biological Sciences, Binghamton University, Binghamton, NY.

Gene by environment interactions occur when various genotypes respond differently to the same environments. Interestingly, members of each sex may respond differently to those environments, even if they share the same genotype. Sex-specific genetic mechanisms are consistently found in traits that aren't obviously affected by sex. For example, in humans, sex-specific progressions of diseases such as cardiovascular disease, depression, rheumatoid arthritis, and type II diabetes have recently been uncovered. Sex-specific genetic effects have also been found in model organisms, but less work has been done to investigate how gene by environment interactions may be different between the sexes. To investigate the interactions between genetics, environment, and sex, the present study investigated paraquat susceptibility in *D. melanogaster*. Paraquat is a commonly applied herbicide which causes oxidative stress of the mitochondria, death of dopaminergic neurons, and motor deficiencies in non-target organisms. First, the naturally occurring genetic variation for paraquat susceptibility was mapped by assaying 100 of the DGRP lines for climbing ability under control and paraquat conditions. With these data, a GWAS was performed that identified 35 genes in males and 17 genes in females that are associated with paraquat susceptibility. Only two associated genes overlap between males and females, despite the significant positive correlation between male and female phenotypes. Forty-seven of all associated SNPs had significant SNP by sex interactions, indicating that these specific SNPs had conditionally neutral effects on paraquat susceptibility, depending on sex. Each SNP with a significant interaction had a stronger effect on susceptibility in males than females. Gene by environment interactions can be different depending on sex. In addition to finding sex specific effects, this study identified several genes not previously known to be involved in paraquat susceptibility. A subset of candidate genes has been verified

using RNAi and candidate gene expression was measured in the DGRP lines to identify mechanisms that may be contributing to differences in susceptibility between the lines.

358T TRiAGE: technique for ranking genes in epistasis. *J. Matthew Mahoney*[1], Anna Tyler[2] 1) Neurological Sciences, University of Vermont, Larner College of Medicine, Burlington, VT; 2) The Jackson Labroatory.

Multilocus statistical models are becoming popular in the analysis of complex traits due to their ability to identify genetic interactions, or epistasis, among alleles. These studies are limited, however, because as with single-locus trait mapping, genetic resolution of QTLs in multilocus models is typically insufficient to identify causal variants. Moreover, the combinatorial nature of multilocus models expands the pool of potential causal variants far beyond that of single-locus models. We reason, however, that epistatic interactions inherently contain more information than single-locus associations, and that this information can be harnessed to generate specific hypotheses about causal variants in interacting QTLs. In particular, a statistical interaction between QTLs implies a functional interaction between variants encoded by the interacting loci. The pool of potential causal variants in the two loci is therefore limited to those that functionally interact to influence the phenotype of interest. In recent years there has been a concerted systems biology effort to predict functional genomic networks of gene-gene interactions across species, tissues, and cell types, as well as to systematically tabulate gene-phenotype associations. Following these efforts, we developed a machine learning strategy called Technique for RAnking Genes in Epistasis (TRiAGE) that integrates functional genomic networks with epistasis to prioritize candidate gene pairs responsible for the observed epistatic interaction. TRiAGE uses gene-gene interaction weights from functional genomic networks and known gene-phenotype associations to construct a novel feature representation of phenotype-associated gene-gene interactions. TRiAGE uses these novel features to train a support vector machine (SVM) classifier to recognize known phenotype-associated gene interactions. It then classifies all putative gene-gene interactions spanning epistatic QTLs. We have performed two proof-of-concept analyses using TRiAGE to predict modifier alleles of seizure severity, first in mouse models of absence epilepsy (AE), and second in chemical induction models of seizure susceptibility (SZS) in the BXD recombinant inbred lines. In AE, TRiAGE predicts that an interaction between a known seizure gene, *Plcb1*, and a transcription factor regulating myelination, *Tenm4*, is responsible for an epistatic interaction between QTLs on Chrs 2 and 7. In SZS, TRiAGE predicts that a functional interaction between a known SZS gene, *Kcnj9*, and another myelin regulating transcription factor, *Myt1l*, is responsible for an epistatic interaction between a QTLs on Chrs 1 and 12. In both cases our predictions are highly plausible candidates, indicating that TRiAGE holds promise to nominate quality candidates for epistatic QTLs on a significantly shorter timescale than previously possible.

359T Genetic variation interacts with experience to determine interindividual differences in learned song. *D.G. Mets*[1,2], M.S. Brainard[1,2] 1) Physiology, Univ California, San Francisco, San Francisco, CA; 2) Howard Hughes Medical Institute, San Francisco, CA.

Learning reflects the influence of experience on genetically determined circuitry, but little is known about how experience and genetics interact to determine complex learned phenotypes. Here, we used vocal learning in songbirds to study how experience and genetics contribute to interindividual differences in learned song. Previous work has established that such differences in song within a species depend on learning, but in principle some of these differences could also depend on genetic variation. We focused on song tempo, a learned and quantifiable feature that is controlled by central neural circuitry. To identify genetic contributions to tempo we computer-tutored juvenile Bengalese finches (*Lonchura striata domestica*) from different genetic backgrounds with synthetic songs in which tempo was systematically varied. Computer-tutored birds exhibited unexpectedly strong heritability for song tempo and comparatively weak influence of experience. We then tested whether heritability was fixed and independent of experience by providing a second group of birds with enriched instruction via live social tutoring. Live tutoring resulted in not only a significant increase in the influence of experience on tempo but also a dramatic decrease in the influence of genetics, indicating that enriched instruction could overcome genetic biases evident under computer tutoring. Our results reveal strong heritable genetic contributions to interindividual variation in song tempo but that the degree of heritability depends profoundly on the quality of instruction. They suggest that for more complex learned phenotypes, where it can be difficult to identify and control relevant experiential variables, heritability may similarly be contingent on the specifics of experience.

360T Digging for genes involved in behavioral evolution. *O. Meyerson*[1,2,3,4,5], H. Metz[1,2,3,4,5], C. Hu[1,2,3,4,5], H. Hoekstra[1,2,3,4,5] 1) Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA; 3) Center for Brain Science, Harvard University, Cambridge, MA; 4) Museum of Comparative Zoology, Harvard University, Cambridge, MA; 5) Howard Hughes Medical Institute, Harvard University, Cambridge, MA.

While we have made much progress in identifying genes contributing to morphological variation, we know far less about the genetic basis of behavioral evolution. Burrowing in deer mice is a compelling system for studying the genetic basis of a natural behavior. Two recently diverged sister species – *Peromyscus polionotus* and *P. maniculatus* – exhibit distinct and heritable burrowing behaviors. *P. polionotus* dig burrows with a long entrance tunnel and a secondary structure that serves as

an escape tunnel from predators, whereas *P. maniculatus* dig burrows with a short entrance and no escape tunnel. *P. polionotus* and *P. maniculatus* are interfertile, and F_1 hybrids dig burrows resembling those of *P. polionotus*. Previous work identified three genomic regions associated with burrow entrance-tunnel length, and one region associated with the presence of an escape tunnel. Here, we further refine the genetic architecture of deer mice burrowing using an advanced backcross design. First, we selected hybrid individuals producing *polionotus*-like burrows and repeatedly backcrossed them to *P. maniculatus*. We then generated a high resolution QTL map by implementing multiplexed shotgun genotyping and using automated image analysis to quantify burrow morphology. This approach revealed QTL regions that replicated the initial mapping, as well as new QTL peaks associated with burrow size and shape. We further investigated the effect of the strongest QTL peak on chromosome 1 by generating a congenic strain through isolating the *polionotus* genotype at this locus on an otherwise *maniculatus* background, and testing its effect on burrowing behavior. Finally, we assayed gene expression of these candidate loci in the brains of both species as well as F_1 hybrids. Together, this study narrows in on a set of exciting candidate genes that contribute to differences in deer mouse burrowing and exemplifies how evolution acts upon the genome to orchestrate the emergence of a complex behavioral trait.

361T Observing viability selection in contemporary humans. *H. Mostafavi*, M. Przeworski Biological Sciences, Columbia University, New York, NY.

A number of open questions in human evolutionary genetics would become tractable if we were able to directly measure evolutionary fitness. The recent availability of large biomedical datasets provides that opportunity. With this in mind, we recently developed a method to examine whether individual genetic variants, or sets of genetic variants, currently influence viability (Mostafavi et al., PLOS Biology 2017). The approach consists in testing whether the frequency of an allele varies across ages, accounting for variation in ancestry. Applying it to the GERA data set, and by proxy, to parents of the UK Biobank participants, we detected signals for variants of the APOE and CHRNA3/5 genes. We also found evidence that variants that contribute to several quantitative traits influence lifespan, notably age at puberty and and age at first child birth. We now analyze the UK Biobank full cohort, also performing survival analysis. We discover several new variants to be associated with father or mother lifespan, including variants near LPA, CHRNA2 and CHRNA4 among others, almost all in loci containing genes previously shown to be associated with deleterious conditions such as high cholesterol levels, risk of heart disease and lung cancer. Intriguingly, despite having considerable power, only a few common variants with large effects at old ages are observed (mostly within APOE), which suggests that late-acting deleterious variants are kept at low frequency by purifying selection, either through inclusive fitness effects, or because they have subtler effects earlier in life (i.e., mutational target size for variants with deleterious effects only at late life stages is very small). Furthermore, across the genome, we find marked differences between males and females for effects on lifespan. More generally, this approach opens the door for investigating possible trade-offs among sexes and age groups, as well between effects on viability and other components of fitness.

362T Genomic rearrangements considered as quantitative traits. *Richard Mott*[1], Martha Imprialou[2], Paula Kover[3] 1) Genetics Institute, University College London, London, UK; 2) Wellcome Trust Centre for Human Genetics, Oxford UK; 3) University of Bath, UK.

To understand the population genetics of structural variants and their effects on phenotypes, we developed an approach to mapping structural variants that segregate in a population sequenced at low coverage. We avoid calling structural variants directly. Instead, the evidence for a potential structural variant at a locus is indicated by variation in the counts of short-reads that map anomalously to that locus. These structural variant traits are treated as quantitative traits and mapped genetically, analogously to a gene expression study. Association between a structural variant trait at one locus, and genotypes at a distant locus indicate the origin and target of a transposition. Using ultra-low-coverage (0.33) population sequence data from 488 recombinant inbred *Arabidopsis thaliana* genomes, we identified 6502 segregating structural variants. Remarkably, 25% of these were transpositions. While many structural variants cannot be delineated precisely, we validated 83% of 44 predicted transposition breakpoints by polymerase chain reaction. We show that specific structural variants may be causative for quantitative trait loci for germination and resistance to infection by the fungus *Albugo laibachii*, isolate Nc14. Further we show that the phenotypic heritability attributable to read-mapping anomalies differs from, and, in the case of time to germination and bolting, exceeds that due to standard genetic variation. Genes within structural variants are also more likely to be silenced or dysregulated. This approach complements the prevalent strategy of structural variant discovery in fewer individuals sequenced at high coverage. It is generally applicable to large populations sequenced at low- coverage, and is particularly suited to mapping transpositions.

363T Environmental clustering strategies to exploit genotype-by-environment interaction for genomewide selection. *Jeffrey Neyhart*, Kevin Smith Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN.

Genotype-by-environment interactions (GEI) are a common challenge for plant breeders working on quantitative traits. Such interactions complicate efforts to develop broadly superior genotypes, but may be reduced or exploited to target genotypes for specific environments by clustering similar environments together. Genomewide selection, or the prediction of the breeding value of new genotypes based on genomewide markers, is a valuable tool for plant breeders and may be extended

to account for GEI and improve predictions. Of special interest for breeding is the prediction of new genotypes in future environments, which has not yet been thoroughly explored. We investigated this problem in barley (*Hordeum vulgare* L.) by gathering phenotypic data for three quantitative traits on 233 individuals grown in more than 40 location-year environments. Genomewide predictions were made using data on 183 individuals as a training population and predictions were validated using data on the remaining 50 individuals. We explored environmental clustering methods that used observed phenotypic data (e.g. environmental distance *D*, principal components, and factor analysis) or environmental covariables such a temperature, rainfall, and soil characteristics. For each of the clustering methods, we simulated the introduction of future environments by masking phenotypic values from that environment and predicting the values using data from environments within the cluster. Generally, clustering methods based on observed phenotypes were more effective at reducing GEI within the cluster and led to more accurate predictions than clustering based on environmental covariables. Nevertheless, the use of environmental covariables permits the prediction of future environments, and although predictions were less accurate using this method, it may still be an effective strategy to select superior genotypes in future growing conditions.

364T Transcriptional response of a multiparent population to adult nutritional change. *E. Ng'oma*, A. Rahman, E.G. King Division of Biological Sciences, University of Missouri, Columbia, MO.

Nutritional quality plays a major role in how resources are allocated to biological structures and function in many organisms. The constraints exerted by food availability in turn dictate the maximal value that fitness components such as reproduction and lifespan can attain. The molecular mechanisms of such plastic responses to changes in diet conditions are yet to be understood in most circumstances. In this study, we compare whole-genome transcript abundances in three tissues (heads, ovaries and bodies) of mated adult female *Drosophila melanogaster* sampled from an outbred multipatental population reared on three diet conditions (high sugar, low protein and control). We relate profiles of gene expression and differential regulation to patterns of resource allocation predicted by evolutionary theory.

365T Coevolved nucleo-mitochondrial interactions and slow growth contribute to mitochondrial genome stability. *T. Nguyen*, H. Fiumera Binghamton University, Binghamton, NY.

Mitochondrial respiration produces nearly all cellular energy in eukaryotes and underlies virtually all eukaryotic ecological and evolutionary phenomena. Given that mitochondrial DNA (mtDNA) plays an essential role for mitochondrial energy production, mtDNA loss leading to respiratory deficiency occurs at a surprisingly high frequency in cells. Damage in mtDNA reduces cell fitness, indicating possible selection for effective mtDNA maintenance in natural populations. The genetic and environmental factors contributing to mtDNA stability and mitochondrial fitness are fundamental to traits with adaptive and speciation potential but are not well understood. We examined rates of mtDNA loss in natural isolates of *Saccharomyces cerevisiae* by measuring the frequency that respiratory-deficient, *petite* colonies arose from replicating cells. We compared *petite* frequencies in strains harboring unique nucleo-mitochondrial (n-mt) combinations and found that mtDNA stability was influenced by both nuclear and mitochondrial genetic variants, and n-mt interactions, in population-specific ways. Strains with coadapted n-mt combinations conferred lower rates of mtDNA loss than those with synthetic genome combinations, indicating that coevolution between n-mt genomic interactions likely occurs to mitigate mtDNA loss. Interestingly, we observed positive correlation between growth rates, a proxy measure for mitochondrial fitness, and mtDNA loss in strains with original n-mt combinations. By growing cells in low-glucose media, reduced growth rates were accompanied with decreased mtDNA loss. Under temperature stress, coadapted n-mt interactions were reported to provide growth advantages over synthetic genome combinations, although the rate of mtDNA loss was exacerbated by the same condition. This suggests that there may be a conflicting tradeoff between genetic interactions that promote growth under stress conditions with the ones that maintain mtDNA integrity. Currently, we are seeking to identify the exact nuclear variants that contribute to mtDNA stability.

366T Networks underpinning symbiosis revealed through cross-species eQTL mapping. *Dahlia Nielsen*[1], Yuelong Guo[2], David Bird[3], Valerie Williamson[4] 1) Biological Sciences, North Carolina State Univ, Raleigh, NC; 2) RTI International, RTP, NC; 3) Plant Pathology, North Carolina State University, Raleigh, NC; 4) Plant Pathology, University of California, Davis, CA.

Interactions between species are pervasive among plants, animals, and microbes, yet the underlying molecular actors are known for only a few interactions. Many techniques have been designed to uncover genes involved in signaling between organisms. Typically, these focus on only one of the partners involved. We developed an expression quantitative trait locus (eQTL) mapping-based approach to identify cause-and-effect relationships between genes from two partners engaged in an interspecific interaction. We leverage a model plant-parasite system to demonstrate the efficacy of this approach. Gene expression measurements for ninety-eight isogenic plants (*Medicago truncatula*), each inoculated with a genetically distinct line of the diploid parasitic nematode *Meloidogyne hapla* were assayed. By sampling infected tissue from each host plant, systematic differences in gene expression across plants could be mapped to genetic polymorphisms of their infecting parasites. The effects of parasite genotypes on plant gene expression were often substantial, with up to 90-fold ($p=3.2x10^{-52}$) changes in expression levels caused by individual parasite loci. Mapped loci included a number of pleiotropic sites, including one 87 kb parasite locus that modulated expression of more than sixty host genes. The 213 host genes identified were

substantially enriched for transcription factors. We distilled higher-order connections between polymorphisms and genes from both species via network inference. To replicate our results and test whether effects were conserved across a broader host range, we performed confirmatory experiments using *M. hapla*-infected tomato and soybean plants. These experiments revealed that host responses are conserved across broad evolutionary distances. Our study demonstrates the efficacy of cross-species eQTL mapping in connecting genetic variation in one organism to gene expression responses in an interacting organism. The power of this approach is its ability to concurrently identify sets of hosts and pathogen genes, rather than focusing on one side of the interspecific dialogue.

367T Developing environmental region-specific genomic predictions for beef cattle breeds. *S. Nilson*[1], T. Rowan[1], C. Seabury[2], R. Schnabel[1], J. Decker[1] 1) Division of Animal Science, University of Missouri, Columbia, MO; 2) Department of Veterinary Pathobiology, Texas A&M University, College Station, TX.

Genotype-by-environment interactions play a significant role in the lifetime productivity of an animal. We hypothesize that including loci responsible for environmental adaptation in prediction models will increase the accuracy and speed of region-specific genomic selection. To develop environment specific genomic predictions, 30-year normals for temperature, precipitation, and elevation were analyzed with K-means clustering to divide the United States into 9 distinct regional environments. Over 12,000 Gelbvieh, 10,000 Simmental, and 20,000 Red Angus cattle distributed across the majority of the environmental regions were included in the analyses. All genotypes were phased with Eagle v2.4 and imputed with IMPUTE2 v2.3.2 up to ~850,000 SNPs. Breed and region-specific Bayesian Sparse Linear Mixed Models were fit in GEMMA v0.94 for 3 phenotypes: birth weight, weaning weight, and yearling weight. Regions with less than 300 samples were removed from the analyses. A separate model was fit for each region, thus for example, 18 models (3 traits by 6 regions) were fit for the Gelbvieh breed. Accuracy of the models were assessed with a three-fold cross-validation and by calculating the square of the correlation between the phenotype and the predicted breeding value. A predicted breeding value was calculated for every animal for all three phenotypes across the environmental regions to investigate reranking across environments. These prediction models will serve as a valuable tool for producers to identify how well an animal will perform in their environment before making breeding or purchasing decisions.

368M Gene, environment and cellular interactions underlying behavioral variance and their relation to fitness in a long-term *Caenorhabditis elegans* evolution experiment. *L.M. Noble*[1], T. Guzella[2], F.M. Mallard[2], M.V. Rockman[1], H. Teotónio[2] 1) Center for Genomics and Systems Biology, New York University, New York, NY; 2) Institut de Biologie, École Normale Supérieure.

By evolving *C. elegans* from multiparental standing genetic variation we have generated a sequenced panel of recombinant inbred lines with which to study the properties and evolution of genetic architectures. We measured fitness, precisely defined under the experimental regime, and morphological and behavioral traits that vary in their alignment with fitness, in more than 400 lines, under familiar and novel conditions. Additive architectures for fitness and closely related traits are extremely polygenic, as expected, but strong sign epistasis with weak marginal effects also accounts for a large fraction of trait variance.

The converse is true for behavioral traits, which show weaker (phenotypic and additive genetic) correlations with fitness. And, consistent with results from other systems, additive effects are relatively consistent across environments while epistatic interactions are much less so. Using whole animal single cell gene expression data we see that the inferred cellular basis of behavioral variance is strongly dependent on environment, evolutionary history, and fitness alignment. The expression of interacting loci underlying variance in behavior and fitness-related traits also differs markedly, consistent with variable pleiotropy.

369T Phenotypic characterization of gigantism in Gough Island mice. *M.J. Nolte*[1], C.N. Dewey[2], B.A. Payseur[1] 1) Laboratory of Genetics, University of Wisconsin-Madison, WI; 2) The Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, WI.

Island-colonizing organisms often evolve gigantism or dwarfism, raising the question of whether common developmental and genetic mechanisms are involved. We use house mice (*Mus musculus domesticus*) from isolated Gough Island as a model system for understanding the evolution of body size on islands. Mice from Gough Island weigh twice as much as their mainland relatives, despite their recent arrival on the island a few hundred generations ago. Here, we report a high-resolution comparison of growth between inbred strains from Gough Island and the mainland. By collecting daily body weights, we discovered that divergence in body size between island and mainland mice begins around two weeks post-birth, a time period associated with a transition from suckling to solid food in classical inbred strains. RNA-Seq data revealed a substantial increase in the number of genes that are differentially expressed between island and mainland mice in the liver at this age. Analyses of serum metabolites and weights of metabolic organs pointed to changes in tissue development and function that underpin gigantism in Gough Island mice. This progress positions us to evaluate the genetic and phenotypic effects of individual quantitative trait loci we previously tied to the evolution of extreme body size.

370T Effect of Transposable Element Copy Number on Maize Phenotypes. *E. Osborn*, J. Ross-Ibarra, M.C. Stitzer University of California, Davis , Napa, CA.

Transposable elements (TEs) are sequences that can increase in copy number in the genome when they jump to new positions. TEs are universally present in plant genomes and, though generally thought to be junk DNA, can in some cases affect the expression of genes important for plant fitness. In this study, we use a well- characterized TE insertion in maize to investigate whether variation in TE copy number can impact gene expression. A TE insertion 65 kb upstream of the *teosinte branched 1 (tb1)* increases the gene's expression relative to maize's wild ancestor, teosinte. This TE copy from the Hopscotch family is found in all modern maize and is largely responsible for branching differences between teosinte and maize. We identified additional copies from the Hopscotch family throughout the genomes of other maize lines using whole genome sequencing data, identifying unique junctions between flanking and TE sequence originating when the copy jumps. Preliminary results show that the 30 analyzed inbred lines each contain from 1 to 3 copies of Hopscotch, that copy number is positively correlated with *tb1* expression, and that maize lines with basal branching in some conditions have lower Hopscotch copy number. Because expression of *tb1* also acts through inflorescence phenotypes, like tassel and ear length, we estimate copy number for 4,500 progeny of a mapping population to identify whether Hopscotch copy number affects phenotypes. Together, these results suggest a role for variation in TE copy number in modulating gene expression of *tb1* and resulting phenotypes.

371T Whole Genome Resequencing to identify QTLs for ascites in chickens. *A. Parveen*[1], S. Dey[1], C. Jackson[2], N. Anthony[3], D. Rhoads[1] 1) Biological Sciences, University of Arkansas, Fayetteville, AR; 2) Biology, John Brown University, Siloam Springs, AR; 3) Poultry Sciences, University of Arkansas, Fayetteville, AR.

We are using whole genome resequencing to identify chromosomal regions associated with resistance or susceptibility to ascites, a form of pulmonary hypertension syndrome, meat-type chickens. Previous Genome Wide Association Studies (GWAS) based on Single Nucleotide Polymorphisms (SNPs) have identified regions on chromosomes 2, 9 and Z. Despite several GWAS and further genotyping, there are no reliable or potential markers for ascites phenotype. We have completed screening of Copy Number Variations (CNVs) and Single Nucleotide Polymorphisms in ascites resistant and susceptible birds from the relaxed, REL, line derived from a commercial elite broiler line. DNA samples from resistant and susceptible birds were purified, quantified and pooled in two pools of 10 DNAs from each phenotype for both genders. Eight pools (2 pools x 2 phenotypes x 2 genders) were generated. Each pool was submitted for bar-coded library generation, and 2x125 paired end reads on Illumina HiSeq 2500 and with 66X genome coverage. The sequence reads were mapped onto Galgal5 using Bowtie for initial CNV mapping cn.mops (R package). Further mapping to chromosomes were done using NGen and ArrayStar (DNAStar ver 13). So far we have identified two potential regions for CNVs and 31 regions for SNPs with potential association with ascites phenotype. CPQ gene on chromosome 2 and LRRTM4 gene on chromosome 22 have been validated for containing ascites QTLs. However, their exact role in ascites is yet to be discovered. Further, we will validate all regions using high-throughput SNP assays in a larger catalog of DNAs from additional commercial broilers.

372T Sexual dimorphism and evolution of crossover patterning in house mice. *A. Peterson*[1], D. Jacobson[1], N. Miller[2], B. Payseur[1] 1) Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; 2) Department of Botany, University of Wisconsin-Madison, Madison, WI.

Homologous recombination shapes genetic diversity and ensures the proper segregation of chromosomes during meiosis. Despite functional constraints on recombination, females and males recombine at different rates in many species. At a genome-wide level, both the number and position of crossovers exhibit sexual dimorphism. In several species, crossover number is greater in females and crossover position is biased to chromosome ends in males. A compelling but unaddressed evolutionary question is whether sexual dimorphism in crossover number (heterochiasmy) and dimorphism in crossover patterning co-evolve. The house mouse, *Mus musculus,* provides a powerful system to answer this question because inbred strains enable comparisons between males and females with highly similar genomes. To determine how sexual dimorphism evolved between two recently diverged subspecies of mice, we used immunofluorescent cytology to visualize recombination proteins in individual oocytes and spermatocytes. To characterize crossover location, we developed an automated image analysis program that estimates the positions of foci along thousands of individual chromosomes. We confirmed that inbred strains of *M. m. musculus* show a male-specific 30% increase in crossover number compared to *M. m. domesticus* strains. We propose that changes in the synaptonemal complex – the proteinaceous structure that tethers homologous chromosomes during meiosis – played an important role in this evolutionary increase. At the same time, females and males display differences in relative crossover position that are similar in the two subspecies. These results indicate that crossover number dimorphism and crossover position dimorphism evolved independently in house mice.

373T Ohtadstat: A new framework for the application of Ohta's D statistics to genome-wide datasets. *P.F. Petrowski*, E.G. King, T.M. Beissinger Division of Biological Sciences, University of Missouri, Columbia, MO.

Patterns of linkage disequilibrium (LD) across the genome may reveal loci that have been under selection. For example,

natural selection can influence LD by increasing levels of LD near the selected site, or via epistatic selection acting on multiple interacting loci. In 1982, Tomoka Ohta capitalized on this tendency by deriving a set of statistics which compare levels of linkage disequilibrium in subpopulations relative to that of the total population. She posited that loci which deviate from expected levels of LD in a subpopulation may be under selection in that subpopulation. This is a powerful approach, but to date no framework exists to support its use on a widespread basis. We present ohtadstat (available on GitHub at https://github.com/pfpetrowski/OhtaDStats), an R package designed to facilitate the implementation of Ohta's D statistics in a variety of use cases. Here, we demonstrate this package's ability to estimate patterns of linkage disequilibrium in structured populations in order to detect signals of selection on two publicly available genomic datasets. Additionally, we discuss the three functions available in the package and show how they may be adapted for use in any experimental setup for which genotype data is available across two or more subpopulations.

374T Distinct distributions of mutational effects on fitness revealed by high-throughput phenotyping. *Y. Plavskin*[1,2], M.S. de Biase[3], N. Ziv[1,2], Y.O. Zhu[4], B. Goulet[1,2], D.W. Hall[5], R.F. Schwarz[3], D.A. Petrov[4], D. Tranchina[2,6], M.L. Siegal[1,2] 1) Center for Genomics and Systems Biology, New York University, New York, NY; 2) Department of Biology, New York University, New York, NY; 3) Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany; 4) Department of Biology, Stanford University, Stanford, CA; 5) Department of Genetics, University of Georgia, Athens, GA ; 6) Courant Institute for Mathematical Sciences, New York University, New York, NY.

Spontaneous mutations provide the raw material on which selection acts. To shed light on the spectrum of mutations that constitute the starting point of selection and their typical effects on fitness, we estimated the distribution of effect sizes of spontaneous mutations affecting growth rate, a phenotype closely related to fitness, in the budding yeast *Saccharomyces cerevisiae*. To capture a snapshot of the effects of mutations that have not yet been filtered by selection, we studied mutational effects in a collection of >80 Mutation Accumulation (MA) strains. Previous work had demonstrated that these strains contain ~300 mutations, primarily single-nucleotide mutations (SNMs); however, conventional sequencing analysis fails to identify mutations in DNA repeats, including simple sequence repeats (SSRs). By combining genotype information with high-precision microcolony-based growth rate measurements from each strain, we are able to infer estimates for the mean mutational effect size, the shape of the mutational effect distribution, and the proportions of mutations with a neutral, positive, and negative effect on growth rate. Interestingly, a single distribution of mutational effects cannot account for the observed distribution of growth phenotypes. Rather, growth rate effects of mutations fall into two categories: single-nucleotide substitutions, which on average have larger effects, and an additional set of frequent, smaller-effect mutations. The high estimated frequency of mutations in the latter category, combined with the fact that they were undetected by conventional sequencing, point to SSR mutations as likely candidates underlying the non-SNM mutational effects we observe. Follow-up experiments confirm that MA strains lacking SNMs indeed contain genetic variation contributing to growth rate differences, and we are currently further probing the properties of these non-SNM spontaneous mutations using sequencing and high-throughput growth measurements. Our work reveals the spectrum of effect sizes of the mutations on which evolution acts, and points to a likely key role of SSRs in shaping natural diversity.

375T Trait correlations and environmental interactions to inform the domestication of *Silphium integrifolium*. *J. H. Price*, K. P. Smith Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, MN.

The wild sunflower relative *Silphium integrifolium* is a candidate for domestication as a new perennial oilseed crop. Beyond food quality oil, silphium offers a number of other benefits, including drought tolerance and pollinator support. In order to maintain these qualities, and to improve silphium as a crop, it is necessary to understand the variation for key traits in available germplasm. To address this question, we have established the silphium domestication panel, a population of 380 genotypes clonally propagated and planted in field trials in three locations. These genotypes have been improved relative to wild populations though several cycles of selection. Here, we present measured ranges, correlations, and estimated heritabilities for a number of traits. For most traits, we found that a moving average method to correct spatial variation in the field generally increased estimates of heritability. Flowering time and stalk color were the most heritable traits, while lodging and diameter of the top of the stalk were among the least heritable. Strong correlations were found among height and stalk diameter measures, while flowering time had a weaker correlation with height than expected (r = 0.15). Additionally, leaf thickness was not generally correlated with other traits related to plant size. We also present information about GxE interactions. Finally, we discuss the implications and uses of this information for breeding.

376T A phylogenetic analysis of the *Drosophila* metabolome. *D. Promislow*[1], J. Hoffman[2] 1) Department of Pathology and Department of Biology, University of Washington, Seattle, WA; 2) Department of Biology, University of Alabama at Birmingham, Birmingham, AB.

In recent years, researchers have begun to incorporate measures of 'omic' domains (metabolome, transcriptome, etc.) in attempts to build more accurate genotype-phenotype maps, capturing at least some of the missing heritability typical of GWA studies. To bridge the gap between genotype and phenotype, we have focused in particular on the metabolome. The metabolome measures the hundreds of small molecules that make up the structural and functional building blocks of all

organisms. Surprisingly little is known about how this complex high-dimensional 'trait' evolves over time. We carried out a comparative analysis of genetic and phylogenetic diversity in the metabolome across 50 million years of evolution in the genus *Drosophila*. We measured both targeted and global metabolomic profiles using mass spectrometry. Comparative analysis of such data is challenging, given the very large number of variables. We present several approaches that attempt to address the statistical challenges of combining phylogenetic comparative analysis and high-dimensional data. We find that the metabolome has a strong phylogenetic signature, and identify some of the first evidence for metabolome gain/loss over evolutionary time. We also identify metabolites with evolutionary conserved patterns of change between between sexes and across ages. In addition to presenting our findings, we will discuss the challenges, opportunities, and some possible solutions in bringing a phylogenetic perspective to high-dimensional physiologically relevant data.

377T High-resolution mapping of candidate loci for social behavior in the red fox. *H. Rando*[1,2], J. Johnson[2], G. Zhang[3,4,5], L. Trut[6], A. Kukekova[2] 1) Illinois Informatics Institute, University of Illinois at Urbana-Champaign, Urbana, IL; 2) Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL; 3) China National Genebank, BGI-Shenzhen, Shenzhen, Guangdong, China; 4) Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark; 5) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; 6) Institute of Cytology and Genetics of the Russian Academy of Sciences, Novosibirsk, Russia.

Statement of Purpose: Sociality is fundamental to behaviors such as group living, parental care, and mate choice and is atypical in many human psychiatric and neurodevelopmental disorders. Though insights can be gained by uncovering the genetic architectures of social behaviors, mapping quantitative traits in wild populations or humans is very challenging. A selective breeding experiment at the Russian Institute for Cytology and Genetics presents a novel model for the genetics of social behavior. Fifty generations of breeding has produced two strains of red fox (*Vulpes vulpes*): a hypersocial "tame" strain and a defensively "aggressive" strain. Though quantitative trait loci (QTL) mapping has previously revealed nine fox behavioral QTL, the resolution was too low to identify candidate genes. The recent assembly of the red fox reference genome presents an opportunity to fine-map these behavioral loci.

Methods: The genomes of 10 tame, 10 aggressive, and 10 conventional farm-bred foxes were each sequenced at 2.5x. Sequences were aligned to the reference genome and pooled by population for SNP calling. SNPs were analyzed within sliding windows for within-population diversity (pooled heterozygosity, H_p) and pairwise between-population divergence (FST). Simulations were conducted to discriminate selective sweeps from drift. Genotypes could not be assigned to individuals due to the low coverage. An independent set of foxes comprising the 20 least-related individuals from each of the tame and aggressive strains were then sequenced with deep coverage (15x each). SNPs were called for each individual and analyzed to identify putative regions under selection within or between populations. The two analyses were compared to identify regions highlighted by both datasets.

Results: In the pooled data, 103 genomic regions were identified as putatively under selection based on H_p and/or FST. Thirty of these regions overlapped 5 previously identified fox QTL, supporting their inclusion as putative candidates for the regulation of behavior. Regions identified by the two independent analyses (low-coverage pooled data versus deep-coverage individual data) are strong candidates for the regulation of social behavior, especially those that refine QTL intervals. The deep sequencing data also presents the first opportunity to characterize haplotypes and even polymorphisms differentiating the tame and aggressive strains in regions likely to be under selection.

378T A model experiment of parallel selection for elucidating the architecture of complex traits. *R.G. Reeves*[1], W Zhang[1], J Großbach[2], M Clement[2], A Beyer[2], C Schmalohr[2], D Tautz[1] 1) Max Planck Institute Evolution, Germany; 2) University of Cologne, Germany.

In an effort to robustly explore the genetic architecture of a highly heritable complex trait we performed a select and re-sequence approach with 28 replicates. This was done within the context of a 15 generation pedigree incorporating > 1700 single pair crosses and > 90,000 phenotyped individuals. The novel trait of interest was the length of the *Drosophila melanogaster* pupal case (h^2= 0.44 ±0.04 SE & H^2= 0.58). The use of an automated phenotyping and data recording system made it possible to document the complete ancestry of every individual within the pedigree. The use of only single pair crosses renders this Drosophila data analogous to that available from most studies on plants and higher animals, including humans.

In addition to identifying statistically significant parallel allele-frequency changes in response to selection, we were also able to conduct multiple analytical approaches, within the same biological context, to assess their relative strengths and weaknesses: association, random-forest, , extended family-analysis and tests of epistasis. This analysis was based on >300 complete sequenced genomes. It is anticipated that this study and phenotype could represent a rich resource with which to examine analytical approaches when it is placed on-line. We propose that pupal size genetics in Drosophila could represent a model complex trait amenable to deep genetic dissection using the open source automated system.

Using established recombinant inbred panels (DGRP2 and DSPR) insights for the same trait of pupal length will be also contrasted.

379T Genetic variation in shoot developmental turnover underlies climate adaptation in the perennial plant *Arabidopsis lyrata*. *D.L. Remington*, B.K. Giri Department of Biology, University of North Carolina at Greensboro, Greensboro, NC.

Life history theory provides predictions for how climate-induced changes in fecundity, survival, and reproductive costs will affect evolutionary trajectories in perennial plants. However, the underlying genetic mechanisms remain largely a "black box," limiting our ability to apply evolutionary models. The perennial rock cress *Arabidopsis lyrata* provides unique opportunities to elucidate these mechanisms due to its wide variation in life history traits across a broad climatic range. Common-garden studies using *A. lyrata* populations with contrasting life histories from Norway and North Carolina (NC) and their crosses reveal local adaptation in response to the contrasting environments. However, QTL effects on these life history differences are inconsistent with predictions based on costs of reproduction; genetic variation in vegetative development affects reproductive output and not vice versa. The rates at which individual shoots transition from vegetative to reproductive growth show major differences between the two populations, consistent with a turnover-based model of perenniality. To establish the developmental mechanisms underlying the contrasting life histories and ultimately their molecular basis, we have analyzed QTL for a more detailed set of developmental traits, analyzed gene expression, and conducted phenocopy experiments. We focused on a QTL region previously shown to have large effects on life history traits, with antagonistic effects on fitness in the two parental environments. NC alleles increased the number of inflorescences primarily via large reductions in vegetative growth during the reproductive season, suggesting that the QTL mainly regulates shoot turnover. NC alleles also altered inflorescence architecture, which may explain why they reduce reproductive output in Norway. The QTL region includes two genes encoding auxin transporters, suggesting that variation in auxin transport might underlie the differences in shoot turnover. Consistent with this hypothesis, one of these candidates showed consistent imbalance of NC vs. Norway allelic expression in F1 hybrids, and treating NC plants with auxin inhibitors resulted in more Norway-like plant architecture. These results suggest that genetic variation in shoot turnover could be a key determinant of climate adaptation in perennial plants, point to potential underlying molecular mechanisms, and have implications for predicting plant life history responses to changing climates.

380T Tools for studying natural variation in Akt signaling in *Drosophila melanogaster*. *D.P. Sarikaya*[1], T. Gillies[2], D.J. Begun[1], J.G. Albeck[2] 1) Evolution and Ecology Department, UC Davis, Davis, CA; 2) Molecular and Cellular Biology Department, UC Davis, Davis, CA.

Cellular signaling pathways regulate the activity of cell and allow for cells, and by extension the organism, to respond accordingly to its environment. Deleterious mutations in these genes underlie many human diseases, yet evolutionary studies have shown that there are adaptive genetic differences in cell signaling pathway genes. How variable is pathway activity within a species? What effect does altering signaling level have on organismal phenotypes?

To address these questions, we created highly sensitive tools to study the activity levels of Akt, a downstream kinase of Insulin signaling (IS), at the single cell level in *Drosophila melanogaster*. Studying pathway activity often relies on tissue-level assays of phosphorylation levels of kinases or changes in downstream gene expression. However, recent studies have shown that single-cell level differences in the kinetics of signaling molecules can influence the cell's response to pathway activation. These findings suggest that there is an additional layer of information coded within the kinetics of signaling molecules that cannot easily be captured using tissue-level studies. A previous study in flies showed that clinal variants in the IS receptor have whole body differences in IS activity level. Building on this and other studies that implicate changes in IS genes in adaptive evolution, *Drosophila* provides an excellent system to investigate natural variation in cell signaling. We designed a reporter for Akt activity, and found that the reporter responds to drug-induced stimulation and/or inhibition of Insulin signaling in S2 tissue culture cells. Transgenic animals carrying the reporters were generated, and fixed tissues of starved adults show expected Akt activity patterns in different cell types. Live-imaging results of adult and larval tissues of transgenic animals, and data analysis methods will be discussed.

381T From Estimation to Prediction of Genomic Variances: Allowing for Linkage Disequilibrium and Unbiasedness. *N. Schreck*[1], M. Schlather[1,2] 1) Stochastics and Its Applications, University of Mannheim, Mannheim, DE; 2) Animal Breeding and Genetics Group, Center for Integrated Breeding Research, University of Goettingen, Goettingen, DE.

The additive genomice variance, the chief ingredient for the heritability, is often underestimated in phenotype-genotype regression models. Various remedies, including different models and estimators, have been proposed in order to improve on what has been coined the missing heritability. Recently, debates have been conducted whether estimators for the genomic variance include linkage disequilibrium (LD) and how to explicitly include LD in estimation procedures of the genomic variance.

Up-to-now, the genomic variance in random effect models (REM) has been estimated as a parameter of the marginal, i.e. unconditional model. As the seminal point in our work, we propose a strict conditioning on the data. This signifies a paradigm shift from the estimation to the prediction of the genomic variance. That is, the genomic variance in REM should be predicted as a conditional random quantity based on the conditional distribution of β. Our approach is structurally in perfect accordance with the Bayesian regression model (BRM), where the posterior genomic variance is estimated based on the

posterior of β. We introduce a novel, mathematically rigorously founded predictor for the genomic variance in (g)BLUP, which is structurally close to the Bayesian estimator, but can be given in a closed form expression.
We tackle, for the first time comprehensively and rigorously, the notion of unbiased estimation of the genomic variance with respect to the defintions in quantitative genetics (i.e. the genomic variance is caused by variability in allele content, whereas the marker effects are population parameters). We show that our approach and the Bayesian method are nearly unbiased. In contrast to that, other estimators are shown to underestimate the genomic variance, which has been observed as missing heritability for a long time. On the way, we show similar results for the fixed effect models (FEM), which is useful when the number of effects is small. An exemplary simulation study based on the commonly used dataset of 1814 mice genotyped for 10346 polymorphic markers substantiates that the bias of our novel predictor is small in all standard situations, i.e. that the predictor for the conditional genomic variance remarkably reduces the missing heritability. Moreover, our predictor performs at least as good as the estimator for the posterior genomic variance in BRM and the estimator in FEM.
Our approach has further crucial advantages. The conditioning of our predictors on the data is intrinsically tied to the inclusion of the contribution of LD and the predicted effects. Our predictor shows much weaker dependence on distribution assumptions than estimators of other approaches, e.g. GREML. Last but not least, our predictor (contrasted with the estimator in the unconditional model) enables an innovative approximation of the influence of LD on the genomic variance in the dataset.

382T Genomic prediction of puberty and reproductive traits in beef cattle and its accuracy in admixed individuals. *W. Shaffer*[2], D. Patterson[1], R. Schnabel[1,2], J. Taylor[1], J. Decker[1,2] 1) Division of Animal Science, University of Missouri, Columbia, MO; 2) Informatics Institute, University of Missouri, Columbia, MO.

For decades, estimated breeding values would be predicted for animal performance based on individual performance, progeny performance, and pedigree data. However, for young, unproven animals or for traits measured late in life, such predictions are subject to low accuracy and improvement. In the case of traits with low heritabilities, genomic prediction serves as a valuable tool to accurately increase selective pressure. Reproductive traits, which are traditionally lowly heritable, are of special importance given their influence on farm and ranch profitability. To represent sexual maturity at time of conception, ability to conceive, and factors limiting dystocia, we performed genomic predictions on the Reproductive Tract Score (RTS), Adjusted Days Pregnant (ADP), and Pelvic Height (PH) and Width (PW) traits in a cohort of 3078 high percentage Angus and Angus crossbred animals. Genotyping was conducted using the GGPF250 assay, which was developed for rare variant inclusion and research applications. Initial analyses included Univariate and Multivariate Genome Wide Association Analyses (GWAA) to detect significantly associated SNPs using the GEMMA software suite. GWAA SNP associations were visualized with Manhattan plots adjusted for multiple testing in R. Each trait was analyzed and multivariate analyses were included for RTS/ADP and RTS/PH/PW; PH revealed one significantly associated SNP and the multivariate RTS/PH/PW analysis revealed several SNPs that were weakly associated. Subsequently, genomic predictions were carried out on 3078 Angus individuals in GEMMA using the Bayesian Sparse Linear Mixed Model algorithm on randomly split, three-fold cross-validation groups (2/3 training and 1/3 prediction). To determine the accuracy of purebred predictions on crossbred individuals, we clustered the individuals to obtain high and low percentage Angus animals and trained on 2,463 high-percentage individuals. The high percentage group is validated within itself (high-percentage three-fold cross-validation) and independently on the 615 low-percentage individuals to evaluate prediction accuracy in crossbred cattle. Pearson correlations for the randomly split, three-fold cross-validation were low to low-moderate ranging from 0.02-0.27. While lowly to moderately heritable, RTS and PH are producing accurate genomic predictions even with our limited sample size.

383T Impact of genetic variation on mRNA splicing in human tissues. *A. Shah*[1], Y. Li[2,3] 1) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 2) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Chicago, Chicago, IL.

Up to 70% of disease-associated genetic variants do not have apparent effects on gene expression levels. This suggests that a large fraction of functional variants may disrupt gene regulation through alternate mechanisms, such as mRNA splicing. However, there exists no comprehensive analysis of the regulatory mechanisms by which genetic variation affects RNA splicing in a large number tissues. Here, we investigated the impact of genetic variation on RNA splicing in 53 tissues from the Genotype-Tissue Expression (GTEx) Project. Using LeafCutter, we identified thousands of splicing quantitive trait loci (sQTLs) in each tissue, and subsequently quantified the patterns of sQTL sharing across all tissues. While over 80% of the sQTLs were shared across tissues, we found that ~20% of genetic variants affected RNA splicing in a single tissue. By systematically analyzing UV crosslinking immunoprecipitation (CLIP) sequencing data for ~100 RNA-binding proteins from ENCODE, we found that tissue-specific sQTLs were enriched in introns harboring splice factor binding sites. Our findings suggest that RNA binding proteins play important roles in mediating sharing and specificity of functional genetic effects between tissues.

384T The Genetic Architecture of Morphology in Peruvian Native Potato. *L.M. Shannon*[1], R. Gómez[2], J. Soto[2], N. Anglin[2], D. Ellis[2], J.B. Endelman[3] 1) Horticultural Science, University of Minnesota, Saint Paul, MN; 2) Genebank, International Potato Center, Lima, Peru; 3) Horticulture, University of Wisconsin, Madison, WI.

Potatoes are the fourth most widely grown crop in the world (after wheat, rice, and corn). However, compared to these grains, much less is known about potatoes, in part because they have complicated clonal autopolyploid genomes. Potatoes were domesticated in Peru. As a result, Peruvian native potatoes which are the most diverse in the world, provide an ideal population for uncovering the genetic basis of morphological traits in potato. The International Potato Center (CIP) in Lima, Peru, houses the global in trust collection of potato with over 5,500 accessions of cultivated potato conserved in their genebank. We have genotyped the CIP collection with version 2 of the SolCAP SNP array, creating the most comprehensive potato diversity panel to date. The panel contains over 1687 predominantly tetraploid cultivars, with morphological descriptors developed as part of genebank curation. Of the 12K SNPs on the array, accurate tetraploid genotype calls were made for 9,292 markers using the ClusterCall package in R. To identify the genetic basis for the morphological descriptors, a Genome Wide Association Mapping (GWAS) approach using linear mixed models was implemented. The resulting QTL for coloring and patterning across tissues exhibited a range of dominance patterns, including single and multiple allele dominance. Many of the resulting QTL's co-localized near two known color loci, *Developer* and *Pigmented Flesh,* on chromosome 10. However, the majority of the QTL's are for traits not known to be controlled by either gene. *Developer* is a MYB transcription factor, one of seven in the QTL region. We hypothesize a series of diverged homologous MYBs and associated cis regulatory elements may be the primary determinants of color and patterning across tissues in Peruvian native potatoes.

385T *Cis*-acting variants affect gene expression dynamics within and between *Saccharomyces* species. *C. Shih*[1,2,3], J. Fay[1,2,3] 1) Department of Biology, University of Rochester, Rochester, NY; 2) Department of Genetics, Washington University in St. Louis, MO; 3) Center for Genome Science and Systems Biology, Washington University in St. Louis, MO.

Variation in *cis*-regulatory sequences has been shown to modulate gene expression levels and developmental patterns. However, temporal control of gene expression may be as important to fitness as steady-state levels. To determine whether *cis*-regulatory sequence variation contributes to gene expression dynamics as well as steady-state levels we measured allele-specific expression within multiple yeast hybrids during the transition from fermentative to respiratory growth.

We found 1225-1645 genes with significant *cis*-acting variation in expression dynamics within species, and 2666-2991 genes between species. As divergence increased we found a greater proportion of genes exhibited *cis*-acting variation in expression dynamics compared to expression levels. Through genome sequencing and analysis of polymorphism, we found that single nucleotide polymorphisms (SNPs) in promoters regions are associated with a 1.02-1.04 odds ratio (OR) of allele specific expression dynamics or expression levels. In contrast, insertions and deletions (InDels) are associated with a 1.11-1.15 OR of allele-specific expression dynamics or expression levels. Using the massively parallel reporter assay, CRE-Seq, with 7288 and 7232 synthetic promoters in the intraspecific and interspecific hybrid strains, we show for a subset of genes that promoter variants can recapitulated endogenous expression dynamics and levels.

We conclude that there is nearly as much *cis*-acting variants that affects temporal changes in gene expression as affects steady-state levels alone. Our observation that InDels have a stronger association than SNPs suggests a difference in how expression levels and dynamics are modulated.

386T Host-Pathogen QTLs underlying susceptibility to infectious disease. *C.M. Smith*[1], M.T. Ferris[2], F. Manuel de Villena[2], R. Williams[3], R. Baker[1], C.M. Sassetti[1] 1) Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, MA, USA; 2) Department of Genetics, University of North Carolina at Chapel Hill, NC, USA; 3) Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, TN, USA.

The complex interplay between host and pathogen determines if an individual controls infection or progresses to disease. While abundant evidence suggests that genetic diversity contributes to the variety of outcomes, the combined effect of variation in the host and pathogen remains unclear. We developed a "dual-genome" system to unravel genetic interactions between *Mycobacterium tuberculosis (Mtb)* and its mammalian host that drive outcome to infection. Host variation was modeled using a panel of ~100 mouse strains, including the Collaborative Cross (CC) and BXD panels. Bacterial variation was concurrently generated using saturated libraries of transposon mutants and panels of diverse *Mtb* clinical isolates.
The wide genotypic variation in the CC panel produced remarkably diverse phenotypes upon Mtb infection, ranging from extreme susceptibility to progressive clearance of the pathogen. Metrics of disease that are tightly linked in the typical C57BL/6 model such as bacterial burden, dissemination, weight loss and inflammation were genetically separable in the diverse strains. We identified individual polymorphic host genome regions (QTLs) underlying lung and spleen bacterial load and host control of infection independently in the CC and BXD panels.
We additionally separated the clinical disease traits into intermediate phenotypes by determining the relative fitness of thousands of bacterial mutants in the mouse panels. Host QTLs underlying differential bacterial fitness modules were identified, many of which mapped to the same host region as the clinical disease metrics. Each interaction between host and pathogen locus was defined as a host-pathogen QTL (hpQTL) that controls a specific aspect of the bacterial microenvironment and collaboratively influences overall susceptibility.

Overall, the strategy of using bacterial fitness profiles as reporters of the underlying host microenvironment is a sensitive and specific method for identifying disease-modifying host polymorphisms, demonstrating the power of a dual-genome systems genetics approach to understand the fundamental drivers of susceptibility to infection.

387T Identifying genetic modifiers of flight performance using the Drosophila Genetic Reference Panel. *A.N. Spierer*, J.A. Mossman, F.A. Lemieux, D.M. Rand Ecology and Evolutionary Biology, Brown University, Providence, RI.

Drosophila flight is a quantitative behavioral trait commonly used to evaluate organismal performance. This highly complex behavior requires a coordinated phenotypic response from many tissues and biological systems. However, the genes underlying flight performance in Drosophila are poorly understood.

Here, we sought to identify the most significant genetic modifiers of flight performance using the native Drosophila Genetics Reference Panel (DGRP) lines. We subjected approximately 100 flies of each sex from 189 DGRP lines to the flight performance assay pioneered by Seymour Benzer and further modified by the Ganetzky lab. We quantified mean landing height, dispersion of landing heights, and the proportion of flies that did not fly as distinct phenotypes.

The DGRP lines demonstrated wide variation in all measured traits; the highest mean landing height was roughly twice that of the lowest mean landing height for both sexes. These phenotype metrics were used as inputs for separate Genome Wide Association Studies (GWAS). We found a strong positive correlation between male and female flight performance ($r = 0.76$, $p < 0.00001$) across the DGRP lines. In addition, we identified a number of loci (20 with LOD score > 8) whose association was significant across the sexes, suggesting these are likely strong candidate genes for flight ability in Drosophila. In contrast, the majority of significant loci associated with flight ability in one sex showed no association in the other sex, suggesting a complex sex-specific genetic architecture to this quantitative trait. In addition to intergenetic regions, candidate regions, and transcription factor binding domains, we identified a number of SNPs in candidate genes corresponding with neuronal patterning and function (*Cadherin-N*, *Snoo*, *alan shepard*), development (*odd skipped*, *Dorsocross 2*, *bric a brac 1*) and transcription factors (*chameau*, *Sox21b*). These genes and associated ontologies suggest the interdependence of various integrated biological systems that likely have a strong genetic basis for flight.

Our next step is to conduct flight performance assays in DGRP lines heterozygous for a Gal4-UAS system, allowing us to investigate specific gene functions in the Drosophila indirect flight muscle.

388T Parallel Evolution of Ethanol Tolerance found in Four Populations of *Drosophila melanogaster*. *Q.D. Sprengelmeyer*, J E Pool Genetics, University of Wisconsin-Madison, Madison, WI.

Ethanol tolerance in *Drosophila melanogaster* has been shown to increase with latitude and linked to different genes, most notably *ADH*. This study focused on ethanol tolerance of flies from their ancestral range (Zambia) compared to populations found in higher altitudes (Ethiopia, South Africa, and Uganda) and latitude (France). To test for ethanol tolerance we exposed the flies to an 8% ethanol solution and measured the survivability over a 12-hour period. Flies from low altitude Zambia were found to be the least ethanol tolerant, whereas the other three sub-Saharan high altitude populations had different levels of tolerance. The flies from France were extremely tolerant, having nearly 100% survivability. To ascertain candidate genes responsible for higher ethanol tolerance, bulk segregant analysis was performed to detect quantitative trait loci (QTL). Each of the ethanol tolerant populations had mainly different QTL peaks and substantial differences were observed between strains from the same population as well. These data suggest that the loci for ethanol tolerance may not be fixed, or else that epistatic interactions significantly alter mapping results. To find evidence of selective sweeps, F_{ST} and the haplotype statistic χ_{MD} were analyzed, comparing each of the higher tolerance population genomes with low-tolerance population Zambia. From this study new insights can be gained in the genetics of adaption; in particular the polygenicity of a trait evolution and its genetic predictability between the different populations.

389T Characterizing translational regulation in genetically diverse mouse strains. *Alexander Stanton*, Steven Munger The Jackson Laboratory, Bar Harbor, ME.

Proteome homeostasis is critical to cell function. Emerging evidence demonstrates that genetic variation affecting gene transcript abundance is frequently not observed at the protein level. Conversely, changes in protein abundance are often not driven by changes to transcript abundance. These findings highlight the importance of posttranscriptional regulation of the proteome. A recent quantitative trait loci (QTL) mapping study from our group analyzed liver transcript and protein abundance in Diversity Outbred (DO) mice and identified a population of genes with uncoupled transcript and protein abundance. Protein expression of these genes is controlled by changes at single distant genetic loci, which could be working through translational and/or posttranslational mechanisms. I am utilizing ribosome profiling in livers of the eight inbred DO founder strains to assess the extent to which differential translation efficiency is responsible for the expression of these genes. We anticipate ribosome profiling to identify two subcategories of transcript-protein discordant genes within the DO founder population: genes where translation efficiency can and cannot account for observed transcript-protein differences.

Genes in the former category exhibiting differential translation that mirrors the founder allele effect at its distant protein-QTL can be further interrogated by dissecting codon-level resolution ribosome stalling, alternative exon inclusion, or other effects detectable from ribosome position. Genes that do not show variation in translational efficiency are most likely regulated by posttranslational mechanisms, including the stoichiometric dependency of physically-interacting protein partners and multimeric complex members we identified in our previous study. By integrating this transcriptome-wide quantification of translation across a diverse panel of mouse strains to our existing transcript and protein QTL data, we will catalog the specific genetic variants – and model classes of genetic variation - that affect individual or multiple steps of the gene regulatory cascade. We expect these insights into the proteome and proteostasis in diverse mice will better inform predictions of drug response and our understanding of disease etiology in humans.

390M Population complexity trumps model complexity in understanding trait variation. *M.G. Sterken*, L.B. Snoek, R.P.J. Bevers, R.J.M. Volkers, J.A.G. Riksen, J.E. Kammenga Laboratory of Nematology, Wageningen University, Wageningen, NL.

The study of expression quantitative trait loci (eQTL) through the use of recombinant inbred lines has yielded detailed information about the transcriptional regulation of complex traits. However, it has proven difficult to apply more advanced genetic models explaining genetic variation underlying gene expression differences. Here, we make use of the difference in genetic complexity of two types of inbred population in the nematode *Caenorhabditis elegans* to estimate the number of loci affecting gene expression.

We measured gene-expression in a recombinant inbred line (RIL) and an introgression line (IL) population constructed from crossing the strains N2 and CB4856. Both populations received a heat-shock treatment and gene-expression profiles were obtained before (48h at 20°C), directly after heat-shock (2h at 35°C), and after a recovery period (2h at 20°C). Making use of the difference in genetic make-up between the populations - few loci from one parent in the IL versus many in the RILs - allowed for the identification of transcripts regulated by multiple loci. By measuring the transcript variance within each population, for over 1,000 genes across the three conditions we found strong evidence for multiple eQTL underlying gene expression variation. Importantly, most of these multi-loci eQTL are environment-specific. Furthermore, we observed over 200 genes where the phenotypic variation in the IL panel significantly exceeded that in the RIL panel, suggesting evidence for complex genetic buffering.

In conclusion, by using two types of inbred populations the complexity of trait architectures can be investigated without reliance on models of higher complexity. The genetic complexity of a trait is directly observed, rather than estimated *post-hoc*. Therefore, relying on *population complexity* rather than *model complexity* can provide valuable insight in the architecture of quantitative traits.

391T Inferring the population genetics of polygenic adaptation through machine learning. *M.G. Stetter*[1], K. Thornton[2], J. Ross-Ibarra[1] 1) Department of Plant Sciences, University of California Davis, Davis, CA; 2) Department of Ecology and Evolutionary Biology, University of California, Irvine, CA.

Understanding the genetic basis of phenotypic adaptation to changing environments is an essential goal of population and quantitative genetics. While technological advances now allow interrogation of genome-wide genotyping data in large panels, our theoretical understanding of the process of polygenic adaptation is still quite limited. To address this limitation, we use extensive forward-time simulation to explore the impacts of variation in demography, trait genetics, and selection on the rate and mode of adaptation and the resulting genetic architecture. We simulate a population adapting to an optimum shift, modeling sequence variation for 20 QTL for each of 12 different demographies for 100 different traits varying in the effect size distribution of new mutations, the strength of stabilizing selection, and the contribution of the genomic background. We then use random forest regression approaches to learn the relative importance of input parameters for statistics of interest such as the speed of adaptation, the relative frequency of hard and soft sweeps, or the final genetic architecture of the trait.
We find that selective sweeps abound, even for slowly-moving traits under relatively weak selection and where the genetic background explains most of the variation. Though most sweeps occur from variation segregating in the ancestral population, new mutations can be important for traits under strong stabilizing selection that undergo a large optimum shift. In spite of this effect, we find that deleterious mutations are more strongly influenced by the strength of stabilizing selection. We also show that population bottlenecks and expansion impact overall genetic variation, but also the relative importance of soft sweeps and the speed with which adaptation can occur.
We then use the matrix of effect sizes and allele frequencies in each population as a target for machine learning and find that demography and the effect size of new mutations have the largest influence on present day genetic architecture. But because a variety of parameter combinations can result in relatively similar genetic architectures, we find that it is not possible to infer much about the process of adaptation from the genetic architecture alone. Overall, our results underscore the population genetic complexity of individual loci in even relatively simple quantitative trait models, but provide a glimpse into the factors that drive this complexity.

392T Behavioral and Life-History Evolution in Island Mice with Extreme Body Size. *Jered Stratton*, Megan Latsch, Mark Nolte, Michelle Parmenter, Bret Payseur Genetics, University of Wisconsin - Madison, Madison, WI.

Common ecological features of islands have the potential to generate rapid phenotypic evolution across a wide array of colonizing populations. The "island syndrome" synthesizes repeated patterns of directional evolution in morphology, behavior, and life history observed in island populations. The syndrome predicts that small-bodied animals will evolve larger size, delayed maturation, smaller litters, decreased mortality, and less aggressive behavior on islands than on the mainland. To determine the degree to which substantial change in body size is accompanied by the evolution of novel life histories and behaviors, we phenotyped house mice (*Mus musculus domesticus*) from Gough Island, which recently evolved to become twice as large as their mainland relatives. Compared to a wild-derived inbred strain from the mainland, Gough Island females delay sexual maturity by approximately one day. Gough Island mice give birth to two more pups per litter on average with decreased juvenile mortality. Finally, Gough Island mice are relatively more active and spend more time in the center during an open field test. Our findings set the stage for genetic dissection of divergence in behavior and life history associated with the evolution of extreme body size in island populations.

393T Reciprocal, sparse diallels of Collaborative Cross mice for identifying effects of parent-of-origin, perinatal exposures, allelic imbalance, and X-inactivation skew. *K. Sun*[1,2], D. Oreper[1,2], S. Schroenrock[1,3], R. McMullan[1,4], L. Tarantino[1,5], F. Pardo-Manuel de Villena[1], W. Valdar[1] 1) Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, North Carolina; 2) Bioinformatics and Computational Biology Curriculum, School of Medicine, University of North Carolina, Chapel Hill, North Carolina; 3) Neuroscience Curriculum, School of Medicine, University of North Carolina, Chapel Hill, North Carolina; 4) Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina, Chapel Hill, North Carolina; 5) Department of Psychiatry, School of Medicine, University of North Carolina, Chapel Hill, North Carolina.

It is well established that genetics, parent-of-origin, and environment have effects on behavior, and that these factors likely modulate the effects of each other. Studying such an interplay - in particular, devising an experimental or population design that enables such study - is, however, challenging. Here we combine two powerful approaches: the reciprocal cross, a design that detects parent-of-origin effects unambiguously; and the replicable genetic reference population, a genetic system ideally suited to investigating how background modulates the effects of environmental exposures. Using a sparse diallel cross of the Collaborative Cross mice, a panel of recombinant inbred mouse strains, we performed a survey of genome-wide and perinatal diet effects on behavioral traits. The reciprocal intercross design of our study population is especially powerful for identifying parent of origin (PO) effects of allelic imbalance in individual CC strains. Prior to mating and throughout gestation, the parental CC females were exposed to different nutrient-deficient diets. Data from a suite of behavioral tests and gene-expression (RNA-seq) assays were collected on the resulting F1 CC-recombinant inbred (CC-RIX) pups. We applied Bayesian hierarchical statistical models to decompose the influence of PO and maternal diet, along with complex interactions of parental line, maternal diet, and PO, on these data. Among the behavioral outcomes, we find significant PO effects for jumping behaviors in the open field and distance traveled during the light-dark assay in two RIX crosses. These initial results help focus our efforts to identify PO effects in differential RNA expression, and ultimately, to identify reciprocal effects on gene transcript levels correlated with downstream changes in these traits. Using RNA-seq data that was quantified for allele-specific expression, we discovered biases in gene expression on the X chromosomes in female CC-RIX pups that indicate uneven X-inactivation in either the paternal or maternal X. We can apply the same methodology to find imprinted genes throughout the genome where an allele corresponding to one parent is overrepresented, indicating epigenetic processes at play. Our matched study design produces reciprocal crosses of parental lines and uniquely allows for an accurate and well-controlled accounting of PO effects, providing a valuable tool for a more complete understanding of how diet *in utero* affects imprinting regulation.

394M The potential of regularized regression to provide more accurate multivariate selection estimates. *J. Sztepanacz*[1,2], D. Houle[1], T.F. Hansen[2] 1) Florida State University, Tallahassee, FL; 2) Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway.

The breeder's equation $\Delta\mathbf{z}=\mathbf{G}\boldsymbol{\beta}$, allows us to understand how genetics (the genetic covariance matrix **G**) and the vector of linear selection gradients selection ($\boldsymbol{\beta}$) interact to generate evolutionary trajectories. Estimation of $\boldsymbol{\beta}$, using multiple regression of trait values on relative fitness, revolutionized the way we study selection in laboratory and wild populations. Multicollinearity, or correlation of predictors, is a major challenge for any multiple regression approach, that can lead to very high variances and covariances between elements of $\boldsymbol{\beta}$. The usual approach to multicollinear predictors is to discard some of them, thereby losing any information that might be gained from those traits. Using simulations, we show how, on the one hand, multicollinearity can result in inaccurate estimates of selection, and on the other how the removal of correlated phenotypes from the analyses can provide a misguided view of the targets of selection. We show that regularized regression, which places a priori constraints on features of $\boldsymbol{\beta}$, generates more accurate estimates of selection in the presence of multicollinearity. We compare standard and regularized regression estimates of selection on sexual pheromone signals in

the Australian fruit-fly *Drosophila serrata*. This analysis, and our simulations suggest that regularized regression should be adopted as a valuable tool in selection analyses.

395T Assessment of sugar beet (*Beta vulgaris* L. subsp. vulgaris) elite inbred lines in Japan. *K. Taguchi*[1], Y. Kuroda[1], K. Okazaki[1], M Yamasaki[2] 1) NARO, Hokkaido Agricultural Research Center , Hokkaido, JP; 2) Graduate School of Agricultural Science, Kobe University, Japan.

In plant breeding, it is essential to investigate the genetic diversity and phenotypical variation of elite inbred lines. Genetic diversity of 63 Japanese sugar beet (Beta vulgaris L. subsp. vulgaris) elite inbred line diversity set (JSBDIV) was investigated using 33 cleaved amplified polymorphic sequence and 38 simple sequence repeat analyses in one nation level. In addition, the data set was evaluated agronomic important traits such as root yield, sugar content, quality relating components and disease resistances in field experiments. The lines were significantly subdivided into 7 (Pedigree information), 7 (NJ method) or 12 (Structure analysis) subgroups according to whether a comparison of three levels of classification pedigree information, the construction of a phylogenetic tree, or an analysis of molecular variance was undertaken. Structure analysis yielded a high PhiPT value, indicating that the explainable variation among subgroups was 32%, higher than that achieved with the other classifications in AMOVA. Given some subgroups close connection with ancestral open-pollinated varieties (OPVs), it was estimated that subgroups were created as their heterotic pool of genetic contribution. A high variation in all phenotypic traits was observed for both OPVs and each subgroup. Variation in Sugar content was significant among subgroups: the 'TA15' (one of ancestral OPV) heterotic pool was characterized as being higher in sugar content than other subgroups; however, secondary subgroups were also characterized as high sugar which bore no genetic contribution from the 'TA15' heterotic pool. Some subgroups showed high resistance to Aphanomyces root rot, Cercospora leaf spot and Rhizoctonia root rot. The conclusions of such assessment would be very informative in designing a breeding scheme or to carry-out maintenance management of this genetic variation. Regarding the relationship among heterotic pools, it was suspected that ancestral OPVs had more or less historically overlapped with each other, and in some cases the wide DNA variation in ancestral OPVs might have been reflected by PhiPT values. However, it was estimated by structural analysis that more robustly discriminated subgroups within the breeding population could be delineated than any other classification method.

396T Epistatic networks jointly influence phenotypes related to metabolic disease and gene expression in Diversity Outbred mice. *A.L. Tyler*, B. Ji, D.M. Gatti, S.C. Munger, G.A. Churchill, K.L. Svenson, G.W. Carter The Jackson Laboratory, Bar Harbor, ME.

Genetic studies of multidimensional phenotypes can potentially link genetic variation, gene expression, and physiological data to create multi-scale models of complex traits. The challenge of reducing these data to specific hypotheses has become increasingly acute with the advent of genome-scale data resources. Multi-parent populations derived from model organisms provide a resource for developing methods to understand this complexity. In this study, we simultaneously modeled body composition, serum biomarkers, and liver transcript abundances from 474 Diversity Outbred mice. This population contained both sexes and two dietary cohorts. Transcript data were reduced to functional gene modules with weighted gene coexpression network analysis (WGCNA), which were used as summary phenotypes representing enriched biological processes. These module phenotypes were jointly analyzed with body composition and serum biomarkers in a combined analysis of pleiotropy and epistasis (CAPE), which inferred networks of epistatic interactions between quantitative trait loci that affect one or more traits. This network frequently mapped interactions between alleles of different ancestries, providing evidence of both genetic synergy and redundancy between haplotypes. Furthermore, a number of loci interacted with sex and diet to yield sex-specific genetic effects and alleles that potentially protect individuals from the effects of a high-fat diet. Although the epistatic interactions explained small amounts of trait variance, the combination of directional interactions, allelic specificity, and high genomic resolution provided context to generate hypotheses for the roles of specific genes in complex traits. Our approach moves beyond the cataloging of single loci to infer genetic networks that map genetic etiology by simultaneously modeling all phenotypes.

397T Extreme QTL Analysis of Lifespan in an Advanced Intercross Population of *Drosophila*. *D. Unselt*[1,2], L. Everett[1,2], T. Morozova[1,2], G. Arya[1], L. Turlapati[1], R. Anholt[1,2], T. Mackay[1,2] 1) Department of Biological Sciences, Program in Genetics, North Carolina State University, Raleigh, NC; 2) W.M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC.

Understanding the genetic mechanisms affecting variation in lifespan in natural populations is crucial for understanding the genetic basis of age-related diseases. Lifespan is known to vary in natural populations due to the segregation of multiple genetic factors as well as to exposure to different environmental conditions. Further, many pathways associated with lifespan, such as the insulin-like signaling pathway, are evolutionarily conserved between humans and model organisms. *Drosophila melanogaster* is a powerful model for assessing naturally occurring genetic variation in lifespan because of the ability to perform genomic analyses on a large scale while effectively monitoring genetic backgrounds and controlling environmental conditions. The *D. melanogaster* Genetic Reference Panel (DGRP), a population of inbred, sequenced lines, facilitates mapping the effects of natural genetic variation on phenotypically variable traits, including

lifespan. We developed an outbred advanced intercross population (AIP) using 37 DGRP lines that were maximally genetically divergent. We selected flies at random as well as the longest 10% surviving flies and sequenced the pools to identify variants with significant changes in allele frequency between the longest lived and control pools (extreme QTL mapping). We identified 363 (458) single nucleotide polymorphisms (SNPs) in 263 (328) candidate genes in females (males) (P-value ≤ 0.05). We have also collected samples of flies at weekly intervals and dissected heads, reproductive organs and carcasses for gene expression analysis. We plan to map SNPs to transcriptional start and end sites of genes whose expression changes with age to infer novel genetic networks associated with variation in aging. Since basic biological processes, such as aging, are evolutionarily conserved, these studies will also provide candidate genes for investigation in other species, including humans.

398T The *Drosophila* model system for analyzing natural variation in resistance to *Metarhizium* spp. *J. Wang*, H-L Lu, R. St. Leger University of Maryland, College Park, MD.

We have used broad and narrow host range *Metarhizium* spp, generalist and narrow host plant specialist *Drosophila* spp, 188 *Drosophila melanogaster* Genetic Reference Panel (DGRP) lines, and a panel of mutant *D. melanogaster* to explore the genetic basis of natural variation in *Metarhizium* host specialization and insect disease resistance. Features of interest include: 1) substantial individual variation in disease resistance between flies collected in the same habitat, with implications for the evolution of disease resistance; 2) strong genetic control over microenvironmental plasticity, particularly in fly species that have a wide environmental range; 3) males are typically more resistant than females; 4) little difference between sporulation capacity on resistant lines and susceptible lines, so there is no penalty for the fungus to kill quickly, and 5) variation between DGRP lines in resistance to *M. anisopliae* was correlated with resistance to *Pseudomonas aeruginosa*, oxidative stress sensitivity, sleep duration and number of nightly 'naps'. We identified a host of candidate genes associated with variation in disease resistance, many of which are known to interact physically and/or genetically which enabled us to place them in a biologically informative genetic network. Overall, our results suggest that flies differ in their ability to control and/or tolerate replicating fungi during infection, which is achieved mostly through the coordinated interplay of morphological and physiological restraints, and phagocytic effectors that function in subtly different ways in different lines. However, the majority of polymorphisms with major effects on disease resistance were rare, suggesting a general cost to defense involving trade-offs.

399T Environment-dependent pleiotropic effects of mutations on growth rate and carrying capacity of population growth. *X. Wei*, Jianzhi Zhang Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI.

Growth rate (r) and carrying capacity (K) are key life history traits that together characterize the density-dependent population growth, and therefore are crucial parameters of many ecological and evolutionary theories. Although r and K are generally thought to be negatively correlated, both r-K tradeoffs and tradeups have been observed. However, neither the conditions under which each of these relationships occurs nor the causes of these relationships are fully understood. Here we address these questions using genetic mappings of r-QTLs and K-QTLs followed by mathematical modeling. We estimated r and K using the growth curves of more than 7000 yeast recombinant diploid genotypes in nine lab environments and found that the r-K correlation among genotypes changes from 0.53 to -0.52 with the rise of the environment quality, measured by the mean r of all genotypes in the environment. Many QTLs simultaneously influence r and K, but the directions of their effects are environment-dependent such that a QTL could show concordant effects on the two traits in a poor environment but antagonistic effects in a rich environment. We propose that these varying trends are generated by the relative impacts of two factors: the tradeoff between the speed and efficiency of ATP production and the energetic cost of cell maintenance relative to reproduction, and demonstrate a good agreement between model predictions and empirical observations. Together, these results reveal and explain the complex environment-dependency of the r-K relationship, which bears on many ecological and evolutionary phenomena.

401T Accurate, ultra-low coverage impuation and association studies in Hybrid Swarm mapping populations. *Cory Weller*, Alan Bergland Biology, University of Virginia, Charlottesville, VA.

Genetic association studies seek to describe how underlying genetic architecture translates to the observable phenotype. Here, we describe a new, resource-efficient approach to fine-scale genetic mapping and compare it to alternatives. Modern mapping studies often rely on reference panels of inbred lines derived from wild-caught organisms, which serve as temporally stable snapshots of genetic variation. One popular method of generating an outbred mapping population from such inbred reference panels involves successive rounds of recombination (e.g., 50 generations) between relatively few founders–eight-way crosses are commonly used in plants, invertebrates, and mammals. We evaluated an alternative mapping approach, which we refer to as the Hybrid Swarm. In the Hybrid Swarm, dozens to hundreds of pre-existing inbred lines from a refence panel are crossed together for few (4-6) generations, allowing researchers to rapidly generate custom mapping populations for which full genomes can be imputed from ultra-shallow sequencing data. We developed and applied an imputation pipeline to thousands simulated of Hybrid Swarm genomes, modeling genetic variation from the *Drosophila* Genetic Reference Panel (DGRP). With as little as 0.05X sequencing coverage, our pipeline correctly estimated Hybrid Swarm genotypes with 99% accuracy at variable sites. Next, we tested whether Hybrid Swarm populations are suitable for high-

resolution association studies by developing a mapping pipeline that accepts any user-defined panel of inbred founding lines, simulates recombination over a specified number of generations, and stochastically models percent variation explained (PVE) for a polymorphism of predetermined frequency. We simulated thousands of Hybrid Swarm populations (derived from both the DGRP and neutrally-evolving coalescent simulations), which yielded comparable accuracy and precision to other mapping population designs. Our analysis demonstrates the feasibility and efficiency of alternative methods for designing and conducting genetic mapping studies.

402T Dissecting the Genetics Basis of Learning, Memory, and Thermal Tolerance in a Multi-parental Population of Fruit Flies. *P. Williams-Simon*, C. Posey, E.G. King, T. Zars Biological Sciences, University of Missouri, Columbia, MO.

Learning, memory, and thermal tolerance are complex traits that are fundamental survival skills in many species. For example, everyday tasks such as: foraging, finding a mate, escaping predation, and acclimating to temperature changes, are highly dependent on how well an organism remembers what was learned. Although, both learning and thermal tolerance mechanisms are evolved traits used to adapt to warmer temperatures, studies examining both mechanisms, and how they relate genetically have not been done. Understanding the natural variants of the genes that control these traits is of high importance if we want to better comprehend how variation in these traits arise within a species. Here, I take a quantitative genetics approach to dissect the genetic basis of learning, memory, and thermal tolerance using *Drosophila* Synthetic Population Resource (DSPR). This multi-parental population consists of approximately 1,800 Recombinant Inbred Lines, which allows for high-resolution genome wide scans, and the identification of loci contributing to naturally occurring genetic variation. Using a behavioral assay known as "place learning", we are able to train flies with a highly sensitive apparatus, the "heat box". Whenever a fly crosses the midline of the chamber within the heat box, the whole chamber either warms or cools. This allows for us to test both how well a fly learns to avoid uncomfortable temperatures, and how well the fly retains this memory. We found that there was approximately a 2-fold difference between the performance index (PI) of flies for learning (0.5 – 1), memory showed a nearly 10-fold range of variation (0.1 – 1), and thermal tolerance ranges from, 60s – 431s. We then performed genome scans using the DSPRqtl R package, which uses a Haley Knott regression to test for an association between phenotypes and genotypes and identified several loci (QTL) affecting each trait. Our results revealed that there is a genetic basis for variability in these traits, and that there are some loci of moderate effect within the fly genome that are important for these traits. These loci have not been previously been implicated in learning or memory. We performed RNAseq to identify differentially expressed candidate genes within these QTL that affect learning, memory, and thermal tolerance. Future work will aim to fine map these loci, and validate the function of these genes in learning, memory, and thermal tolerance.

403T Recombination in mtDNAs reveals negative mito-mito epistasis between *Saccharomcyes cerevisiae* populations. *J. Wolters*[1], G. Charron[2], A. Gaspary[1], C.R. Landry[2], A. Fiumera[1], H. Fiumera[1] 1) Biology, Binghamton University, Binghamton, NY; 2) Institut de Biologie Intégrative et des Systèmes, Département de Biologie, Département de biochimie, microbiologie et bio-informatique, Université Laval, Québec, Canada.

Mitochondrial haplotypes contribute to functional diversity in natural populations. Uniparental inheritance makes it difficult to characterize the genetic architecture of mitochondrially driven phenotypes. We used mitochondrial recombination that naturally occurs in *Saccharomyces cerevisiae* to explore how variants contribute to growth. We generated mitochondrial recombinants with novel phenotypes by crossing two strains with identical nuclear genomes but haplotypes from divergent populations. Intermediate growth of recombinants in conditions that tax mitochondrial function suggests multiple loci with additive effects. Surprisingly, recombinants showed poor growth in ecologically relevant conditions consistent with negative epistasis between mitochondrial loci (mito-mito epistasis). We created a large collection of recombinants between four divergent mitochondrial haplotypes from various populations and assessed growth in multiple conditions. Positive estimates of mito-mito epistasis affecting growth rate were found in all conditions. Negative mito-mito epistasis may act in concert with mitochondrial-nuclear epistasis as a post-zygotic barrier contributing to speciation.

404T Genome-wide association for testis weight in the diversity outbred mouse population. *J.T. Yuan*[1], D.M. Gatti[2], V.M. Philip[2], S. Kasparek[1], A. Kreuzman[1], B. Mansky[1], K. Sharif[1], D. Taterra[1], W.M. Taylor[1], M. Thomas[1], J.O. Ward[1], A. Holmes[3], E.J. Chesler[2], C.C. Parker[1] 1) Middlebury College, Middlebury, VT; 2) The Jackson Laboratory, Bar Harbor, ME; 3) National Institute on Alcoholism and Alcohol Abuse, NIH, Bethesda, MD.

Testis weight is a genetically mediated trait associated with reproductive efficiency across numerous species. We sought to evaluate the genetically diverse, highly recombinant Diversity Outbred (DO) mouse population as a tool to identify and map quantitative trait loci (QTLs) associated with testis weight. Testis weights were recorded for 502 male DO mice and the mice were genotyped on the GIGAMuga array at ~143,000 SNPs. We conducted a genome-wide association analysis and identified one significant and two suggestive QTLs associated with testis weight. Using bioinformatic approaches, we developed a list of candidate genes and identified those with known roles in testicular size and development. We performed targeted exon sequencing on a subset of candidate genes for 10 mice and compared the resulting sequence data amongst the animals. Candidates of particular interest include the RNA demethylase gene *Alkbh5*, the cyclin-dependent kinase inhibitor gene

Cdkn2c, the dynein axonemal heavy chain gene *Dnah11*, the phospholipase D gene *Pld6*, the trans-acting transcription factor gene *Sp4*, and the spermatogenesis-associated gene *Spata6*, each of which has a human ortholog. Our results demonstrate the utility of DO mice in high-resolution genetic mapping of complex traits, enabling us to identify developmentally important genes in adult mice. Understanding how genetic variation in these genes influence testis weight could aid in the understanding of mechanisms of mammalian reproductive function.

405T The genetic basis of evolutionary transitions in early development. *C. Zakas*, M. Rockman Biology, New York University, New York, NY.

Phenotypic evolution in animals is constrained by the mechanics of early development. Large-scale evolutionary changes are initially shaped by developmental program, where simple trade-offs can ultimately result in a vast spectrum of physiological, morphological, and ecological differences. How do these major transitions in development occur? The polychaete annelid *Streblospio benedicti* provides a unique opportunity to use forward genetics to experimentally dissect a major transition in animal development. *S. benedicti* produces two distinct offspring types that differ in egg size, early development, and larval morphology. Thus it is an ideal genetic model for the evolutionarily common transition between indirect and direct development. Using genetic crosses between these types, we constructed the first annelid genetic map, which reveals the distribution of genetic factors affecting a suite of genetically separable developmental phenotypes. Because early development is strongly influenced by maternal effects, our cross design disentangles maternal and zygotic genetic effects and shows that a transition from indirect to direct development requires contributions from both the zygotic and maternal genome; an increase in egg size alone is not sufficient to change development mode. By identifying the loci responsible for early developmental phenotypes, we begin to uncover how the transition from indirect to direct development proceeds.

406T Natural variation in C. elegans arsenic-induced toxicity is explained by differences in branched chain amino acid catabolism. *S. Zdraljevic*[1,2], B.W. Fox[3], O. Panda[3], S.C. Brady[1,2], T.A. Crombie[2], F.C. Schroeder[3], E.C. Andersen[1,2,4] 1) Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA; 2) Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA; 3) Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA; 4) Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL 60611, USA.

Organisms have long been forced to adapt to the environmentally ubiquitous toxic metalloid arsenic. Currently, an estimated 100 million people are at risk of chronic exposure to arsenic. Recent evidence from genome-wide association studies suggest that polymorphisms in the *AS3MT* gene, which are present in human subpopulations exposed to elevated arsenic, result in an increased ability to metabolize arsenic. Given the difficulty in functional characterization of human variation, we have no experimental evidence that links polymorphisms in the *AS3MT* gene to arsenic metabolism or toxicity. However, these results do suggest that standing genetic variation can be used to identify mechanisms by which organisms tolerate arsenic exposure. In the present study, we take advantage of the genetic diversity present in the tractable model organism *Caenorhabditis elegans* to identify a novel mechanism of arsenic toxicity that might have arisen as a result of differential exposure to arsenic among subpopulations of this free-living nematode species.

Using two genetic mapping approaches, we show that a major source of variation in *C. elegans* responses to arsenic trioxide is caused by natural variation in the *dbt-1* gene. This gene encodes for the E2 subunit of the branched-chain α-keto acid dehydrogenase (BCKDH) complex, which is a core component of branched-chain amino acid (BCAA) catabolism. We used CRISPR/Cas9-mediated allele editing to show that a single non-synonymous variant (C78S) in the highly conserved lipoyl domain of DBT-1 is the causal polymorphism underlying variation in response to arsenic trioxide. Next, we used unbiased metabolomics and chemical supplementation experiments to show that differences in *C. elegans* responses to arsenic trioxide result from differential depletion of mono-methyl branched chain fatty acids, metabolites with a central role in developmental progression. We hypothesize that the presence of the thiol group in the sensitive DBT-1 C78 allele mediates the coordination of arsenic binding to the reduced lipoic acid cofactor, thereby inhibiting the catalytic cycle of the BCKDH. Our study marks the first time that the BCKDH complex and BCAA metabolism have been implicated in the response to arsenic. These results demonstrate the power of using natural genetic diversity of *C. elegans* in combination with comparative metabolomics to identify mechanisms by which environmental toxins affect organismal physiology.

407T Genetic analysis on multi-parental populations. *L. Zhang*, J. Wang, L. Meng, S. Zhang Institute of Crop Science, CAAS, Beijing, CN.

Genetic populations derived from multiple parents are becoming more important because of the increased genetic diversity. Compared with bi-parental populations, genetic analysis methods are less investigated in multi-parental populations. In the past few years, we conducted studies on linkage analysis and QTL mapping methods for clonal F_1, four-way cross F_1, and pure-line populations derived from four-way and eight-way crosses. For linkage analysis, 1) markers categories were classified based on the number of distinguishable alleles in parents and genotypes in the progenies. 2) For

two linked marker loci, theoretical frequencies of identifiable genotypes were derived, from which the maximum likelihood estimates of recombination frequency could be estimated. 3) Linkage maps were built using the estimated recombination frequencies. Three sets of linkage maps could be built for clonal F_1 and four-way crosses, i.e. female, male and combined maps. Only one set of linkage map could be built for pure-line populations, i.e. combined map. For QTL mapping, 1) incomplete and missing markers were imputed as fully informative markers using the combined linkage map. 2) Orthogonal variables were created for each marker and used in an inclusive linear model, so as to completely absorb the genetic variation in the mapping population. 3) Inclusive composite interval mapping (ICIM) approach was implemented for one-dimensional scanning, during which the inclusive linear model was employed to control the background variation. The efficiency of the proposed methods were demonstrated by extensive simulations and by comparisons with other software packages. The proposed methodology for linkage map construction can build more accurate linkage maps in shorter time. The proposed methodology for QTL mapping is efficient when considering high detection power, low false discovery rate (FDR) and high accuracy in estimating QTL locations and effects. For software development, the package GACD was developed for linkage map construction and QTL mapping for clonal F_1 and four-way crosses. The package GAPL was developed for genetic analysis for pure-line populations derived from four-way and eight-way crosses. Three functionalities were integrated in each software with user-friendly interface: binning of redundant markers, linkage map construction, and QTL mapping, which acted as a pipeline. The two packages are freely available from www.isbreeding.net.

408T Using automated phenotyping to study the genetic architecture of pupation site selection in Drosophila melanogaster. *W. Zhang*, G. Reeves, D. Tautz Department of evolutionary genetics, Max Planck Institute for Evolutionary Biology, Plön, Schleswig-Holstein, DE.

Pupation site selection of Drosophila melanogaster third-instar larvae is critical for the survival of individuals, as pupae will be exposed to many biotic and abiotic dangers while immobilized during several days of metamorphosis. Pupation site selection is sensitive to both genetic and environmental factors, but the specific mechanisms of this behavior still remain largely unknown. In this present study, we developed a semi-automatic systematic phenotyping approach to assay the variation on pupation height, measured as the distance between pupation site and food medium, in approximately 200 inbred lines derived from a wild Drosophila melanogaster genetic reference panel (DGRP2). Despite being such an essential trait, there is substantial variation on pupation site selection among D. melanogaster strains. We did not observe significant difference on pupation site selection between different sex in Drosophila melanogaster. Importantly, we found pupation site selection has a strong genetic component and identified multiple associating genetic loci. Identified candidate genes are enriched in those with high expression in larvae central nervous system, and may indicate the involvement of these genes in maintaining normal functions of the larvae nervous system. Our study not only identified candidate genes for pupation site selection behavior that is relevant to individual fitness, but also shed new light on the initial stage underlying the evolution of this behavior.

409T Admixture statistics for populations in space. *G. Bradburd* Integrative Biology, Michigan State University, East Lansing, MI.

Admixture is a fundamental process that shapes patterns of genetic variation within and between populations. Admixture can affect prediction of disease risk or an understanding of how natural selection operates, but it can be difficult to detect in empirical population genetic data. Recently, a series of statistics (*f*-statistics) have been introduced for identifying population admixture. These statistics assume that populations are related via a population phylogeny, and seek to identify departures from that tree-like null; these departures are interpreted as evidence for admixture. Though powerful, these *f*-statistics may therefore provide misleading results when relatedness between populations is better described by a spatial process, rather than a tree-like structure. Because isolation by distance (IBD) is expected to be common across many taxa, the tree basis of these *f*-statistics may be a serious limitation.

To address this concern, I propose a serious of spatial *f*-statistics, in which populations are assumed to be related via a spatial process, with nearby populations more closely related than more distant ones. After fitting a spatial model, admixture can then be identified against this spatial null. I apply these statistics, and demonstrate their utility, using a large sample of modern humans.

410T Development of Genomic-Based Strategies for Screening and Selection of Accesssions from a Carrot (*Daucus carota)* Germplasm Collection. *K. Corak*[1], S. Ellison[1,2], C. Luby[1], D. Senalik[1,2], M. Iorizzo[3], D. Spooner[1,2], I. Goldman[1], P. Simon[1,2], J. Dawson[1] 1) Department of Horticulture, University of Wisconsin- Madison, Madison, WI, USA; 2) Vegetable Crops Research Unit, US Department of Agriculture- Agricultural Research Service, Madison, Wisconsin, USA; 3) Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina, USA.

Germplasm collections are rich sources of both genetic and phenotypic diversity but are difficult to comprehensively screen for desired traits and/or alleles. To efficiently utilize the diversity present in large germplasm collections, plant breeders often attempt to identify a subset of accessions that represents the larger collection. Methods for creating these “core collections”

rely on partitioning collections into sub-clusters based on geographic, morphologic or neutral genetic similarity. These methods do not consistently capture functional diversity and may be inappropriate for highly admixed species. We are using a collection of domesticated carrot (*Daucus carota*) accessions to test genomic-based strategies that will allow breeders to create custom subsets of germplasm collections that maximize trait values of interest. Our preliminary work has established carrot as an appropriate species in which to study these strategies. We used a large dataset of genotyped cultivars and wild accessions to study the genetic structure of carrot germplasm resources available to breeders. We found a genetic distinction between Eastern and Western domesticated carrots, as well as between cultivated and wild accessions. Within Western cultivars, genetic diversity is present but there has been continuous gene flow and admixture. Principal component analysis and hierarchical clustering of Western domesticated carrot accessions based on the genetic differences between them support the conclusion that they form one large breeding pool. While partitioning the accessions according to either geographic, morphologic or genetic similarity resulted in core sets that adequately represented the whole, these cores did not differ significantly from a random sample. We plan to develop two genomic-based selection schemes that a) balance genetic and phenotypic diversity and b) incorporate genomic prediction models to identify interesting accessions. While these strategies will likely not identify subsets that maximize the diversity of the subset and are a departure from the traditional core collection concept, we expect that they will help us identify accessions and develop breeding populations that are more relevant for specific breeding goals.

411T Bayesian Inference of the Allelic Series in Multiparental Populations. *Wesley Crouse*, Samir Kelada, William Valdar Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Multiparental populations (MPPs) are experimental populations in which the genome of every individual is a random mosaic of a set of known founder haplotypes. Such populations provide distinct advantages for detecting quantitative trait loci (QTL) because tests of association between phenotypes and genetic variation can leverage inferred founder haplotype descent. Once a QTL is detected, however, further analysis is required to determine how the haplotypes group into distinct functional alleles, termed the allelic series (AS). We introduce a Bayesian framework for inferring the AS that takes into account sources of uncertainty found in typical MPPs, including individual haplotype states at the QTL, the number and composition of functional alleles, and the size of the allele effects.

We evaluate our method via simulation and consider two prior distributions for the AS, both derived from the coalescent: a default prior based on the Chinese restaurant process and an informative prior based on the local sequence phylogeny of the founder haplotypes. We find that posterior inference of the AS is generally uncertain even when the sample size and allele effects are large. Despite this uncertainty, our method still permits high-confidence allelic contrasts, and it improves effect estimation when the true number of alleles is small. We also find that leveraging phylogenetic information improves posterior certainty of the AS, effect estimation, and genetic signal, but that this is sensitive to misspecification of the phylogenetic tree. We apply our method to real data from the Collaborative Cross and the Drosophila Synthetic Population Resource, demonstrating that our approach provides a robust framework for inferring the AS and prioritizing downstream experiments of contrasting alleles.

412T Assessment of population sub-structure in genetic association studies. *N. DeVogel*, T. Wang. Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI.

STATEMENT OF PURPOSE: The purpose of this study is to explore statistical methodology for adjusting and hypothesis testing for population sub-structure (PS) in genetic association studies. Quite often, only the additive effects of PS are considered while the dominance effects of PS are ignored. In this study, we consider simultaneous adjustment and hypothesis tests of both the additive and dominance effects of PS. METHODS: For quantitative traits, PS can usually be adjusted or tested as random effects via a linear mixed model (LMM), where other covariates or candidate genetic variants can be treated as fixed effects. Using ancestral informative markers (AIM), the additive genomic relationship matrix can provide an estimate of the covariance matrix of the additive random effects of PS between individuals. Similarly, the dominance genomic relationship matrix can be treated as the covariance matrix of the dominance effects of PS between individuals. However, fitting this type of LMM and hypothesis tests of the variance components could be challenge. We propose a reformulated LMM based on the principal components of the additive and dominance relationship matrices with random regression coefficients. Methods for fitting this type of LMM will also be explored. In addition, both asymptotic and exact hypothesis tests on the variance components will be proposed. SUMMARY OF RESULTS: The proposed linear mixed effect models can incorporate both the additive and dominance genomic relationship matrices to adjust for the PS effects. The reformulated LMM can be easily fitted using standard statistical software. The proposed tests of variance components can also provide appropriate testing procedures for assessing the effects of PS in a quantitative trait based on the proposed LMM.

413T A multivariate linear mixed model for detecting GxE. *H. Kim*[1], J. Lovell[2], T. Juenger[3], S. Sen[1] 1) Biostatistics, Preventive Medicine, UTHSC, Memphis, TN; 2) Genome Sequencing Center, HudsonAlpha Institute for Biotechnology,

Huntsville, AL; 3) Integrative Biology, University of Texas at Austin, TX.
We develop a multivariate linear mixed model for detecting gene-environment interactions when there are many environments, and we have information annotating the environments. Our prototype example datasets are on segregating plant populations grown in multiple sites in multiple years. We will have information on the weather in each year as well as site-specific information such as latitude. The goal is to find QTLs that depend on latitude accounting for weather patterns that vary by year. We formulate a linear mixed model where traits can be correlated due to genomewide similarities (genetic kinship) and due to weather similarities ("climate kinship") between environments. We implement an efficient algorithm that uses an Expectation Conditional Maximization (ECM) algorithm in conjunction with an acceleration step. We tested the method for detecting gene by gene interactions in Arabidopsis recombinant inbred lines grown in Sweden and Italy for 3 consecutive years. In this dataset, our performance is comparable to univariate linear mixed models (FaSTLMM). Further evaluation of the method in a switchgrass dataset growin in 10 locations is in progress. Julia package implementing the methods is in development.

414T Realized heritability in panmictic populations - evolution of the concept. *M. Lstiburek* Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Praha 6, CZ.
We investigated the effect of phenotypic selection in offspring population of forest trees that originated in seed orchards. Seed orchard's pollen pool was subjected to external gene flow, thus the proportion of the siring pollen was the product of contamination. We formulated mathematically that gametic contribution of external pollen donors is reduced in the top-phenotypic subset of the offspring. This reduction (*r*) is a function of trait's narrow-sense heritability, level of external gene flow, selection differential of the seed orchard, and the selection intensity in offspring population (Lstibůrek et al. 2012).

We can utilize the above function as a standalone method in genetics to solve for the narrow-sense heritability. Assuming arbitrary population in H-W equilibrium, offspring population (F_1) is derived from the parental population (P) by random union of gametes. We then rank individuals in the P and F_1 populations by respective phenotype of a quantitative trait and then we evaluate the top-ranking phenotypic subset of the F_1 population and calculate the percentage of cases when one or both parents of the top-ranking offspring individuals belong to the corresponding phenotypic subset of the P population. This percentage corresponds to the *r* parameter in our earlier analysis. It follows that narrow-sense heritability can be estimated from the parameter *r* and the selection differential (Lstibůrek et al. 2018).

Our approach is demonstrated using a simulated population under two extreme scenarios, i.e., the narrow-sense heritability equal to 0 and 1, and graphically compared to a standard regression analysis.

The method is fundamentally different from existing experimental schemes as the heritability can be estimated from arbitrary phenotypic segments of the P and F_1 populations. It follows that the realized heritability can be calculated in panmictic populations with no artificial manipulation of the population structure (i.e., no selection).

Lstibůrek M., Bittner V., Hodge G.R., Picek J., and Mackay T.F.C. (2018). Estimating realized heritability in panmictic populations. Genetics 208: 89-95.

Lstibůrek M., Klápště J., Kobliha J., and El-Kassaby Y.A. (2012). Breeding without Breeding: effect of gene flow on fingerprinting effort. Tree Genetics & Genomes 8: 873-877.

415T Variance QTLs as a means to identify epistatic interactions in genome-wide association studies. *A.R. Marderstein*[1,2], E.R. Davenport[3], A.G. Clark[2,3] 1) Tri-Institutional Program in Computational Biology & Medicine, Weill Cornell Medicine, New York, NY; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.
Many human genome-wide association studies have focused on identifying the particular genetic loci that impact the phenotype. Although model organism studies have often been successful in identifying epistatic (GxG) and genotype-by-environment (GxE) interactions, few studies in human genetics have had adequate power. To address this gap, we screened for genetic associations with the variance of a phenotype (vQTLs), which can not only reveal direct genetic control over the

variance of a trait, but also a potential mean-based gene x gene (GxG) interaction or genotype x environment (GxE) interaction that underlies the variance association. Since exhaustive all-pairwise interaction testing is a computationally intensive task and decreases statistical power via multiple hypothesis testing correction, screening for vQTLs can provide a powerful inroad to discovering genetic interactions by generating a subset of loci that serve as promising candidates for an interaction, whether GxG or GxE. Using simulations, we found that two variants affecting the mean of a phenotype through an interaction are also highly likely to be associated with the phenotypic variance. We then adopted a sequential approach to identify epistatic interactions impacting gut microbiome composition in UK Twins, a cohort of over two thousand twins. First, genome-wide association studies with the variance of bacterial relative abundances were performed across 935 taxa. 1.3 million SNPs were tested using a two-step squared-residual linear mixed model approach. Second, to discover novel interactions with the phenotypic means, we tested significant vQTL loci in mean-based interaction models with all other variants. In this dataset, we discover 885 genome-wide significant interactions ($P < 5 \times 10^{-8}$) across the 935 microbial phenotypes. We identify a host genetic interaction between *SLC2A13,* which has been significantly associated with Crohn's disease and Parkinson's, and long intergenic non-coding RNA *LINC00877* impacting *P. copri* abundance, a taxon previously correlated with autoimmune disease and autism. In summary, vQTLs provide a practical approach to perform genome-wide, population-based analysis of epistatic interactions using human genome-wide association studies.

416T Relatedness and differentiation in arbitrary population structures. *A. Ochoa*[1,2], J.D. Storey[1,2] 1) Lewis-Sigler Inst, Princeton University, Princeton, NJ; 2) Center for Statistics and Machine Learning, Princeton University, Princeton, NJ.

Several important biomedical applications, including genome-wide association studies and heritability estimation for complex traits, require accurate modeling of the covariance structure of genetic variants. This dependence structure between individuals is parametrized by kinship coefficients, which are defined as the probability that random alleles are "identical by descent" (IBD). The fixation index "F_ST" is also an IBD probability that measures the overall population structure. My work is focused on extending current models and estimation approaches for kinship and F_ST to arbitrary population structures, where individuals are not assumed to belong to subpopulations that are disjoint, homogeneous, and statistically independent. African-Americans and Hispanics are two examples of complex population structures without independent subpopulations that require my novel approaches. I will first show how my approach improves upon previous approaches in real human datasets containing world-wide samples and in simulations where the true parameters are known. I then present my theoretical findings that several widely-used previous approaches yield biased kinship matrices and F_ST estimates, which return abundant negative values and other systematic distortions. My work led to a novel kinship and F_ST estimation framework with greatly improved accuracy, implemented in the R package 'popkin' available on CRAN. These results have direct implications in the estimation of heritability, association studies, and other analyses where population structure is a confounder.

417T A Bayesian approach to quantitative genetics for high-dimensional traits. *D. Runcie*[1], J. Ta[1], L. Crawford[2], S. Mukherjee[3] 1) Plant Sciences, University of California Davis, Davis, CA; 2) Department of Biostatistics, Brown University, Providence, RI; 3) Statistical Science, Duke University, Durham, NC.

Statistical models for Genome-Wide Association Studies, QTL analysis, and Genomic Prediction, are the foundation of modern quantitative genetics and crop improvement. Driven by the explosion of whole-genome genotype data, recent improvements to these models allow for analyses of millions of markers at a time. However, similar advances for modeling large phenotype datasets is lacking. New phenotyping technologies collect thousands of observations on each individual plant or line – changes in morphology through time, molecular phenotypes such as gene expression or metabolite levels, or performance measures across multiple environments. Jointly modeling these high-dimensional traits can provide insight into developmental and physiological mechanisms that link genotype and phenotype. We propose a robust and efficient method for modeling the genotype-phenotype relationship of high-dimensional traits. The key idea underlying our model is that groups of traits will be highly correlated due to genetic and developmental pleiotropy. We leverage these correlated modules to prioritize the most important signals in big data. We will demonstrate how our method provides powerful and interpretable estimates of genetic architecture using two high-dimensional datasets: a time-series analysis of growth curves, and a dataset of genome-wide gene expression.

418T The contribution of homeologous gene interactions to the genetic variance of allopolyploid wheat. *N. Santantonio*[1], J. Jannink[1,2], M. Sorrells[1] 1) Plant Breeding and Genetics, Cornell University, Ithaca, NY; 2) Robert W. Holley Center for Agriculture & Health, USDA-ARS, Ithaca, NY.

The sub-genomes of an allopolyploid will each contain complete, yet evolutionarily divergent, sets of genes. With the availability of affordable genome-wide markers, breeders of allopolyploids now have the opportunity to manipulate individual sub-genomes and investigate interactions of homeoalleles across sub-genomes. We present theory and a statistical framework for partitioning genetic variance and predicting breeding values for each sub-genome and their inter-genomic interactions. Using an allohexaploid wheat breeding population for demonstration, sub-genome main effects and interactions were fit using multi-kernel mixed models for variance component estimation and genomic prediction. Strictly modeling inter-

genomic interactions resulted in equivalent increases in genomic prediction accuracy as modeling all pairwise marker interactions. Using the IWGSC RefSeq v1.0 wheat genome sequence, 18,184 triplicate and 5,612 duplicate homeoallelic gene sets were identified and anchored to the nearest GBS marker, forming 10,172 unique sets of homeologous markers. Homeologous marker interactions for each homeoallelic marker set were used to predict whole genome breeding values, as well as estimate homeologous main and interaction effects. Using gain in genomic prediction accuracy as a proxy for importance of marker interactions, we show that homeologous marker interactions can explain up to 60% of the additional genetic signal from the additive model. Negative relationships observed between homeologous marker main effects and interaction effects point to a pattern indicative of homeoallelic subfunctionalization. Homeoallelic sites exhibited a higher degree of linkage disequilibrium than other sets of loci across chromosomes, suggesting selection is acting to fix important homeoallelic combinations. Thus, we provide new tools for breeders of allopolyploid crops to characterize the genetic architecture of existing populations, determine breeding goals, and develop strategies for selection of sub-genome additive effects and inter-genomic epistasis.

419T Regulatory network inference from a dense time-course RNA-seq study of *Drosophila* innate immune response. *M.F. Schlamp*[1], A. Early[1], M.T. Wells[2], S. Basu[2], A.G. Clark[1,2] 1) Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Biological Statistics and Computational Biology, Cornell University, Ithaca NY.

Following microbial infection, *Drosophila* launch a rapid and efficient immune response that is crucial to survival. However, these responses are costly for the organism, consuming energy and resources that could be used for other life processes such as metabolism, reproduction, and environmental stress responses. Therefore, organisms must tune their immune response to strike a balance between the advantage of a rapid and robust ability to fight infection and the costly side-effects of an over-prolonged or unnecessary immune response. The dynamics of this immune response have not yet been studied at high temporal resolution. Discovering and reconstructing gene regulatory networks from high-throughput time-course gene expression data constitutes a key problem in functional genomics and bioinformatics. Common statistical methods to analyze static RNA-seq data are not ideal for time-course RNA-seq data, since they do not take into consideration the correlations of genes across previous and subsequent time points. New approaches for analyzing time-course data are pivotal to reveal dynamic behaviors in organisms and discover regulatory interactions among genes. In this study, we perform a dense time-course RNA-seq analysis of the *Drosophila* immune response to learn more about the dynamics of activation and shutdown of the innate immune response. Flies were injected with commercial lipopolysaccharide, a known non-pathogenic elicitor that can stimulate a robust yet transient immune response, while avoiding the confounding effects from a growing and changing internal population of pathogens. Flies were sampled for RNA-seq analysis pre-infection as a control and post-infection for 20 time points throughout 5 days. We used gene-wise linear models to fit polynomial trends with time, and standard empirical Bayes F-tests to select genes whose expressions altered significantly across the time course. Clustering analysis on this subset of differentially expressed (DE) genes show strong temporal patterns of early and late induction of immune processes, as well as transient and sustained responses to infection. We also constructed networks of bivariate and multivariate Granger causality (GC) relationships among this subset of DE genes. GC relationships present in these networks point to several novel interactions governing temporal gene regulation of the immune response and trade-offs between immune response, metabolism, and reproduction.

420T Effect of fitness landscape, population structure and linkage disequilibrium on the detection of local adaptation. *Bertrand Servin*, Claude Chevalet Institut National de la Recherche Agronomique, Laboratoire de Génétique, Physiologie Systèmes d'Elevage, Castanet-Tolosan, France.

Detecting genetic factors that have been involved in the adaptation of populations is a fundamental question in population and evolutionary genetics. Several methods aimed at detecting such loci have been proposed to leverage the rich information provided by genomic datasets acquired in multiple populations. While some approaches only exploit genetic information, others aim at testing for adaptation to a particular covariate, being a trait or an environment. We will introduce two new statistics to detect adaptation to a covariate one of which leverages information on Linkage Disequilibrium. Using both simulated and real data we study the statistical power of these new as well as previous methods to detect the effect of a QTL on populations evolving under polygenic adaptation. While we confirm that using covariate information usually increases the power to detect adaptation, we show this is not always the case as some scenarios exist where the combined effects of structure and fitness landscape greatly limit their performance. In addition to introducing new adaptive tests, our results can help in the interpretation of genome-wide scans for selection and the sampling design of future studies.

421T Genetic variation and ancestry in *Helicobacter pylori* and gastric cancer susceptibility in Central America. *Gloria Tavera*[1], Roberto Dominguez[3], Jaquelaine Bartlett[1], Barbara Schneider[2], Dawn Israel[2], Juan Romero-Gallo[2], Robertino Mera[2], Blanca Piazuelo[2], Richard Peek[2], Keith Wilson[2], Scott Williams[1], Doug Morgan[2] 1) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Ingram Cancer Center, Division of Gastroenterology, Hepatology, and Nutrition, Vanderbilt University, Nashville, TN; 3) Hospital de Occidente, Santa Rosa de Copan, Copan, Honduras.

Helicobacter pylori, a chronic gastric pathogen that infects more than half of all humans, is the principal cause of gastric cancer, the second leading cause of cancer mortality worldwide. A least half of the world's population is infected with *H. pylori* and approximately 85% in our study populations, yet only 1-3% of those infected develop Cancer. Whether *H. pylori* acts as commensal bacteria or a pathogen is strongly dependent upon as yet unknown host-bacterial genetic factors. Our study is designed to discover what motivates the progression to cancer in patients with *H. pylori* infection, in a patient population disproportionately impacted by gastric cancer mortality. As part of an ongoing, population-based, case-control gastric cancer initiative in the high-incidence region of western Honduras, *H. pylori* was successfully cultured in 95 of 136 patients (70%), with gastric histology distributed across the gastric cancer cascade. *H. pylori* whole genome sequences were obtained via Illumina HiSeq 2500 Rapid Run. Sequences were aligned to a reference genome using Bowtie 2 and GATK and genetic ancestry was determined using multi-locus sequence typing (MLST), using STRUCTURE, with an admixture model. High quality genomic data was available for 89 subjects, with a mean age 50 (SD +/-15) and 55% male. Gastric histology included 9.0% atrophy, 19% intestinal metaplasia, and 10.7% adenocarcinoma. Nearly all *H. pylori* strains were CagA+ and VacAs1m1+ (99% and 91%, respectively). Using MLST data from 5 diverse global regions (n=746), and assuming 4 distinct ancestral groups in the Honduran samples, we were able to show that these 89 H. pylori had mixed European and African ancestry. The finding of African genetic ancestry in *H. pylori* strains, in conjunction with presumed Mayan host genetic ancestry, may support the role of disrupted coevolution as a contributor to the severity of gastric disease in the high-incidence gastric cancer regions of Central America. Further characterization of *H. pylori* and host genetic ancestry may help identify patients at high risk for developing gastric cancer.

422T Pedigree Analysis of a large dog population. *S.R. Urfer*[1,2], M. Kaeberlein[1,2], K.E. Creevy[2,4], A. Steiger[5], C. Gaillard[5], D.E.L. Promoslow[1,2,3] 1) Medicine Pathology, University of Washington, Seattle, WA; 2) Dog Aging Project, University of Washington, Seattle, WA; 3) Biology, University of Washington, Seattle, WA; 4) Clinical Veterinary Medicine, TAMU, College Station, TX; 5) Institute of Genetics, University of Bern, Switzerland (emeritus).

The purebred domestic dog is an interesting quantitative genetic model due in part to the availability of extensive pedigree and phenotype data. While a number of canine quantitative genetic pedigree studies exist, such studies are generally based on data across relatively few generations, even though systematic breeding started during the second half of the 19th Century in most modern breeds. Here we describe analysis of a pedigree database of 59,833 purebred Irish Wolfhounds born between 1845 and 2017, covering 69 generations, and including geographic origin, age and censoring data on 5,048 dogs. We analyzed these data using R, Pedigree Explorer, and the PEDIG software package.

We found that the modern Irish Wolfhound breed is based on ~30 founders born between 1845 and 1862. The maximum effective population size of 15.46 was reached between 1920-29. Effective population size remained fairly constant until 1960 and then decreased steadily, reaching 6.82 for 2008-17. Over the same period, fewer individual ancestors accounted for increasing amounts of genetic variability: In dogs born 1920-29 the most important ancestor accounted for 13.13% of genetic variability, while it accounted for 27.48% in dogs born 2008-2017. Both of these findings could be traced to two major bottleneck events between 1939 and 1954; taken together, the two bottleneck ancestors accounted for 46.86% of genetic variability in dogs born 2008-17. This was accompanied by an increase in average inbreeding coefficients over time; however, when considering inbreeding over 10 generations, this effect was masked due to exponential population growth starting around 1965.

Using Kaplan-Meier Analysis, we found that females lived significantly longer than males (P=3E-8), and that dogs in North America and Ireland lived significantly longer than in Great Britain and Continental Europe (P=9E-11). Interestingly, dogs from 1951-2000 lived significantly longer than both dogs from 1845-1950 and dogs from 2000-2017 (P

To our knowledge this is the first pedigree analysis of a modern dog breed going back to the mid-19th Century. Our data show the feasibility of such a study and illustrate some potential issues with pedigree analyses using fewer generations.

423T Integrated software packages of map construction and QTL mapping in various kinds of genetic populations. *J. Wang*, H. Li, L. Zhang, L. Meng Institute of Crop Science, CAAS, Beijing, CN.

In past ten years, we have developed three integrated software packages for linkage map construction and QTL mapping. QTL IciMapping was designed for various bi-parental populations, GACD for clonal F_1 and double cross, and GAPL for pure-line population from multiple parents. The three packages are freely available from http://www.isbreeding.net/. QTL IciMapping has nine functionalities. (1) AOV: Analysis of variance for multi-environmental phenotyping trials; (2) BIN: Binning of redundant markers; (3) MAP: Construction of genetic linkage maps in biparental populations; (4) CMP: Consensus map construction from multiple genetic linkage maps sharing common markers; (5) SDL: Mapping of segregation distortion loci in biparental populations; (6) BIP: Mapping of additive and digenic epistasis genes in biparental populations; (7) MET: QTL by

environment interaction in biparental populations; (8) CSL: Mapping of additive and digenic epistasis genes with chromosome segment substitution lines; and (9) NAM: QTL mapping in nested association mapping populations. GACD has three functionalities. (1) BIN, binning of redundant markers; (2) CDM, construction of female, male and combined linkage maps; and (3) CDQ, mapping of additive and dominance genes. GAPL has three functionalities. (1) BIN, binning of redundant markers; (2) PLM, construction of linkage maps; and (3) PLQ, mapping of additive genes. These software packages are first genetic analysis tools which can conduct map construction and QTL mapping simultaneously. The latest programming technology and concept of project was adopted, to increase their user-friendliness. In a project, users can easily load input data, use the functionalities, review the outputs, and use some outputs as inputs of other functionalities. The integration of the software provides a seamless pipeline from treatment of redundant markers, to estimation of recombination frequency, to building of linkage maps, and finally to mapping of quantitative trait genes. Most genetic populations from self-pollinated, cross-pollinated and clonal species can be analyzed by the three packages.

424T Ancestral life history and natural genetic variation shape starvation survival in *C. elegans*. *Amy K. Webster*, James M. Jordan, Jonathan D. Hibshman, Brad T. Moore, Rojin Chitrakar, Rebecca E. W. Kaplan, Anthony Hung, Megan Jiao, L. Ryan Baugh Duke University Biology Department, Durham, NC.

Essentially all organisms face fluctuations in food availability in the wild. In response to these fluctuations, the development of the nematode worm *C. elegans* is tightly coupled to food availability. In the presence of low food and high population density, worms enter an alternative larval stage called dauer diapause which enables them to survive starvation. While the worm life cycle typically takes only three days when food is plentiful, worms can remain in dauer diapause for months. Once conditions improve, worms exit dauer diapause to become reproductive adults. Aside from dauer diapause, worms can simply arrest development during larval stages in the complete absence of food and survive in the arrested state for weeks. Modulation of the ability to survive starvation during larval arrest is of particular interest because the trait is closely linked to fitness. Historically, loci important for survival have been mainly identified using mutants in isogenic populations. However, few studies have taken into account the factors that may influence starvation survival in the wild, including 1) ancestral life history and 2) natural genetic variation. Here, we present evidence that isogenic worms starved long-term during dauer diapause produce fewer, more variable progeny that die quicker during starvation. However, these worms produce great-grandprogeny that survive starvation better, suggesting transgenerational epigenetic memory of life-history traits. In parallel, we have used a population-based sequencing approach with wild isolates of *C. elegans* and identified natural variation at particular genomic locations associated with the ability to survive starvation. We are following up on this by generating nearly isogenic lines. Our data show that both ancestral life history and genetic variation are important for starvation survival, and we aim ultimately to understand the interaction of these factors in shaping this fundamental trait.

425T Inferring parental genomes and parent-of-origin using genotypes from siblings. S. Basu-Roy[1], John Blangero[2], *A.L. Williams*[1] 1) Biological Statistics & Computational Biology, Cornell University, Ithaca, NY; 2) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX.

Children inherit two chromosome copies, one from each parent, with both formed via recombination. While each child inherits only half of the genomes of each parent, independent segregation and recombination are randomized such that n siblings will inherit on average a proportion of $1\text{-}1/2^n$ of both parents' genomes. Analyses of sibling data can recover (partial) parental haplotypes, yet which parent—mother or father—each pair of haplotypes belongs to is ambiguous. Moreover, reconstruction of each chromosome is independent, leading to $2^{22} > 4$ million possible parental haplotype configurations in humans.

Male- and female-generated crossovers differ in location and frequency, a fact we exploit to infer which parent carried each haplotype. Specifically, we use a hidden Markov model to analyze sibling data, inferring the parental haplotypes (without determining which parent each belongs to) together with the positions of crossover events. Crucially, since crossovers only occur between haplotypes carried by a single parent, in most cases, linkage enables the partitioning of the haplotypes and all their associated crossovers into two sets corresponding to the two parents. Next, from the two sets of crossovers transmissions, we compute the joint likelihood of the crossovers being transmitted by a male or female parent, determining the probability of each parental assignment. We adopt a Poisson model of crossover with the probabilities of the number and location of the events based on Morgan positions in sex-specific genetic maps.

We evaluated this framework using 69 families from the San Antonio Mexican American Family Studies that include data for three or more siblings and genotypes for 918,917 SNPs. Initially, we performed an analysis using two families with 12 and 10 children and find that the true parent assignments are 1,000-times more likely than the alternative (LOD > 3). To learn about the feasibility of the approach for smaller families, we phased the siblings with parental data and inferred the parent configuration using the resulting two sets of transmitted crossovers (avoiding potential errors in phase). We find that 80% of the 1,518 analyzed chromosomes also showed high confidence very (LOD > 2) while only 4.5% were inferred incorrectly.

These results are likely to be representative of inference quality in the presence of joint family- and population-based phasing, and will open the door to novel analyses such as parent-of-origin testing even without parent data.

426T Genetically-informed association analysis identifies risk factors for late-onset Alzheimer's disease. *D. Yan*[1], Q. Lu[2] 1) University of Wisconsin - Madison, Madison, WI; 2) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA.

Genome-wide association studies (GWAS) have provided great insights into the genetic basis of complex human diseases and traits. Increasingly accessible GWAS summary statistics and biobank data have made it possible to further dissect the complex relationship between traits. Here, we introduce GAACT (Genetically-informed Association Analysis for Complex Traits), an efficient method to identify associations between a disease of interest and thousands of genetically-imputed traits using GWAS summary statistics. Compared to traditional association analysis, GAACT does not require traits to be measured on the same individuals and effectively removes reverse causation from associations. We applied GAACT to more than 2,000 traits from the UK Biobank (N~337,000) and summary statistics for late-onset Alzheimer's disease (LOAD) from the International Genomics of Alzheimer's Project (IGAP; N=54,162). We replicated our findings in an independent dataset from Alzheimer's Disease Genetics Consortium (ADGC; N=7,050). Results from two stages were then meta-analyzed. We identified a total of 72 traits that reached Bonferroni-corrected significant associations with LOAD and showed consistent effect directions in two independent analyses. Unsurprisingly, family history of dementia was strongly associated with a higher LOAD risk (p=2.0E-77 and 1.6E-35 for illnesses of mother and father, respectively). Diagnosis of vascular dementia (p=2.2E-12) and cognitive impairment (p=2.5E-06) were also significantly associated with LOAD. In addition, traits for intelligence and education showed consistent associations with LOAD. Fluid intelligence score (p=6.5E-17) and age completed full-time education (p=1.8E-08) were strongly associated with a lower LOAD risk. College (p=1.4E-13), A levels (p=3.6E-12), and O levels education (p=1.4E-06) were negatively associated with LOAD while CSE (p=9.7E-06) and other education (p=4.5E-09) were associated with a higher risk. High cholesterol (p=3.2E-14) and usage of cholesterol-lowering medication (p=6.2E-06) were positively associated with LOAD. Leukemia and lymphoma traits also showed highly consistent associations. Chronic (p=4.7E-08) and acute lymphocytic leukemia (p=3.5E-06), self-reported leukemia diagnosis (p=8.3E-06), and Hodgkin's disease (p=5.1E-08) all reached significance in our analysis. These results demonstrate the effectiveness of GAACT and provide novel insights into LOAD etiology.