PEQG26

Population, Evolutionary and Quantitative Genetics Conference

ABSTRACT BOOK

June 9–12, 2026
Pacific Grove, CA | #PEQG26

GSA | GENETICS G3

1 Coalescent structure through time: A non-parametric approach

John Novembre University of Chicago

The patterns of biological variation we observe today, from deep phylogenetic splits to isolation by distance, arise from how genetic ancestry is shared among individuals. Recent progress in the inference of ancestral recombination graphs (ARGs) provides new opportunities to understand genetic ancestry with unprecedented detail; but it remains challenging to uncover the layers of genetic structure that ARGs can in principle reveal. In this talk, I will share recent progress from my group in this area, including a new conceptual framework that provides a flexible, individual-based, time-varying view of genetic ancestry. A feature of this approach is that it emphasizes variation in the rates at which genetic ancestors coalesce in a way that reflects the consequences of population size change and migration without requiring pre-defined demographic models. Using simulations, I will demonstrate the approach's utility and subtleties in interpreting its outputs. Finally, I will share applications of the method to chimpanzees, bonobos, and humans, illustrating how the method can provide new resolution on time-varying genetic ancestry in these species.

2 Visualizing the shared nature of human genetic variation

James Kitchens, Ally Angel-Clark, Graham Coop University of California, Davis

Genetics education can inadvertently lead students to an overly simplified view of the human genetic variation and the differences between groups of people. Therefore it is important to include conversations around genetics and race within the curriculum. However, many visualizations of human population structure, which are useful in research context, are readily misinterpreted in the classroom and in broader public outreach. Here, we present an Euler diagram based visualization technique which better captures the shared nature of common human genetic variation. We applied this approach to samples from the 1000 Genomes Project (1KGP) and measured the interpretability of these diagrams using an online, out-of-class activity (IRB approved) given to undergraduate students in evolutionary biology and human genetics classes. We find that, after a brief introduction to the figures, students can reasonably estimate the percentage overlap shown by a diagram (+ or - 15%), supporting that such figures are readily interpretable. Prior to a class discussion on the topic, when students were asked to estimate the percentage overlap in common genetic variation between two sample groupings from the 1KGP, their estimates generally fell well below the true percentage overlap – a bias that has been similarly observed across other previous studies. Illustrating this bias in students' expectations then provided a useful basis for in-class discussion and explaining the high degree of overlap in human variation amongst groups. We show how Euler diagrams and online activities offer an approachable technique for demonstrating the scale of genetic diversity across humans and can be easily incorporated into an introductory genetics curriculum.

3 Building an AI-assisted, Quarto-based digital course companion textbook to support a lab-based advanced bioinformatics course

Laurie S Stevison Auburn University

Bioinformatics courses must keep pace with rapidly evolving tools and data types while supporting active hands-on learning. In a combined upper-division undergraduate and graduate bioinformatics course centered on a semester-long research project, I developed an AI-assisted, Quarto-based digital course companion textbook that replaces both a commercial textbook and various supplemental resources typically shared with students via Canvas Modules. The textbook is authored in markdown within RStudio and rendered as a password-protected website, remaining free to students while giving full instructor control over content, sequencing, and formatting. Each chapter consolidates foundational textbook material, from a textbook last published in 2015, with up-to-date primary literature into concise narratives introducing bioinformatics concepts and workflows, with embedded citations to source material. End-of-chapter prompts explicitly connect in-class lecture content to students' ongoing research projects and foster metacognitive reflection. A dedicated lab manual section provides step-by-step, code-focused protocols. Large language models accelerate drafting and iterative refinement of narrative, figures, and discussion prompts, which I curate and integrate through Quarto's reproducible publishing pipeline. The site prioritizes accessibility compliance, including alt text, semantic heading structure, and color contrast. Because materials are version-controlled and mirrored on GitHub, they remain portable across institutional moves and enable collaborative development without sacrificing instructor ownership. This infrastructure supports rapid, low-friction updates in response to new software tools, evolving file formats, and lessons learned from each cohort—avoiding the multi-year lag and recurring costs of textbook adoptions. While the initial development required significant effort, semester-to-semester reuse and refinement of chapters and labs is substantially less time-consuming than repeatedly sourcing new commercial textbooks as the field evolves. The companion significantly improved alignment between lecture, lab, and the semester project, and reduced student confusion about where to find course information. My experience demonstrates that AI-assisted, Quarto-based materials provide a lightweight yet robust alternative to LMS-centered course design in computationally intensive life science courses.

4 The Genomics Education Partnership: Leveraging comparative evolutionary genomics to engage biology students through course-based undergraduate research experiences

Laura Reed[1], Wilson Leung[2], Logan Katie Cohen[3], Amanda Stalke[4] [1]Biological Sciences, The University of Alabama, [2]Washington University in St. Louis, [3]The University of Alabama, [4]Colorado Mesa University

Founded in 2006, the Genomics Education Partnership (GEP) is a nationwide collaboration of over 260 faculty from 200+ institutions that engage several thousand students annually in Course-based Undergraduate Research Experiences (CUREs). GEP provides a free web-based platform (thegep.org) with curated curriculum that can easily be incorporated into existing courses, making student research experiences broadly accessible. Students participate in projects such as characterizing molecular evolution of genes in a specific pathway (e.g., insulin signaling, oxidative stress), analyzing the evolution of novel venom proteins in parasitoid wasps, identifying candidate genes for selective conservation breeding in the endangered Puerto Rican Parrot, and distinguishing genes involved in the evolution of introduced biocontrol beetles. Students learn to leverage multiple, often conflicting, lines of evidence from related reference species (e.g., *Drosophila melanogaster*), expression data (e.g., RNA-Seq), gene prediction algorithms, evolutionary conservation, and basic molecular biology rules to construct gene models best supported by the available evidence. For example, students participating in the Pathways Project use comparative analysis of genomic neighborhoods (microsynteny) and sequence conservation to identify and create models of gene orthologs. Students interact with the evidence using a custom mirror of the UCSC Genome Browser, thereby enabling students to focus on conceptualizing fundamental properties of the genome, at the level of computational experience typically found in undergraduate students, instead of the mechanics of bioinformatics analyses (e.g., learning the command-line). The GEP also provides web-based tools to allow students to verify and correct their proposed gene models, thus students learn the critical element of iteration in research, and to experience growth through formative frustration. Some students proceed to explore more complex evolutionary genomic research questions following their initial research experiences. The gene models are also collated to test evolutionary genomic hypotheses (e.g., test whether the rates of molecular evolution are correlated with the connectivity of a given gene within biological networks). The GEP's approach to CUREs provides a scalable and cost-effective way to make genomics research accessible to all institutions and students. Supported by NIH R25GM160660 and R25GM130517, and NSF 1915544 to LKR.

5 Rhapsody of Genetics: An Algorithmic Composition System that Converts Genetics into Music and Web-Based Browser for Genomics Education

Siyuan Feng[1], Daria Tennikova[2] [1]Department of Genetics, University of Wisconsin Madison, [2]Mead Witter School of Music, University of Wisconsin - Madison

Genetic sequences encode the fundamental blueprint of life, yet this data remains abstract and inaccessible to non-specialists. While sonification has been proposed to bridge this gap, existing methods often rely on arbitrary mappings that result in monotonic, unmusical outputs, failing to capture the structural richness of biological molecules. We present "Rhapsody of Genetics," an automatic compositional system and web-based platform that renders the invisible language of genetics into an immersive, scientifically accurate auditory experience, and transforms the teaching of genomics through multimodal engagement.

We developed a multi-dimensional orchestral composition engine that uses measurable biochemical properties, such as amino acid hydropathy, DNA molecular weight, and hydrogen bonds, to determine musical parameters. These biological characteristics directly influence rhythm, choice of percussion instruments, relationship between the pitch classes, and instrumentation, allowing genes to generate their own distinctive musical identities and listeners to intuitively perceive genetic sequences and their molecular characteristics.

We then assembled the composition engine into the world's first interactive Genetic Music Browser, a web interface that parses user-selected genes to generate multi-track orchestral compositions. By synchronizing audio playback with a visualization of the DNA and amino acid sequences, the browser serves as a powerful new medium for public engagement. It demonstrates how computational biology and algorithmic composition can converge to make complex scientific data not only perceptible but emotional, offering a scalable tool for communicating the complexity of the genome to diverse audiences, including K-12 and undergraduate students.

6 Elevated mutation near crossovers inhibits the evolution of recombination

Bret A. Payseur[1], Sarah P. Otto[2] [1]University of Wisconsin-Madison, [2]University of British Columbia

Recombination diversifies offspring genomes and helps ensure chromosome segregation during meiosis. Mutation rates are elevated near crossovers due to the induction of double-strand breaks and their imperfect repair, a byproduct of recombination typically ignored by theory designed to explain its evolution. To examine the evolutionary role of mutagenic recombination, we analyze a population genetic model in which a modifier locus controls both the rate of recombination between two loci experiencing viability selection and the rate of mutation at those loci. Analytical and numerical results demonstrate that the advantage of recombination conferred by its capacity to accelerate removal of epistatic, deleterious variants is overcome by the selective cost of even small increases in the mutation rate. Our findings suggest that higher recombination evolves by altering steps later in the crossover pathway that are less likely to inflict mutational damage. These results underscore the potential for models that consider the cellular and molecular features of recombination to alter our view of how this process that governs inheritance evolves.

7 The repeated evolution of hybrid melanoma across swordtail fish

Kelsie E Hunnicutt, Molly Schumer Department of Biology, Stanford University

Hybridization between closely related species is a common process in the evolutionary history of many species, and as a result, the genomes of modern species are a mosaic of regions derived from past hybridization. Species with admixed genomes must contend with the negative consequences that can arise from mixing two divergent genomes, such as the exposure of genes or loci that do not interact properly in hybrids. Uncovering the evolutionary forces that drive the formation of these “hybrid incompatibilities” is crucial to understanding the function of modern genomes, but we have rarely been able to identify the genetic architecture of vertebrate hybrid incompatibilities. Here, we investigate the evolution of a repeatedly evolved hybrid incompatibility in *Xiphophorus* swordtail fish where hybrid offspring from develop malignant melanoma from pigmentation spots that are benign in parental species. In multiple species pairs, hybrid melanoma is thought to be driven by misregulation of the oncogene *xmrk* and the tumor suppressor genes that normally control *xmrk* in parental species. While *xmrk* has long been implicated in hybrid melanoma, the identities of tumor suppressor genes remain largely unknown. First, we test the hypothesis that hybrid melanoma has repeatedly arisen across *Xiphophorus* due to the breakdown of interactions between the *xmrk* gene and lineage-specific tumor suppressor genes. Through genetic mapping crosses across multiple species pairs, we find that melanoma severity is determined in all cases by *xmrk* but by different interacting repressor genes in each species pair. Next, we investigate the evolutionary history of *xmrk* across swordtails using long-read whole genome and transcriptome sequencing to characterize structural variation and allelic diversity. We demonstrate that the flanking regions of *xmrk* across *Xiphophorus* are enriched in structural variants compared to other genomic regions and that repetitive DNA content is elevated around *xmrk,* consistent with the hypothesis that higher rates of ectopic recombination and genomic instability may drive the exceptional levels of structural variation in this region of the genome. This research is one of the first investigations into the mechanisms that govern vertebrate hybrid incompatibilities beyond individual pairs of species and provides key insights into the genetic and evolutionary mechanisms underlying hybrid incompatibilities and their consequences for genome evolution.

8 Following the process of sunflower domestication through space and time with ancient DNA

Nathan Wales[1], Julia Kreiner[2], Melis Akman[3], Peter Stokes[4], Keric Lamb[4], Hsiao-Lei Liu[5], Kate Christensen[4], Bruce Smith[5], Kristen Gremillion[6], David Lentz[7], Logan Kistler[5], Tom Gilbert[8], Benjamin Blackman[4] [1]Archaeology, University of York, [2]Ecology & Evolution, University of Chicago, [3]Decibel Bio, [4]Plant & Microbial Biology, University of California, Berkeley, [5]National Museum of Natural History, [6]Anthropology, Ohio State University, [7]Biological Sciences, University of Cincinnati, [8]Globe Institute, University of Copenhagen

Where and how many times crop plants were domesticated, how quickly and severely genetic diversity is lost during domestication, and how domestication syndromes are assembled through polygenic evolutionary change are questions of active research and frequent debate. We are addressing these questions with archaeological DNA approaches in the common sunflower, *Helianthus annuus*. Native American farmers living ~5000 years ago began transforming the common sunflower from a highly branched wild plant with small disks and seeds into a staple oilseed crop that sports a single large head with large seeds on an unbranched stalk. We have assembled a time series of 69 archaeological samples spanning from ~3800 to ~400 years before present and obtained endogenous DNA sequences from these samples through whole-genome sequencing and sequence capture approaches, revealing how human cultivation altered sunflower genetic diversity through time. Combining these data with available and new datasets for wild *H. annuus* and extant traditional varieties, we find that sunflower’s cultivation history involved shifting agricultural practices over this period, as revealed by multiple genetic bottlenecks, major changes in crop-wild gene flow over time, and intriguing patterns of organellar haplotype turnover. Through obtaining DNA sequence from an archaeological sample identified as cultivated sunflower that was recovered at a Mexican site as well as from herbarium records of several wild Mexican sunflower relatives, our analyses also resolve a longstanding controversy about where and how many times sunflower was domesticated. Through targeted resequencing of ~425 candidate domestication genes we defined through population genomics and transcriptomics approaches with extant germplasm, our archaeological times series reveals the timing and order of selective sweeps, providing insight into how the sunflower domestication syndrome was assembled by Native American farmers through time. Finally, ~1300 years before present, we observe concurrent shifts in morphology of the archaeobotanical remains, the extent of gene flow from wild *H. annuus* into cultivated sunflower, and the frequency of selective sweeps. We interpret these temporally coinciding shifts as signatures of a transition into a period of more relaxed selection, likely reflective of reduced cultural connectivity and/or the rapid expansion of maize agriculture in Eastern North America at this time.

9 Comparing the evolutionary consequences of large regions of suppressed recombination in Mimulus and Panicum

David B. Lowry Plant Biology, Michigan State University

Variation in meiotic recombination is a fundamental component of the genetic architecture of adaptive traits. While chromosomal inversions are well known to locally suppress recombination and facilitate adaptation, other genomic features, especially pericentromeric regions, also experience extreme suppression of recombination and frequently harbor loci underlying ecologically and agronomically important traits. Despite their prevalence, we lack a unified framework for understanding how different sources of suppressed recombination shape genotype–phenotype relationships, linked selection, and evolutionary responses to selection. Here, I present recent and ongoing work in *Mimulus* (monkeyflowers) and *Panicum* grasses that integrates functional genetics, quantitative trait mapping, and population genomics to contrast the evolutionary consequences of recombination suppression caused by chromosomal inversions versus pericentromeric regions. In *Mimulus*, I focus on a well-characterized locally adaptive chromosomal inversion and describe progress toward functional molecular validation of candidate genes underlying adaptive phenotypes. Leveraging recent advances in transformation efficiency, this work directly tests the supergene hypothesis by evaluating whether adaptive effects arise from the combined action of multiple linked loci or from single major-effect genes. These efforts address long-standing questions about additivity, epistasis, and the mechanistic basis of inversion-associated adaptation. In *Panicum*, I present quantitative trait locus analyses showing that multiple disease-resistance loci colocalize with large pericentromeric regions of suppressed recombination. These results reveal parallels with inversion-associated architectures while highlighting key differences. Unlike inversions, pericentromeric regions remain recombinationally inert even in homozygous states, with important implications for Hill–Robertson interference, the efficacy of selection, and long-term genome evolution. By integrating these empirical results with emerging theoretical expectations, I compare how distinct forms of recombination suppression influence the maintenance of adaptive variation and the predictability of evolutionary outcomes. Together, this work positions *Mimulus* and *Panicum* as complementary systems for understanding how variation in recombination rates and genome structure jointly shape quantitative traits and adaptive evolution.

10 The molecular axis of endemism is governed by the opposing forces of purifying and balancing selection

Alan Bergland[1], Luke Proud[2], Jingqi Liu[1], Julia Beets[3], Dylan Hyland[1], Emily Longman[2], Paul Schmidt[4], Thomas Flatt[5], Joaquin Nunez[2] [1]University of Virginia, [2]University of Vermont, [3]Vrije Universiteit Amsterdam, [4]University of Pennsylvania, [5]University of Friboug

Understanding the forces that shape genetic variation in nature is a central goal of evolutionary biology. The prevailing “Standard Model” of population genetics posits that genome-wide variation is governed primarily by mutation-selection balance, with genetic drift and purifying selection as dominant forces, adaptation leading to selective sweeps as an occasional process, and balancing selection occurring only rarely. Here, we leverage population genomic data from many eukaryotic species sampled across their range to provide a novel test this of model by focusing on the prevalence of synonymous and non-synonymous mutations found in varying numbers of populations. Consistent with the Standard Model, non-synonymous mutations are most abundant among SNPs that are endemic and found in a limited number of populations. However, contrary to predictions from the Standard Model, the abundance of nonsynonymous mutations displays a noticeable uptick among the most cosmopolitan mutations. Using forward simulations, we show that the abundance of non-synonymous cosmopolitan SNPs can only be explained by models where balanced polymorphisms are common, rather than models relying solely on unconditionally beneficial mutations sweeping across metapopulations. Focusing on *D. melanogaster*, we find that the relative abundance of endemic deleterious mutations among populations is negatively correlated with mean annual temperature, consistent with theoretical predictions that effective population size determines the strength of purifying selection. In flies, cosmopolitan SNPs explain more variance in complex traits than endemic ones across hundreds of phenotypes, and nonsynonymous cosmopolitan variants are less likely to be classified as harmful based on integrative functional scores. Cosmopolitan non-synonymous SNPs are more likely to be trans-species polymorphisms, and are enriched for immunity-related genes. Together, these findings show that that the opposing action of purifying and balancing selection jointly contributes to the distribution of genetic variation across the species range.

11 Life history adaptation to an extreme habitat via a novel nonstructural supergene

Joanna Rifkin[1], Kory Kolis[2], Colette Berg[2], Thom Nelon[2], Findley Finseth[2,3], John Lovell[1,3], Lila Fishman[2] [1]Hudson Alpha, [2]University of Montana, [3]Scripps College

Environmental mosaics can generate strong divergent selection on complex multi-trait syndromes, not just single traits. When multi-trait adaptation occurs in the face of gene flow, distinct strategies are often packaged into alternative haplotypes at chromosomal “supergenes” with structural differences (often inversions) that suppresses recombination in heterozygotes. A classic case is the widespread life history polymorphism in yellow monkeyflowers (*Mimulus guttatus*), in which major chromosomal inversions define alternative annual (summer-dry habitats) vs. perennial (summer-wet habitats) life history strategies across W. North America. Here, we investigate the genomic basis of a parallel transition from perennial to annual across an extreme microgeographic habitat mosaic generated by geothermal activity in Yellowstone National Park. Four new chromosome-scale yellow monkeyflower genomes reveal that Yellowstone thermal annuals (as well as neighboring nonthermal perennials) are structurally “perennial”. QTL and population genomic analyses identify a novel major locus (*out6*, ~25 genes) underlying multi-trait adaptation to the unique temporal niche created by winter snowfall on thermally-heated soils. The two thermal annual vs. nonthermal perennial haplotypes at *out6* sort strongly by habitat and exhibit near zero recombination with each other, extremely low diversity, and deep sequence divergence, despite collinearity and high levels of cross-habitat gene flow and recombination genome-wide. In addition, a third rare and equally distinct thermal-associated *out6* haplotype appears to have recently introgressed into Yellowstone *M. guttatus* from *M. micranthus*, an annual selfer found only outside of the geothermal mosaic. Together, these findings suggest that divergent natural selection, potentially facilitated by universally low recombination in the sub-telomeric *out6* region, maintains multi-gene, multi-trait life history strategies adaptive on the extreme geothermal soil mosaic. Despite no re-use of the widespread chromosomal polymorphism, this system adds to the growing evidence for multi-trait local adaptation via recruitment of old genomic variation under long-term balancing selection species-wide.

12 The genetic basis of species persistence in two monkeyflowers with weak reproductive isolation

Henry Arenas-Castro, Cage Cochran, Quinn Evans, Hongfei Chen, Jenn M Coughlan Yale University

Sympatric species often exhibit permeable reproductive barriers, yet many of them coexist as phenotypically distinct entities. One way to account for this apparent paradox is that, during divergence with gene flow, selection on loci that underlie divergent traits render them less likely to cross species boundaries. Chromosomal rearrangements can also help to solve this paradox, as they capture sets of locally adaptive alleles, suppressing recombination among them. Here, we leverage genomic information of two young monkeyflower species, *Mimulus glaucescens* and *M. guttatus*, to ask how they keep their morphological and physiological integrity in sympatry despite lacking strong reproductive barriers. Even though both species exhibit little genetic differentiation, we confirmed that *M. glaucescens* is a distinct lineage of the *M. guttatus* complex. We then integrated independent population and quantitative genomics datasets and approaches to identify *i*) loci that have resisted introgression, ii) genomic regions associated with variation in divergent traits, and iii) signatures of selection along the genome, with keen interest on loci that fit all three criteria. Additionally, we detected low-recombination regions that bear the signature of chromosomal inversions. Combined, this information allowed us to elucidate important features of the genetic architecture that enable these species to persist in face of recurrent hybridization. Our results revealed a polygenic basis of species persistence, where multiple loci across most chromosomes have resisted introgression. Many of these putative barrier loci bear the signature of selection and lie in the vicinity of regions associated with variation in distinctive traits between both species. We also discovered several inversions that are segregating between species, many of which have captured an important number of barrier loci. Some of the genes that have resisted introgression have been implied in phenotypes that are highly divergent between *M. glaucescens* and *M. guttatus*, such as wax production. This set of genes was also enriched in subunits of a transcription factor that control the expression of numerous phenotypes, suggesting that selection acting on few highly pleiotropic loci can facilitate divergence with gene flow. Together, our findings suggest that both selection and recombination play an important role in the coexistence of hybridizing species and advance the molecular characterization of barrier loci in plants.

13 Introgression and parental conflict shape repeated occurrences of postzygotic isolation

Megan Frayer[1], Hagar Soliman[1,2], Pia Schwarz[1], Jenn Coughlan[1] [1]Yale University, [2]Cairo University

Postzygotic reproductive isolation is often thought to accumulate as a byproduct of neutral divergence. Yet, it frequently evolves rapidly, in line with non-neutral evolution. A major driver of intrinsic postzygotic reproductive barriers are intragenomic conflicts, such as conflict between maternal and paternal interests in resource allocation to offspring (i.e. parental conflict). Parental conflict may underlie hybrid seed inviability, a common and rapidly evolving reproductive barrier in angiosperms. Yet, in closely related, hybridizing species, it remains unclear how intragenomic conflicts and introgression interact to determine the fate of incompatibility alleles in nature. Here, we explore repeated incidences of hybrid seed inviability in a rising model: the *Mimulus guttatus* species complex. Using an extensive, range-wide crossing survey, we discover patterns of hybrid seed inviability within the widespread *M. guttatus* that are better described by geography than phylogeny. These patterns of reproductive isolation transgress species boundaries, as geographically-proximate but phylogenetically-distant species also exhibit similar patterns of hybrid seed inviability with allopatric populations of *M. guttatus*. We find strong support that patterns of reproductive isolation are consistent with parental conflict. Lastly, we provide evidence that introgression may underlie shared patterns of hybrid seed inviability between two species within this complex. Such introgression could have led to cascading reproductive isolation with other closely related species, creating a complex landscape of incompatibility. Overall, this work suggests that parental conflict and introgression can interact to shape the rapid and repeated evolution of strong reproductive isolation in the wild.

14 Functional connectivity of a threatened avian habitat specialist in a fragmented landscape

Shailee Shah[1], Faye Romero[2], Elizabeth Flesch[3], Tram Nguyen[4], Sahas Barve[5], Reed Bowman[5], John Fitzpatrick[1,4], Angela Tringali[5], Nancy Chen[6] [1]Cornell Lab of Ornithology, Cornell University, [2]Biology, University of Rochester, [3]Montana State University, [4]Ecology & Evolutionary Biology, Cornell University, [5]Avian Ecolgoy, Archbold Biological Station, [6]Ecology and Evolutionary Biology, University of California, Los Angeles

In light of widespread, ongoing habitat loss and fragmentation, maintaining functional connectivity—*i.e.*, effective dispersal—between discrete demes of vulnerable species is a conservation priority. Effective dispersal can be affected by landscape features, such as distance and size of suitable habitat patches and habitat composition and configuration of the intervening matrix, as well as deme attributes, such as population density and sex ratio, that may shape individual dispersal decisions. Our understanding of the drivers of functional connectivity is limited by the challenges of tracking dispersing individuals and their subsequent reproductive success across large geographical extents. Here, we investigated changes in functional connectivity over time and its effects on population trajectories in the Federally Threatened Florida Scrub-Jay (*Aphelocoma coerulescens*), a habitat specialist with short-distance dispersal living in a highly fragmented landscape. We combine whole genome resequencing data from 196 individuals across nine subpopulations and two time periods, high-resolution (10 m) land cover data, and census data to infer recent and historical patterns of gene flow, model the permeability of different land cover types, and examine the combined effects of functional connectivity and deme-level genomic and demographic attributes on changes in deme population size and density over time. Counterintuitively, citrus orchards appear to be greater barriers to Florida Scrub-Jay dispersal than urbanized areas. While careful habitat management practices increased population size in many demes, we found evidence of increasing isolation over time, emphasizing the importance of maintaining connectivity between populations. Our results shed light on the ecological and evolutionary consequences of anthropogenic landscape alterations, with important implications for conservation management priorities.

15 Dynamics and consequences of co-evolution between *Wolbachia* and *Drosophila recens*

Kelly Dyer, Cameron Seitz, Katherine Billmyre Genetics, University of Georgia

Wolbachia is a maternally inherited bacteria that can manipulate host reproduction to enhance its own transmission. *Wolbachia*-host interactions are highly dynamic and can co-evolve rapidly. The North American mushroom-feeding fly *Drosophila recens* harbors a *Wolbachia* infection that induces cytoplasmic incompatibility, where offspring of uninfected females and infected males die as embryos. We discovered that *D. recens* harbors two *Wolbachia* strains. First, we found that one strain is recently derived from the other, and they both occur in natural populations. Second, we assayed the phenotype of each *Wolbachia* strain in *D. recens* hosts to ask if the derived strain is replacing the original strain. Third, we introgressed each *Wolbachia* strain into the closely related host *D. subquinaria*. We find that while both strains cause cytoplasmic incompatibility in *D. recens*, the two *Wolbachia* strains cause different reproductive phenotypes in *D. subquinaria*, where one causes cytoplasmic incompatibility and the other induces male-killing, or the death of sons of infected females. Finally, we use the genomic and phenotypic differences between *Wolbachia* strains to investigate the genetic basis of male-killing. Our system provides a rare opportunity to observe host-endosymbiont co-evolution in real time and to investigate the genetic basis of *Wolbachia*-induced phenotypes.

16 A draft *Pan*-pangenome reveals the diversity and selection landscape of humans, chimpanzees and bonobos

Joana Rocha[1,2], Peter Sudmant[3] [1]Biology, New York University, [2]Integrative Biology, University of California Berkeley, [3]University of California Berkeley

Chimpanzees (Pan troglodytes) and bonobos (Pan paniscus) are sister species to humans, sharing more than 98% of our genetic makeup despite millions of years of divergence. The genetic background of these species provides a much-needed evolutionary context for capturing and representing the extensive structural diversity of human haplotypes. Chimpanzees comprise a geographically structured group of populations with divergence times between over 300 to 100,00 years ago, which are widely diverse and distinct. While critical to the study of ancestrally inherited genetic variation in humans, the origins and mechanisms driving and maintaining structurally complex loci in our endangered closest living relatives remains unknown. As a result, the diversity and context in which adaptive or disease-causing structural haplotypes have repeatedly emerged across human-ape evolution remain a mystery. Here we present a resource that consists of phased genome assemblies corresponding to 58 haplotypes from a comprehensive sampling of a diverse set of chimpanzees and bonobos throughout their geographic range in Africa. These include some of the most contiguous near telomere-to-telomere assemblies to-date for critically endangered and endangered Western, Central and Eastern chimpanzees, complementing and extending recent efforts to obtain a near-complete long-read based representation of the full spectrum of global genetic variation in humans and non-human primates. By combining our resource with existing human haplotype assemblies, we have characterized the structure, composition, function, and evolutionary trajectories of genomic variation that is both unique to and shared between chimpanzees, bonobos and humans. These integrated analyses – including graph-based representation of Pan genomic diversity - allowed us to identify SV still segregating in human and non-human primate populations, establish their population frequency, selective impact and level of recurrence, and represent extensive shared structural diversity in regions associated with local adaptation and resistance to disease, including malaria and environmental pathogens.

17 From Smeagol to Gollum: the trajectory, drivers, and consequences of Y chromosome degeneration

Kevin HC Wei[1,2] [1]Zoology, University of British Columbia, [2]Life Sciences Institute

Differentiated sex chromosomes like the X and Y typically originate from a pair of homologous autosomes that become sex-linked. The cessation of recombination of a new male-limited chromosome, the neo-Y, initiates a series of evolutionary changes that typically lead to gene loss and repeat accumulation. While population genetic theories can describe this process in broad strokes, empirical support and understanding of the underlying molecular mechanisms are often lacking. Across different Drosophila lineages, repeated fusions between autosomes and sex chromosomes have created neo-sex chromosomes; the neo-Ys are currently at different stages of the degeneration process and each of them offers a unique snapshot of the trajectory of Y chromosome decay. First, I will focus on a young neo-Y chromosome (~100 thousand years old) to explore the emergence and distribution of recessive deleterious and male beneficial mutations (including a male-essential locus) predicted to accumulate on the neo-Y. Second, I will showcase another neo-Y chromosome of intermediate age (~2.5 million years old), and discuss the genomic, epigenetic, and fitness consequences of transposable element activity and regulation, specifically when the neo-Y can act as a transposable element refugium. Together, these snapshots of differently aged Ys not only illuminate the drivers and mechanisms underlying Y degeneration, they also offer novel predictions regarding the trajectory of gene and repeat content evolution on the chromosome.

18 The spatiotemporal dynamics of plant adaptation to anthropogenic environments

Julia Kreiner Ecology & Evolution, University of Chicago

The nature and tempo of adaptation depends on the scales at which selection acts—across landscapes and through time. Yet this heterogeneity is rarely measured, limiting our ability to predict evolution and inform management. My lab leverages landscape genomics, herbarium time-series, and common garden experiments to study these adaptive dynamics in agricultural weeds, which evolve rapidly under intense human-mediated selection. Using century-spanning collections of common waterhemp (Amaranthus tuberculatus), we quantified exceptionally strong selection on resistance alleles across the landscape since the introduction of herbicides in the 1960s, reflecting the average response to selection over these scales. Extending this framework to study climate adaptation, combined with a drought experiment and ancestry mapping to resolve the genetic architecture of drought tolerance, temporal genomics reveal how polygenic responses have shifted over the past century. Critically, these allele frequency trajectories depended on local precipitation and temperature regimes, providing evidence that fluctuating selection through space and time determines the pace of adaptive responses to climate. To further resolve these dynamics, we are expanding the resolution of our herbarium sampling and leveraging farmer-driven in situ evolve-and-resequence experiments to track selection across ecological timescales. Because structural variants have proven to be particularly important in adaptive responses to extreme selection, we are also performing population-level long-read sequencing to better captures the complete spectrum of genomic variation, and have already resolved multiple, distinct structural forms of resistance alleles for integration with our temporal datasets. Together, this work aims to advance our understanding of how real world selection regimes shape the architecture and dynamics of rapid adaptation.

19 Navigating Growth–Survival Trade-offs in Fluctuating Environments

Megan G Behringer Biological Sciences, Vanderbilt University

Microbes are evolutionarily robust organisms capable of rapid adaptation to complex stress, enabling them to colonize harsh environments. In nature, microbes are regularly challenged by starvation, a particularly complex stress because resource limitation often co-occurs with changes in pH, osmolarity, and toxin accumulation created by metabolic waste. Often overlooked are the additional complications introduced by eventual resource replenishment, as successful microbes must withstand rapid environmental shifts before swiftly capitalizing on renewed resources to avoid invasion by competing species.

Using longitudinal population-genomic analysis of *Escherichia coli* populations evolved under extreme feast–famine conditions, we found that adaptation is characterized by narrow evolutionary trajectories with high mutational parallelism and reproducible mutational order. Genetic reconstructions reveal that early-arising mutations impose trade-offs for biofilm formation and motility but trade-ups for growth and survival, as these mutations confer positively correlated advantages during both short-term growth and long-term starvation.

To expand these investigations and directly quantify how fitness effects vary across distinct periods of the growth cycle, we developed a Bayesian multilevel framework for longitudinal randomly barcoded transposon sequencing (RB-TnSeq). Applying this framework to a 10-day feast–famine starvation regime in *Escherichia coli* yields a compact, genome-wide description of growth–survival trade-offs, linking short-term competitive fitness effects to long-term adaptive outcomes.

20 Role of genome structural variation in phenotypic diversity and adaptations

Mahul Chakraborty Department of Biology, Texas A&M University

A central goal of research in our lab is to uncover the genomic and molecular mechanisms underlying phenotypic variation and adaptation. Despite the rapid expansion of genomic datasets, trait-mapping studies often fail to pinpoint causal mutations—suggesting that many functionally important variants remain undetected by conventional genotyping approaches. To address this gap, we developed a new strategy that leverages highly contiguous genome assemblies to detect comprehensive population-level variation. This approach reveals that a substantial fraction of mutations—particularly those involving large (>100 bp) structural changes such as duplications, deletions, transpositions, and inversions—escape detection by standard short-read sequencing methods. Our resulting variant map uncovers numerous previously hidden structural variants (SVs) that offer new insights into the genetic basis of complex traits, visible phenotypes, and ecological adaptations in *Drosophila*. I will highlight several examples, including a tandem duplication coupled with a transposable element insertion that enhances the expression of detoxification genes linked to nicotine resistance. Another SV involving a TE insertion alters the expression pattern of a key insulin signaling gene, with downstream effects on life-history traits. These findings underscore the importance of SVs in shaping phenotypic diversity and demonstrate the power of assembly-based approaches in revealing the hidden layers of genome evolution.

21 Chromosome-scale drift under stabilizing selection

Carl Veller, Pavitra Muralidhar University of Chicago

For a complex trait under stabilizing selection, we study how the mean effects of different subregions of the genome, such as the chromosomes, change over time. We show mathematically that, although the genome-wide mean is constrained to remain near its optimum, the mean genetic value of any subregion of the genome is free to drift far from its initial value. This chromosome-scale drift under stabilizing selection is qualitatively identical to neutral genetic drift, but is slower by a factor that depends on the fraction of the trait's genetic variance contributed by the subregion in question. An immediate implication is that the chromosomes should differ substantially in their mean effects on any quantitative trait under stabilizing selection. Chromosome-scale drift is not impeded by occasional directional selection on the trait or by pleiotropic effects of genetic variants on other traits under stabilizing selection, but symmetric mutation between trait-increasing and trait-decreasing variants imposes a weak long-term brake. We explore two evolutionary implications of chromosome-scale drift under stabilizing selection. First, we show that it generates reproductive isolation between divergent populations, leading to purging of the minor-parent ancestry when the populations mix. Second, we show that chromosome-scale drift generates barriers to evolutionary transitions between sex-determining systems.

22 A General FST Framework Reveals the Variability of Rare Versus Common Alleles

Maike L Morrison[1], Noah A Rosenberg[2] [1]Santa Fe Institute, [2]Biology, Stanford University

Genetic variation across populations reflects evolutionary phenomena such as population structure, migration, and local adaptation. However, summarizing the shared variation of populations is challenging and question dependent. The statistic FST is a standard measure of such variation, but its reliance on heterozygosity (which squares every allele frequency) can obscure evolutionary phenomena whose signatures lie in patterns of rare genetic variation, such as recent population structure and gene flow. Inspired by the Hill numbers, a family of ecological diversity statistics, we present a generalized FST framework, FST-alpha, with tunable emphasis on rare versus common genetic variation. FST-alpha replaces heterozygosity with the Tsallis entropies, which use the parameter alpha to synthesize diversity measures such as richness (alpha=0; maximum emphasis on rare alleles), Shannon information (alpha approaches 1), heterozygosity (alpha=2), and fixation (alpha approaches infinity; maximum emphasis on common alleles). When alpha equals 2, FST-2 simplifies to standard, heterozygosity-based FST. We demonstrate that analyses using a range of alpha values can reveal important patterns of genetic variation that are overlooked when only standard FST (FST-2) is considered. In application to a global human genetic dataset, we find that the well-established negative relationship between genetic diversity and FST-2 reverses when more emphasis is placed on rare variation. We also identify scenarios where FST-2 values are nearly identical, but FST-alpha values for alpha not equal to 2 are very different. For example, when comparing a group of high-diversity populations and a group of low-diversity populations with similar FST-2 values, FST-0 and FST-1 captured substantial variation of rare alleles present only in the high-diversity population. Finally, simulations demonstrate that FST-alpha is better powered than FST-2 alone for detecting phenomena such as recent population structure. This framework expands the toolkit of FST-like statistics, enabling future studies to utilize the full range of genetic variation.

23 Leveraging Ancestral Recombination Graphs to Detect Adaptive Differences Among Gene Duplicates

Charlotte M LeMay[1], Liaoyi Xu[2], Arbel Harpak[2] [1]Computer Science, University of Texas at Austin, [2]University of Texas at Austin

Segmental gene duplicates offer a golden opportunity for evolutionary innovation, with famous examples including gene families coding for hemoglobin, opsins, and antifreeze proteins. Until recently, the reliance on short read sequencing limited our ability to resolve differences between highly-similar duplicate gene sequences, study their evolution and detect such adaptations. With the advent of long-read sequencing, there is new potential for this study. However, appropriate genealogical inference methods for the study of gene families using these data are underdeveloped. In particular, segmental gene duplicates experience rapid interlocus gene conversion, the copying of sequence tracts from one gene and "pasting" onto the paralogous gene, which violate assumptions made by state of the art genealogical inference methods such as Ancestral Recombination Graphs (ARG). The local absence of observed gene conversions can also indicate beneficial differences between duplicate genes: if differences are beneficial, selection should act against gene conversion events that homogenize these differences. Here, we develop an ARG-based method for studying duplicated genes. Our method incorporates within-species polymorphism data which provides a resolution improvement over phylogenetic methods. We apply our method to simulated data, and show that it may be used to infer interlocus gene conversion rate, as well as detect selectively maintained "islands" with little gene conversion. Finally, we applied our method to detect IGC islands in human long-read sequencing data from the 1000 Genomes Project to catalogue gene conversion islands in the human genome and hypothesize about the adaptation they may point to.

24 An Evolutionary Multi-Omic Framework to Assess the Evidence for Adaptive Regulatory Evolution Across Diverse Molecular Traits

Siyuan Feng, Jamie Freeman, John Pool Laboratory of Genetics, University of Wisconsin - Madison

Local adaptation may lead to elevated population differentiation not only in DNA sequences and organismal phenotypes, but also in intermediate molecular traits. However, functional multi-omic data remains under-utilized in evolutionary biology, and most studies do not analyze such data in an evolutionary framework. A few studies have compared quantitative trait differentiation (Q_{ST}) to levels of genetic differentiation (F_{ST}), to find outlier regulatory traits that may have been shaped by local adaptation. However, widespread usage has been hindered by demanding breeding designs, or under-powered due to estimator biases and extrinsic environmental and measurement variance confounding Q_{ST} estimation..

Here, we introduce REDQuanTA (Replication-Enhanced Detection of Quantitative Traits under Adaptation), a statistical framework and experimental design that leverages replicated trait measurements from the same genotypes and an Approximate Bayesian Computation (ABC) approach, to separate extrinsic variance from genetic variance. For any given organismal or molecular trait, REDQuanTA tests whether it is a candidate of local adaptation. This framework enables less biased Q_{ST} estimation without formidable experimental demands. It also ensures robust power and consistent false positive rates across traits with different levels of extrinsic variance, facilitating comparisons across diverse layers of molecular phenotypes.

We apply REDQuanTA to a large new multi-omic data set from the brains of Drosophila melanogaster populations adapted to ancestral warmer or derived cooler natural environments, motivated by population genomic evidence for an important role of nervous system genes in local adaptation. We analyze tens of thousands of molecular traits, including chromatin accessibility (ATAC-seq), coding and non-coding RNA abundance, alternative splicing, RNA editing, protein abundance, and post-translational modifications. We thus identify numerous -omic traits that may have contributed to environmental adaptation, and for the first time, we are able to ask which levels of gene regulation show the clearest evidence for adaptive evolution, and to test hypotheses like: whether selection leaves its strongest footprint at upstream regulatory steps or downstream protein phenotypes proximal to fitness. Overall, REDQuanTA offers an improved toolkit to detect adaptive trait evolution, and our results may suggest which types of -omic data offer the best investments for future studies of molecular-level adaptation.

25 Increased Genetic Diversity and Residual Stratification Contribute to Polygenic Score Prediction Accuracy

Nicole Kleman[1], Arslan A Zaidi[2] [1]University of Minnesota, [2]Genetics, Cell Biology, & Development, University of Minnesota

Recent efforts in human genetics have increasingly focused on analyzing data from diverse cohorts, given their potential for variant discovery and improved polygenic score (PGS) prediction. While this is a welcome shift, the complex genetic structure of diverse cohorts can bias genome-wide association studies (GWAS) effect sizes and PGS prediction accuracy (i.e. residual stratification). However, systematic analyses of residual stratification are lacking. To address this, researchers typically include ancestry components as covariates in PGS prediction; yet, the effect of this practice on PGS accuracy remains unclear.

To assess this, we derived the analytical expectation of PGS prediction accuracy of quantitative traits in the presence of stratification, with and without ancestry correction, in admixed cohorts. As expected, PGS r2 can be inflated in the presence of residual stratification; however, it can also be inflated or deflated, even with unbiased effect sizes, if the true genetic value is correlated with ancestry and, therefore, with the environment. Including ancestry as a covariate mitigates this bias, but it can also underestimate accuracy by removing genetic variation along the ancestry axes.

Empirically, we asked (i) does multi-ancestry GWAS improve PGS r2 in admixed cohorts? And (ii) is this increase driven by higher genetic diversity or residual stratification? To disentangle between the two, we computed PGS r2 in 48,586 self-identified African American individuals in All of Us (AoU) using summary statistics derived from multi-ancestry, European-ancestry, and sibling-based GWAS (which are robust to environmental stratification). All traits exhibited equal or higher r2 when using multi-ancestry GWAS, but some traits (e.g. height and systolic blood pressure) showed inflated accuracy due to residual stratification, as evident from decreased accuracy with sibling-based GWAS. Including ancestry components as covariates did not change the r2 for these traits, suggesting the bias due to stratification is orthogonal to these ancestry components. Sibling-based PGS r2 for other traits (e.g. BMI) remained comparable to the r2 based on multi-ancestry GWAS, suggesting that the improvement in PGS accuracy is not because of stratification but due to increased genetic diversity.

Even though we show increased genetic diversity in GWAS can improve PGS prediction accuracy, rigorous testing is needed to distinguish this improvement from residual stratification.

26 The Impact of Proxy and Missing Contexts on Polygenic Score Calibration

Shevaughn Holness[1], Sohini Ramachandran[1,2], Lorin Crawford[3,4] [1]The Center for Computational Molecular Biology, Brown University, [2]Department of Ecology, Evolution and Organismal Biology, Brown University, [3]Biostatistics, Brown University, [4]Microsoft Research

Recent studies have shown that an individual's environment, ranging from sociodemographic to biological, can have a significant impact on the prediction accuracy of their polygenic risk scores (PGS). Privacy concerns, fear of stigmatization, and survey design often result in noisy, missing self-reported information; however, the downstream consequences of using this incomplete, missing information to improve PGS accuracy is unknown. Analyzing the All of Us biobank, we find significant differences in the skipping frequencies per question within environments that vary by survey (1-28%) as well as that underserved communities are up to twice as likely to skip questions than both the global average and more represented communities. We also acknowledge that researchers use information from surveys as proxy variables for a true, latent variable of interest (i.e. income as a proxy variable for a social determinant of health (SDOH)). To investigate if using proxy variables could also have environment specific downstream consequences on PGS accuracy, we test the correlation between a proxy variable and its latent variable stratified by demography. We find that the correlation between income and various SDOH environment contexts vary significantly by demography, for instance the correlation between income and worries about access to food varies between -0.12 and -0.35 across gender. To investigate how the use of proxy variables affects PGS, we simulate scenarios where the proxy becomes decreasingly correlated with the latent environment context, the proxy has missingness that is dependent on the distribution of the latent environment variable, and the proxy itself is misspecified all at rates similar to what we found in surveys. Furthermore, we simulate environment specific (i.e non-random) missingness on its own. Our results show that poorly correlated contexts, proxy contexts with dependent missingness, and misspecified contexts significantly reduce the accuracy of PGS depending on the simulated gene by environment context scenario (~1-15%). In addition, the decrease in accuracy is often environment specific.

27 Microevolutionary cophylogeny using ancestral recombination graphs

Rowan Hart[1], Matthias Steinruecken[2] [1]Ecology and Evolution, University of Chicago, [2]University of Chicago

Cophylogeny provides insights into the specificity and shared evolutionary history of host-symbiont interactions. While cophylogenetic analyses have primarily investigated macroevolutionary patterns, increasing population genomic data of host-symbiont systems opens opportunities to explore microevolutionary cophylogeny. This remains challenging, however, since the evolutionary history between individuals within a host species cannot be well represented with a single phylogeny. Here, we measure microevolutionary cophylogeny by comparing the ancestral recombination graph of the hosts with the phylogeny of their paired symbionts. This approach provides cophylogenetic measurements of genome-wide trends, as well as signals specific to host genomic loci. Through simulations, we investigate the effects of transmission mode, population structure, admixture, and allelic incompatibility on cophylogeny. We find that genome-wide cophylogenetic signals arise through shared population structure, and demonstrate how genetic interactions can cause transient locus-specific signals. We measure mitonuclear cophylogeny within the 1000 Genomes Project and observe substantial variation in genome-wide signals across human populations, as well as locus-specific enrichment for host nuclear genes involved in mitochondrial pathways. Using paired host-microbe genomic datasets, including the Human Microbiome project, we also explore microevolutionary cophylogeny between humans and their microbiome, elucidating how host-genetics shapes microbial strain diversity.

28 Jointly inferring population sizes and the fitness effects of new beneficial mutations

Solomon Sloat[1], Parul Johri[1,2,3] [1]Department of Biology, University of North Carolina Chapel Hill, [2]Department of Genetics, University of North Carolina Chapel Hill, [3]Integrative Program for Biological and Genome Sciences, University of North Carolina Chapel Hill

The relative contribution of adaptive vs. non-adaptive evolutionary processes in shaping population-genomic variation has long been debated. Two challenges have made this problem difficult: (a) disentangling the joint contributions of mutation rate heterogeneity, recombination rate heterogeneity, population history, and background selection; and (b) characterizing the proportion and strength of beneficial mutations, which are transient and rare. The effects of beneficial mutations provide an additional inference challenge because site frequency spectra (SFS) can be explained equally well by either frequent small-effect selection or infrequent large-effect selection. To disentangle these parameters, we model non-adaptive and adaptive processes using forward-in-time simulation and perform joint inference of population history and the distribution of fitness effects of new beneficial mutations using an approximate Bayesian computation framework. By utilizing the variance in summary statistics across genomic windows, we take advantage of the linked effects of selection, producing reasonably accurate inference. For each of our parameters, we test the performance of SFS-derived statistics in addition to those based on linkage disequilibrium, divergence, and statistics that summarize the ancestral recombination graph. We show that inference of demographic history remains biased, even when accounting for recurrent positive selection, when the mean advantageous selection coefficient exceeds a certain threshold ($s > 0.002$), independent of the frequency of new beneficial mutations or the population size. We then apply our method to a Zambian population of *Drosophila melanogaster* and infer a large mean selection coefficient of beneficial mutations ($2Nes \sim 4500$) with little evidence of population growth, which contrasts with previous studies.

29 Quantifying selection on the nonsynonymous human mutation spectrum

David Castellano, Ryan N Gutenkunst Molecular and Cellular Biology, University of Arizona

Mutation rates and fitness effects are often treated as independent, but mutation rates are variable and evolve under indirect selection. For example, human European populations experienced a transient increase in the 3-mer mutation rate TCC -> TTC in the past 20,000 years. To quantify indirect selection on mutation spectra, we developed an approach to estimate the distribution of fitness effects (DFE) of nonsynonymous 3-mer mutation types, by analyzing pairs of complementary mutation types to account for GC-biased gene conversion and ancestral state misidentification. We then applied this approach to all 96 possible 3-mer mutation types in humans, using data from the 1000 Genomes Project. We found widely varying DFEs among mutation types and that inferences from a few hundred genomes could predict pN/pS in samples of tens of thousands of genomes. Our DFE estimate for TCC -> TTC mutations is consistent with recent theoretical predictions by Milligan, Amster, and Sella of scenarios under which a moderate number of modifier loci could yield population-specific transient bursts of a specific mutation type.

30 The dawn of phylogenetic comparative graph neural networks (PCGNNs): Applications to quantitative trait imputation and ancestral state reconstruction

Austin Patton, Ryan York Arcadia Science

Phylogenetic comparative methods (PCMs) are foundational tools in evolutionary biology and beyond, enabling researchers to account for evolutionary non-independence in comparative studies across species. Chief among these are methods to study trait evolution, infer ancestral character states, and impute missing species traits. However, conventional approaches rely on fitting explicit evolutionary models, often via maximum likelihood optimization, which scale poorly to many traits or large phylogenies and can suffer from model inadequacy when applied to real biological data. Here, we present a phylogenetic comparative graph neural network (PCGNN) architecture designed to jointly perform ancestral state reconstruction and trait imputation from partially missing species trait data and a phylogenetic tree. Our model is based on a GPS Transformer architecture with relative positional encodings that capture phylogenetic distances between nodes. We trained the model on 500,000 simulated phylogenies spanning five tree diversification models, with ten traits per tree simulated under five trait diversification models. We compared PCGNN's performance to mvMORPH, a widely used maximum likelihood implementation of multivariate trait evolution models. For trait imputation, performance was largely comparable between methods, with modest differences limited to specific tree and trait model combinations. In contrast, PCGNN consistently outperformed likelihood-based methods for ancestral state reconstruction, regardless of prediction depth. Importantly, inference with PCGNN was orders of magnitude faster than likelihood-based approaches and scaled to large phylogenies for which likelihood methods were computationally intractable. Together, these results highlight the potential of graph neural networks as fast, flexible, and scalable alternatives to classical PCMs in phylogenetic comparative biology.

31 Balanced polymorphism in a floral transcription factor underlies an ancient rhythm of daily reciprocal sex alternation in avocado

Jeffrey S. Groh[1,2], Marllon F. Soares dos Santos[3], Emmanuel Avila de Dios[3], Gracie Ackerman[4], Edwin Solares[5], Rodrigo A. Iturrieta[3], Eric Focht[3], Danelle Seymour[3], Brandon S. Gaut[6], Mary Lu Arpaia[3], Graham Coop[4] [1]University of California Berkeley, [2]UC Davis, [3]University of California Riverside, [4]University of California Davis, [5]University of California San Diego, [6]University of California Irvine

Many hermaphroditic plant species temporally separate male and female function within individuals, limiting reproductive trade-offs and inbreeding. In avocado, a pair of dominant and recessive alleles controls stable polymorphism for the direction of male-female temporal separation; one flowering type presents female-phase flowers in the morning and male-phase flowers in the afternoon, while another type exhibits the complementary pattern, thus promoting disassortative outcrossing. We map this dimorphism to haplotype variation at *SDMYB*, a member of floral transcription factors established as key regulators of floral maturation in diverse species with links to circadian hormone signaling. Rhythmic *SDMYB* expression over the day-night cycle is closely linked to a biphasic rhythm of floral opening, and the dominant allele, with nonsynonymous changes in conserved functional domains, exhibits a *cis*-regulated phase delay, underlying the offset sex timing between flowering types. The *SDMYB* alleles form an ancient trans-species polymorphism originating in the common ancestor of a clade of ca. 400 species, and have been maintained by negative frequency-dependent selection for over 42 million years. Drawing from additional research from my PhD thesis on the convergent evolution of heterodichogamy in other systems, I discuss how recent research on this mating system sheds new light on the genetic underpinnings of angiosperm sexual diversity.

32 Uncovering the Dynamics of Population Structure Through Time Using Genome-Wide Genealogies

Yun Deng[1], Jeffrey Spence[2], Jonathan Pritchard[1] [1]Stanford University, [2]University of California, San Francisco

Understanding population structure and the underlying demographic history is a central goal in evolutionary genetics. Popular ancestry decomposition methods, such as STRUCTURE and ADMIXTURE, provide a summary of individual ancestry proportions. But they are extremely compressed representations of the much richer evolutionary processes. In particular: (1) they do not resolve local ancestry patterns along the genome; (2) they provide no information about when ancestral groups diverged or how drift accumulated over time; and (3) when many components are specified, the interpretation of "latent ancestries" becomes unclear, when no present-day population is a pure representative of some components.

Here we introduce ARGmixture, a framework that transforms ancestry decomposition into a comprehensive reconstruction of demographic history. ARGmixture jointly calls local ancestry tracts and performs time-resolved PCA/MDS, enabling direct visualization of the temporal dynamics of population divergence and admixture. Crucially, ARGmixture can also infer the local ancestry and demographic history of "ghost ancestries"—ancestral groups lacking modern representatives—without requiring reference genomes. This is particularly important because many ancestries involved in human admixture events are extinct or only partially sampled in present-day populations.

We applied ARGmixture to the Middle Eastern populations of the Human Genome Diversity Project and reconstructed their demographic histories. Notably, ARGmixture identifies and characterizes the Levantine Farmer ancestry, which is represented in ancient DNA but absent as a pure modern population. ARGmixture not only recovers the corresponding local ancestry tracts in modern individuals, but also reconstructs its divergence and admixture history—without using any ancient DNA. This demonstrates ARGmixture's capability to enable a richer, temporally explicit understanding of population structure than traditional ancestry decomposition methods.

33 Small population size and isolation drive low genetic diversity, inbreeding depression, and an elevated mutation rate in the endangered Devils Hole pupfish (Cyprinodon diabolis)

David Tian[1], Nicolas Alexandre[1], Elise Kerdoncuff[1], Naomi Siu[1], Sean Lema[2], Bruce Turner[3], Olin Feuerbacher[4], Kevin P. Wilson[5], Michael R. Schwemm[4], Priya Moorjani[1], Jennifer Gumm[4], Christopher H. Martin[1] [1]University of California Berkeley, [2]California Polytechnic State University, [3]Virginia Tech, [4]US Fish and Wildlife Service, [5]National Park Service

Small populations with limited geographic ranges are predicted to be threatened by inbreeding and low genetic diversity, which can reduce fitness and accelerate population decline. One of the most extreme examples is the Devils Hole pupfish (*Cyprinodon diabolis*), a critically endangered species with the smallest known habitat range of any vertebrate, whose population has fallen from 200–500 historically to as low as 35 individuals in 2013. To assess the genomic and evolutionary consequences of small population size and isolation, we generated a chromosome-scale *de novo* reference genome of *C. diabolis* and 200 resequenced genomes across ten closely related desert pupfish species in Death Valley. We find that effective population size has declined across Death Valley pupfishes over the last ten thousand years, coinciding with the drying of pluvial lakes during the late Pleistocene. Genetic diversity is low across desert pupfishes, but up to an order of magnitude lower in wild *C. diabolis* (π = 0.000021), with the captive population having 40% less diversity than the wild, despite consistent introductions of wild individuals. Despite recent population decline, *C. diabolis* genetic diversity remains similar to pre-bottleneck levels assessed from genomes of 1940s formalin-fixed museum specimens. We also find strong evidence of inbreeding depression in *C. diabolis*: sequencing of embryos that died prior to 5 days post-fertilization, often with a heart defect, reveals that a 10% increase in inbreeding is associated with a 75% decrease in the odds of survival. While there is overall purging of strongly deleterious loss-of-function alleles relative to other desert pupfish species in *C. diabolis*, the remaining genetic load is dominated by fixed realized load. Finally, we leverage long-term low effective population size to test the drift barrier hypothesis and estimate a mutation rate of 8.09×10^{-9} per bp per generation, consistent with the hypothesis. Our results demonstrate that long-term small population size and isolation have driven low genetic diversity, strong inbreeding depression, realized load to fixation, and altered the mutation rate in *C. diabolis*, thus informing future conservation, how small populations evolve, and insights into the mechanisms driving mutation rate evolution in vertebrates.

34 Gene-specific selective sweeps are pervasive across human gut microbiomes

Richard T Wolff[1], Nandita R Garud[2] [1]University of California San Diego, [2]Ecology and Evolutionary Biology, University of California Los Angeles

The human gut microbiome is composed of a highly diverse consortia of species that are continually evolving within and across hosts. The ability to identify adaptations common to many human gut microbiomes would show not only shared selection pressures across hosts but also key drivers of functional differentiation of the microbiome that may affect community structure and host traits. However, the extent to which adaptations have spread across human gut microbiomes is relatively unknown. Here we develop a new selection scan statistic named the integrated linkage disequilibrium score (iLDS) that can detect sweeps of adaptive alleles spreading across host microbiomes by migration and horizontal gene transfer. Specifically, iLDS leverages signals of hitchhiking of deleterious variants with a beneficial variant. Application of the statistic to around 30 of the most prevalent commensal gut species from 24 human populations around the world showed more than 300 selective sweeps across species. We find an enrichment for selective sweeps at loci involved in carbohydrate metabolism, indicative of adaptation to host diet, and we find that the targets of selection differ significantly between industrialized populations and non-industrialized populations. One of these sweeps is at a locus known to be involved in the metabolism of maltodextrin—a synthetic starch that has recently become a widespread component of industrialized diets. In summary, our results indicate that recombination between strains fuels pervasive adaptive evolution among human gut commensal bacteria, and strongly implicate host diet and lifestyle as critical selection pressures.

35 Ultra-soft sweeps drive viral escape from broadly neutralizing antibodies and generate novel diversity signatures

Elena V Romero[1], Abigail E Clyde[2], Dylan Clark[1], Elena E Giorgi[2], Dylan H Westfall[2], Walker Azam[2], Megan L Taylor[1], Lillian B Cohn[2], Alison F Feder[1,2] [1]Genome Sciences, University of Washington, [2]Fred Hutchinson Cancer Center

Human immunodeficiency virus (HIV) rapidly generates variants which escape treatments and host immune responses. This rapid escape and subsequent treatment failure has been recently observed in clinical trials of broadly neutralizing antibody (bNAb) monotherapies. These clinical samples hold crucial information about how HIV rapidly adapts to selective pressures, which can be leveraged to improve design of bNAb combination therapies. However, access to this information has been historically limited by sequencing technologies with tradeoffs between sampling depth and preserving genetic linkage information.

Recent innovations have eliminated this tradeoff, enabling deep sampling while preserving full genetic linkage. Applying these innovations, we re-sequenced full-length viral envelopes from a longitudinal clinical trial of the bNAb 10-1074 to gain a 3.5X depth increase. Using novel linkage-aware analyses, we then performed the most in-depth genomic characterization to date of intra-host escape from bNAb monotherapy. We observed that 10-1074 escape occurred via an “ultra-parallel” process, with many different genetic backgrounds escaping concurrently. We also found that distinct genetic backgrounds within the same participant often escaped via the exact same amino acid change. This finding implicates recurrent mutation and/or recombination in driving escape, revealing that escape mutations arose on many more backgrounds than had been previously observed. Importantly, our finding indicates that bNAb combination therapies will likely need more components than previously hypothesized to successfully prevent escape.

We then investigated if such ultra-parallel adaptation is identifiable via conventional methods for detecting selection, or if it may be easily missed. To do so, we measured how 10-1074 escape impacted intra-host HIV diversity and linkage: two metrics commonly used for identifying selective events. Established expectations predict that selective events decrease diversity and increase linkage. However, the sheer number of genetic backgrounds escaping 10-1074 preserved pre-escape diversity and linkage levels across most trial participants, with diversity even increasing in the immediate vicinity of the escape locus. These unexpected patterns reveal a critical blind spot in current identification methods, underscoring the need to develop new approaches tailored to highly parallel escape.

36 Human Y-specific satellite repeat variation is associated with trans-acting modulation of gene expression

Shane F Warland, Andrew G Clark Molecular Biology and Genetics, Cornell University

Many eukaryotic genomes contain over 100 Mbp of satellite DNA repeats in pericentromeric and subtelomeric regions, and this abundance can vary between individuals by tens of megabase pairs. Multiple proteins have been identified that bind to and stabilize these heterochromatic regions. Furthermore, numerous pioneer transcription factors have been shown to function effectively within heterochromatin. In Drosophila and mice, variation in satellite DNA abundance produces a variable “heterochromatin sink”, which sequesters these proteins and alters their availability for the remainder of the genome. This sequestration subsequently modulates chromatin accessibility and global gene expression patterns. In this study, we investigate whether variation in human Y-linked satellites similarly drives differences in *trans*-acting gene expression. Using T2T genome assemblies, we identified 31 Y-specific *k*-mers embedded within tandem arrays of varying copy numbers. By characterizing this variation and its population structure, we identified significant associations with the transcript abundance of numerous autosomal genes. Furthermore, these satellite repeats are enriched with motifs matching the sequences flanking the promoters of the differentially expressed genes. The discovery that Y-linked satellite repeats function as key *trans*-regulators of gene expression may explain previously observed selective sweeps and the high inter-population heterogeneity seen on the Y chromosome. Collectively, these results support a more dynamic interpretation of Y-chromosome repetitive content as a potent, genome-wide modulator of transcription.

37 Detecting parallel adaptive changes in the gut microbiome

Aina Martinez i Zurita[1], Sharon Greenblum[2], Nandita Garud[3] [1]Human Genetics, University of California, Los Angeles, [2]Genome Technologies Group, Joint Genome Institute, [3]Ecology and Evolutionary Biology, University of California, Los Angeles

The human gut microbiome contains a diverse ecosystem of microorganisms that impact many aspects of human health. Due to the large population size of bacteria in each human host and their short generation time, it is estimated that an average human microbiome experiences billions of de-novo mutations every day. With such a large mutational input, the same mutation may arise in multiple hosts, and if beneficial, could rise to high frequency in these hosts. Mutations changing in frequency in parallel across many hosts in response to similar selection pressures may be adaptive and thus have important functional relevance to the microbiome. Statistical methods to detect such parallelism have proved insightful in metazoan populations, yet complementary methods that account for the unique features of complex bacterial populations have not yet been developed or rigorously tested. Here we develop a statistical framework to detect cases of parallelism in temporally sampled metagenomic datasets across many hosts. We first demonstrate our method is capable of detecting adaptive loci in a previously identified antibiotic resistance gene, DNA gyrase subunit A (gyrA), using a temporally sampled metagenomic dataset of 60 patients undergoing a 5-day ciprofloxacin antibiotic treatment. We further uncover elevated rates of parallelism across multiple other genomic loci during the antibiotic exposure and post-exposure in several bacterial species. Our work begins to uncover the dynamics of adaptive variants that may have been previously missed due to subtle yet consistent allele fraction changes.

38 Recovering signatures of archaic introgression using ancestral recombination graphs

Yulin Zhang[1], Arjun Biddanda[2], Sarah Johnson[1], Colm O’Dushlaine[3], Priya Moorjani[1] [1]UC Berkeley, [2]Johns Hopkins University, [3]insitro, Inc

The sequencing of the Neanderthal and Denisovan genomes has reshaped our understanding of archaic gene flow into modern humans. However, the limited availability of archaic genomes from deeper timescales or other regions (especially outside Eurasia), together with a lack of methods that can reliably identify archaic ancestry without unadmixed outgroup populations, has left the evolutionary history and impact of past introgression events largely unknown. We introduce TRACE, a novel approach to identify archaic ancestry in humans, by leveraging features of ancestral recombination graphs (ARGs) constructed from contemporary genomes alone, without requiring an archaic reference or an unadmixed outgroup. By performing extensive simulations, we show that TRACE has high sensitivity and specificity, comparable to existing methods. Applying TRACE to 1000 Genomes Project data reproduces known signatures of Neanderthal and Denisovan introgression in non-Africans. Moreover, TRACE reveals novel signals of "ghost admixture"–archaic gene flow from an uncharacterized hominin lineage–in both African and non-African populations, pointing to an introgression event predating the out-of-Africa expansion. Ghost ancestry is significantly depleted in conserved and low-recombination regions, yet notably persists within many deserts of Neanderthal and Denisovan ancestry. In Oceanian genomes, TRACE identifies significant enrichment of deep lineages within Denisovan–but not Neanderthal–introgression tracts, supporting a model of super-archaic gene flow into Denisovans and modern humans. Our results demonstrate the power of ARG-based approaches to recover hidden episodes of gene flow in our past and offer a scalable path towards mapping archaic ancestry in modern humans, even in the absence of archaic genome sequences.

39 Pervasive cryptic selection in the human noncoding genome

Swetha Ramesh, Chenlu Di, Kirk Lohmueller University of California, Los Angeles

Comparative genomics approaches identify conserved sequences across distantly related species and have estimated that approximately 5% of the human genome experiences negative selection. It is often assumed that mutations within constrained sequences are deleterious across the phylogeny and mutations outside are neutrally evolving. However, sites that have biological function in certain lineages but not in others, i.e. functional turnover, may violate this assumption since these sites may be invisible to comparative genomics approaches. It is unclear whether the cryptic, or hidden, negative selection from functional turnover affects patterns of genetic variation.

Here, we developed a statistical test to detect cryptic selection by comparing site frequency spectra of variants with different functional annotations. Applying our approach to simulated data, we observed statistically significant differences between site frequency spectra of regions with and without cryptic selection. Detection power depended on the proportion of mutations under cryptic selection, the amount of sequence tested, and the sample size. We applied our method to polymorphism data from the 1000 Genomes Project, comparing variants in putatively functional noncoding regions to those in quiescent (putatively neutral) regions. We detected pervasive signals of cryptic selection in putatively functional regions, even after filtering out the top 70% of conserved sites. Using simulations with varying levels of cryptic selection as benchmarks, we estimated genome-wide constraint in the human genome. Strikingly, we discovered that estimates were highly sensitive to how quiescent regions were filtered, suggesting that even these regions harbor deleterious mutations. Our approximation suggests that mutations in at least 7% of the human genome are under negative selection, which is greater than the estimates from conservation-based methods.

Our findings demonstrate that comparative genomics approaches miss considerable deleterious variation, and that the true extent of constraint is obscured when baseline “neutral” data retains residual negative selection. Cryptic selection has measurable effects on the site frequency spectrum, which may create consequences for downstream population genetic studies. These results emphasize the need to account for functional turnover when identifying putatively neutral variants for evolutionary analyses.

40 The effects of super-exponential growth on the human site frequency spectrum

Mingrui Yu[1], Kole Butterer[1], Deepjyoti Ghosh[1], Margaret C Steinet[1], John Novembre[1,2], Yuval B Simons[1,2] [1]University of Chicago, [2]National Institute for Theory and Mathematics in Biology (NITMB)

Human global population size has increased by over 10,000 fold over the past 100,000 years – from hundreds of thousands to nearly 10 billion. How has this period of super-exponential population growth affected the human site frequency spectrum (SFS)?

We introduce a simple model of super-exponential growth, borrowed from economic history literature, into population genetics. This model has been used in economic history to study the positive feedback between economic development and population size and approximates well the trajectory of human global population size. Using a branching process approach, we show that this model predicts a scaling law in the site frequency spectrum of ultra-rare alleles. The scaling parameter ranges from -2 for exponential growth to -1 for constant population size where, counterintuitively, the more super-exponential the growth the closer the SFS is to the constant population size limit. This happens because under super-exponential growth most change in population size happens in recent generations. We show that such scaling indeed exists in a dataset of 500,000 human exomes and this allows us to estimate the SFS scaling parameter and therefore the parameters of our super-exponential growth model.

With this model in hand, we can estimate what would be the benefits and challenges of obtaining population scale genetic datasets, e.g. a billion genomes project. We show that when study sizes get to the hundreds of millions, every possible substitution at every given site will be present in the dataset, i.e. all sites would be tetra-allelic, requiring a shift in bioinformatic pipelines. Since only strongly deleterious substitutions would be missing from such a dataset, it would be allow for easy identification of every highly-deleterious single nucleotide mutation in the human genome. Furthermore, by looking at the distributions of loss-of-function variants, it should become feasible to accurately estimate the deleterious effect of the knockout of any human gene in heterozygous state and, by looking at gene-wide Hardy-Weinberg disequilibrium, identify all recessive lethal genes in the human genome. We can therefore think of such a billion genomes project as quivalent to reading out the results of a full human mutagenesis experiment.

41 From individual life outcomes to evolutionary genetic change: lessons from multidecadal field studies of social mammals

Jenny Tung Max Planck Institute for Evolutionary Anthropology

Long-term field studies provide one of our most direct sources of insight into the traits that shape variance in fitness in natural populations. They therefore offer a ringside seat to watching evolution in action. Until recently, however, it has been very difficult to link individual-level field data with changes at the genetic level. Here, I will discuss our attempts to overcome this challenge by complementing multidecadal studies of social mammals in the wild with population-scale resequencing information on known individuals. In the 55-year study of the Amboseli baboons of Kenya, these data have helped us interpret and investigate the basis of natural selection against gene flow in this naturally hybridizing population. In the 33-year study of the Kalahari meerkats of South Africa, they have allowed us to probe the genetic architecture of body mass, an important fitness-related trait, and investigate changes in the frequency of body mass-associated alleles across generations. An emerging theme of this work is the importance of genetic data for identifying subtle patterns of change that may go unnoticed in the field, where overt determinants of fitness are typically social or environmental. Reciprocally, we have found that predictions from genetic data alone can lead to erroneous conclusions about phenotypic change. Together, our emerging results thus underscore the value—and mostly untapped potential—for long-term field studies to reveal how evolutionary genetic processes operate in nature.

42 Cis-regulatory change underlies a balanced flower color polymorphism

Zachary J Radford[1], Carolyn A Wessinger[2] [1]University of South Carolina, [2]Department of Biological Sciences, University of South Carolina

The persistence of discrete within-population polymorphisms suggests evolutionary mechanisms may actively maintain genetic diversity. We investigated an apparently balanced flower color polymorphism (FCP) found in the wildflower *Penstemon whippleanus* by linking molecular genetic variation to ecological processes. Populations can be polymorphic in flower color (purple, white) or fixed for either color. Importantly, polymorphic populations can be found throughout the geographic range of high-elevation areas within the Rocky Mountains. Our work examines whether balancing selection maintains variation within this system.

Using a genome-wide association scan (GWAS) approach, we localized the genetic basis of the flower color polymorphism to a genomic region containing two annotated genes: an R2R3 MYB (MYB1) that regulates anthocyanin biosynthesis and a C2-ABA-RELATED (CAR) gene involved in abscisic acid signaling. Both genes are significantly downregulated in developing floral buds of white flowers compared to purple flowers and show allele-specific expression consistent with a cis-regulatory mutational basis. We identified complex structural genetic variation and repetitive element proliferation (TEs) distinguishing the two haplotypes in the interval between these two genes, suggesting that the shared cis-regulatory effects could relate to local TE silencing of the white haplotype. Indeed, a gene involved in RNA-directed DNA methylation (Factor of DNA methylation 4) is significantly upregulated in white flowers. Differential expression analysis of light-stressed, pigmented leaves confirms that expression differences at the MYB1 and CAR is specific to floral tissues.

Genealogical reconstruction of MYB1 across 14 geographically widespread populations revealed deep divergence between the purple and white alleles, supporting a single origin of the FCP. Sequence analyses of the MYB1-CAR region indicate balancing selection, as shown through ancestral recombination graphs, positive values of Tajima's D, and measurements of F_{st}, π, and D_{xy} greater than genome wide average values. Together, these results are consistent with long-term balancing selection acting on flower color variation. Although field observations of pollinator visitation find no obvious role of pollinator-mediated selection, we found a strong positive correlation between elevation and the frequency of purple morph, suggesting that abiotic selective agents may contribute to maintaining this polymorphism.

43 Abundant Recurrent Mitochondrial Mutations and Widespread Mitonuclear Epistasis in *Caenorhabditis elegans*

Tuc Hiep Minh Nguyen, Daniel A Klein, Olivia S Weklar, Erin R Wengrow, Matthew V Rockman New York University

Coordinated genetic and physical interactions between mitochondrial and nuclear gene products regulate ATP production in the mitochondria. Linking mitochondrial genotypes and mitonuclear genetic interactions to phenotypes remains a complex challenge. Here, we have developed *Caenorhabditis elegans* as a model for mitonuclear epistasis studies. In a sample of 540 genetically distinct wild isolates, 10% of sites in the mitochondrial genome vary, with hundreds of missense mutations segregating in the species. Recurrent mutations and triallelic sites are common. Phylogenetic analyses of mitogenome sequences identified 8 distinct lineages, each with diagnostic variants. Principal component analysis of the nuclear genomes showed considerable concordance between mitochondrial and nuclear genomes in *C. elegans* populations, suggesting that disrupting coevolved mitonuclear genetic combinations could reveal substantial epistasis. We used GPR-1 overexpression, which disrupts the first mitotic division, to efficiently exchange nuclear and mitochondrial genomes between all pairs of 18 naturally isolated *C. elegans* strains, generating the largest-to-date animal mitonuclear exchange panel, with 323 unique viable mitonuclear genotypes. We phenotyped development of a subset of strains, with 30 unique genotypes, under 6 different environmental conditions, including high temperature and exposure to heavy metals. Mitonuclear epistasis contributed significantly to phenotypic variance across all tested conditions. We also tested for mitonuclear coadaptation by comparing the stress resistance of matched and mismatched cybrids. Interestingly, some mismatched strains exhibited greater resistance, highlighting the complexity and context dependence of mitonuclear interactions.

44 Single-cell eQTL Mapping Reveals Environment-dependent Genetic Regulation

Akriti Agrawal[1], Ina Suresh[1], Anne Friedrich[2], Joseph Schacherer[2], David Gresham[1] [1]Biology, NYU, New York University, [2]University of strasbourg

Expression quantitative trait locus (eQTL) mapping identifies genetic variants that regulate gene expression and contribute to phenotypic diversity. eQTL studies in controlled crosses have revealed the genetic architecture of natural variation in gene expression within specific genetic backgrounds. More recently, population-based eQTL studies have identified genetic variants in diverse *S. cerevisiae* strains isolated from various ecological niches. However, how genetic variants influence gene expression across different environments within a natural population remains unexplored. Traditionally, bulk RNA-seq is used to identify eQTLs, but it averages expression across entire populations, losing cellular heterogeneity and cell state information. To address these limitations, we developed a multiplexed single-cell RNA sequencing (scRNA-seq) approach to map eQTLs across ~ 96 genetically diverse *S. cerevisiae* strains under multiple conditions including rich media, fluconazole treatment, and synthetic rich media.

We used scRNA-seq for simultaneous cell genotyping, gene expression measurements, and cell cycle phase classification. We mapped eQTLs to SNPs, indels, and structural variants (SVs) using linear mixed modeling to comprehensively assess genetic effects across variant types, as well as stratified by cell cycle phase. We identified both local and distal eQTLs, with more trans-regulatory eQTLs across conditions. Notably, SVs exhibited particularly strong trans-regulatory effects, suggesting that SVs potentially drive broad regulatory changes. Trans-eQTLs exhibited phase-specific patterns, with non-overlapping sets of trans-regulatory loci in G1, S, and G2/M phases. Comparing rich media to fluconazole-treated conditions revealed that many of the eQTLs are environment-specific in these two environments, likely reflecting the distinct physiological demands of nutrient-rich versus stress conditions. These findings demonstrate that natural genetic variation shapes gene expression in environment-dependent ways. Thus, providing insights relevant to understanding phenotypic diversity in natural populations.

45 Unravelling the genetic basis of domestication in a multigenerational experiment with barley

Ruth Sarahi Perez Alfaro[1], Eyal Fridman[2], Daniel Koenig[3] [1]Plant Sciences, University of California, Riverside, [2]Institute of Plant Sciences, ARO Volcano, [3]University of California, Riverside

The domestication of plants is one of the most important developments in the history of human civilization and a classic example of rapid phenotypic and genetic adaptation. The genetic basis of domestication traits has been intensively studied for decades in most major crop species. The best studied domestication traits are easily observable characteristics such as grain shattering, plant architecture, and grain size. It remains unclear whether these visible traits are the only or even the most important changes that occurred during domestication. To study the dynamics of adaptation to domestication and to identify its genetic basis, we conducted a multiyear composite experiment using thousands of individuals segregating for wild and domesticated alleles, the barley CMPP composite (CMPPC). CMPPC was founded with equal contributions from 950 recombinant inbred lines resulting from intercrosses between wild and domesticated barley. Each recombinant carries between 10 and 20% of their genome introgressed from a wild relative. The CMPPC was planted in the fall in Riverside, CA in population sizes greater than 5000 individuals, allowed to compete over the winter, and then harvested in the spring en masse. The harvested seed was then planted in the following fall, allowing the population to evolve and adapt to the domesticated environment over time. Over the course of five years, we characterize extremely rapid evolution in response to selection against wild alleles genome-wide, with rare genomic regions showing evidence of selection favoring wild alleles. We pinpoint several loci that show exceptional evidence for selection, including both well characterized domestication loci and several that have not yet been described. Our experiments provide insight into the genetic basis of domestication and serve as a model for rapid adaptation to novel environments.

46 Shifting Genetic Architecture in Human Complex Phenotypes

Ana I Vazquez[1], Yifei Li[1], Harish Neelam[1], Hongsheng I Gui[2], Davorka Gulisija[3] [1]Michigan State University, [2]Henry Ford Health, [3]University of New Mexico

Medical genetic studies often focus on high-risk groups to enrich for targeted complex phenotypes. Yet most complex phenotypes are influenced by thousands of small-effect variants, with environmental and lifestyle factors further obscuring genetic signals. Together, these effects diminish statistical power and hinder replication. Untangling this complexity has become one of the central challenges of statistical genetics. We propose investigating the genetic architecture of complex, continuous phenotypes in low-risk groups, such as relatively younger individuals. We argue that although the incidence of many conditions often increases with age, this trend may be largely environmentally driven, thus inflating non-genetic variance and masking genetic effects. Studying younger, low-risk groups, therefore, may reduce residual variance and improve power to isolate genetic causes. To illustrate this, we present nine complex traits in both younger and older White European individuals from the UK Biobank. For most traits, environmental noise appears to increase with age: younger groups show a greater number of SNPs surpassing genome-wide significance, with notable overlap in associated variants. Correspondingly, heritability declines with age, reflecting a growing relative contribution of non-genetic variance for the same traits. Taken together, these patterns support the idea that genetic signals for complex traits may be more readily detected in younger participants, traditionally considered a low-risk group. This perspective highlights the importance of strategic participant selection, challenges the traditional focus on high-risk groups, and emphasizes the value of targeting higher-heritability groups to advance genetic discovery.

47 The genetic basis of neural circuit evolution for mate preferences

Emily Behrman Biological Sciences, Dartmouth College

Behaviors that arbitrate interspecies courtship are at the crux of evolutionary barriers between species.Species- and sex-specific pheromone profiles in *Drosophila* have rapidly diversified that signify species' identity to encourage courtship among conspecifics and discourage interspecies courtship. The dominant pheromone produced by *Drosophila melanogaster* females is an aphrodisiac to their males, but the same chemical cue inhibits courtship in male *D. simulans*, whose females do not produce that pheromone. We identify a gene that has evolved to reshape the neural circuit that processes this pheromone and change the male's response from attraction to repulsion. We assess how this gene changes the anatomical and physiological properties of specific neurons in the circuit between the species to produce divergent behaviors. Additionally, we examine intra-species variation in genetic and environmental factors that alter female production of and male attraction to the pheromone cue. Together, this provides a unique mechanistic understanding of the genetics underlying neural circuit evolution and how the corresponding behavior affects evolutionary processes in the wild across different timescales.

48 Evolution and genomic architecture of male reproductive strategies in swordtail fishes

Gabriel A Preising[1,2], Tristram O Dodge[1,2], Daniel L Powell[2,3], John J Baczenas[1,4], Theresa R Gunn[1,2], Alexandra E Donny[5], Rhea Sood[1], Paola Fascinetto-Zago[1,2], Ryan Cross[6], Samantha M Mason[6], Emmarie P Alexander[7], Andrew J Harris[7], Kang Du[8], Carla Gutiérrez-Rodríguez[9], Oscar Rios-Cardenas[9], Molly R Morris[6], Molly Schumer[1,2,4] [1]Stanford University, [2]Centro de Investigaciones Científicas de las Huastecas "Aguazarca", [3]Louisiana State University, [4]Howard Hughes Medical Institute, [5]University of Washington, [6]Ohio University, [7]Texas A&M University, [8]Texas State University, [9]Instituto de Ecología A.C.

The Y-chromosome has long been thought of as primarily playing a role in male sex determination. However, due to its non-recombining nature, the Y-chromosome can also potentially shield loci that benefit certain male phenotypes from recombination, preserving them over time. Here, we report a case of this occurring in a teleost system with alternative reproductive tactics. In swordtails (genus Xiphophorus), males exhibit a variety of reproductive morphs: some species exhibit large courting males, some exhibit small coercive males, and some exhibit a polymorphism for courting and coercive male morphs. We use X. multilineatus, which is polymorphic for both reproductive tactics, to map the genomic architecture underlying these different male morphs. By generating PacBio genome assemblies for each morph and performing a genome-wide association analysis, we identified a large, structurally variable region on the Y-chromosome that corresponds to reproductive tactic. Notably, this region exhibits a higher number of copies of the melanocortin-4 receptor (mc4r) gene in courting male Y-haplotypes. This gene regulates feeding behavior and growth in vertebrates, making it a strong candidate for regulating differences in body size between large and small male morphs. We also explore other duplicated genes and genomic features within the region, including transposable elements. Finally, we compare the structure of this region across different swordtail species and elucidate the evolutionary history of coercive male Y-haplotypes. These results highlight how structural variation of the Y-chromosome can be a source of variation that benefits different male reproductive strategies.

49 The genetic legacy of a single immigrant behind the Big Bird hybrid lineage in Darwin's finches

Rachel Goodridge[1], Matteo Sebastianelli[2], Leif Andersson[2], Peter R Grant[3], Rosemary Grant[3], Carlos Valle[4], Philipp W Messer[5], Erik D Enbody[5] [1]Computational Biology, Cornell University, [2]Uppsala University, [3]Princeton University, [4]Universidad San Francisco de Quito, [5]Cornell University

Hybridization can introduce novel genetic variation that alters evolutionary trajectories and may play an important role in rapid speciation during adaptive radiations. In the Darwin's finch radiation, hybridization between species is traceable through heritable variation in beak morphology, reflected in increasing morphological and genomic similarity between the medium ground finch, *Geospiza fortis*, and the common cactus finch, *G. scandens*, on the small island Daphne Major. However, the evolutionary outcomes of gene flow can be unpredictable, with rare events sometimes leaving lasting genomic signatures. In 1981, a single male *G. conirostris*, a cactus finch with a characteristically larger, deeper beak, made a rare long-distance migration to Daphne and founded a reproductively isolated hybrid lineage with *G. fortis*, known as the "Big Bird" lineage. The long-term fate of this lineage, including whether it persisted as a distinct lineage, went extinct, or introgressed into the population, has remained unknown. We used whole-genome data from 2022 to assess the genetic legacy of this lineage in the contemporary population, in the context of over 30 years of genomic monitoring on the island. Surprisingly, ADMIXTURE analyses revealed increasing *conirostris* ancestry in both *fortis* and *scandens* populations, such that by 2022, every sampled *scandens* individual carried at least 1% *conirostris* ancestry. Using identity-by-descent (IBD) tracts with the founder, we show that while the Big Bird lineage remained reproductively isolated from its parental species, it has more recently outbred with the *scandens* population. Remarkably, over 70% of *scandens* individuals now share at least one IBD segment with the founder, enabling reconstruction of approximately 80% of the founder's genome from the contemporary population. Consistent with a phenotypic effect of introgression, individuals with a higher fraction of their genome in IBD with the founder tend to exhibit blunter beaks. Together, these results reveal the extraordinary genetic legacy of a single immigrant and coincide with the emergence of highly variable beak morphology in *scandens*.

50 Beyond Fixation: Persistent Genetic Variation Under Intense Selection

Kenneth Arnold[1], Zachary S Greenspan[2], David L Hubert[1], Ryan D Robinson[1], Laurence D Mueller[2], Michael R Rose[2], Mark A Phillips[1] [1]Integrative Biology, Oregon State University, [2]Ecology & Evolution, University of California, Irvine

A central puzzle in evolutionary genetics is why substantial genetic variation persists within sexually reproducing populations despite long-term, intense directional selection, under which classical models predict rapid fixation and loss of diversity. Evolutionary experiments show that adaptation typically proceeds through highly polygenic responses, with selection acting on standing variation via coordinated, genome-wide allele-frequency shifts rather than classic hard sweeps. This raises the question of how such extensive variation is maintained. In natural and mesocosm experiments, fluctuating selection is often invoked as one such mechanism. By contrast, experimental evolution imposes sustained, well-defined selection that precludes fluctuating environments, requiring alternative explanations for the persistence of variation. While recent work has emphasized genetic redundancy, balancing selection remains a relevant and underexplored mechanism.

Here, we use long-term experimental evolution in D. melanogaster to test whether balancing selection via antagonistic pleiotropy can maintain cryptic genetic variation under strong life-history selection and enable rapid evolutionary responses when selection is reversed. Populations adapted for accelerated (A) or delayed (C) reproduction for hundreds of generations were shifted to the opposing regime, imposing explicit age-structured fitness trade-offs within a constant environment. Prolonged A-type selection is known to cause severe genome-wide reductions in heterozygosity due to sustained selection on early-life fitness.

We observe rapid phenotypic reversibility in pupation and age-specific mortality, with derived populations converging toward founder phenotypes of their target regime within relatively few generations. Genome-wide pooled sequencing reveals highly repeatable, anti-parallel genomic trajectories along shared axes of variation. Linear mixed-effects models identify ~175,000 SNPs with significant anti-parallel responses, indicating a coordinated polygenic shift. Notably, populations derived from A-type populations show a pronounced rebound in heterozygosity following selection reversal, approaching levels observed in founder C-type populations. Deep sequencing of founding A-types further uncovers ultra-rare alleles persisting below standard detection thresholds, revealing a hidden reservoir of functional variation retained despite prolonged selection.

These results demonstrate that even under intense directional selection, substantial genetic variation can persist and be rapidly redeployed when selection changes. We find that the combination of explicit age-structured trade-offs, mirrored genomic trajectories, and rapid recovery of heterozygosity is most consistent with balancing selection acting through antagonistic pleiotropy.

51 Measuring history dependence along evolutionary trajectories

Caroline Holmes[1], Misha Gupta[1], Alexandra Poret[2], Michael Desai[2] [1]Organismic and Evolutionary Biology, Harvard University, [2]Harvard University

Evolutionary trajectories follow complex paths through genotype–phenotype landscapes. When initiated from identical starting conditions, replicate populations fix different mutations, but it remains unclear whether such trajectories are functionally interchangeable. In particular, we do not know whether points at similar fitness along different evolutionary paths correspond to equivalent underlying phenotypes and positions in genotype–phenotype space, or instead reflect distinct evolutionary solutions. Here, we address this question using high-throughput CRISPR-based methods to introduce large numbers of targeted genomic perturbations. We evolved many parallel Saccharomyces cerevisiae populations under identical conditions and systematically reintroduced mutations that naturally arose during these trajectories both into their original line and into other replicate lines. This design allows us to measure the effects of real, historically accumulated mutations in diverse genetic backgrounds and at multiple evolutionary stages. By quantifying how the consequences of the same mutation depend on where and when it occurs, we reveal the structure of functional divergence among trajectories and track how local genotype–phenotype landscapes change over the course of evolution.

52 Y chromosome evolution shapes male reproduction in the world's highest-dwelling mammal

Daniel E. Shaw[1], Schuyler Liphardt[1], Naim M. Bautista[2], Timothy B. Wheeler[1], Guillermo D'Elía[3], Zachary A. Cheviron[1], Jay F. Storz[2], Jeffrey M. Good[1] [1]University of Montana, [2]University of Nebraska, [3]Universidad Austral de Chile

Species occupying extreme environmental gradients provide powerful models for linking genotypes to locally adaptive phenotypes. Extreme elevation is a particularly strong environmental barrier for organisms, requiring complex physiological adaptations to cope with hypoxia and cold. Exposure to extreme hypoxia also impairs reproductive physiology, including disruption of sperm development and quality in mammals. However, few studies have tested for phenotypic or molecular signatures of sperm adaptation in high elevation species. Here, we combine extensive comparative and population genomic data with common garden physiological experiments to test for genetic and phenotypic signatures of adaptation in male reproduction at extreme elevation. The Andean leaf-eared mouse (*Phyllotis vaccarum*) is the highest living animal in the world (exceeding 6700m) and spans the broadest range as it can be found as low as sea-level. We found that mice exposed to extreme hypoxic conditions in laboratory experiments exhibit decreases in testis size and sperm counts. We then used a chromosome-scale genome assembly of *P. vaccarum* and extensive population genomic resequencing spanning >200 low and high elevation individuals, as well as 8 other Andean species from the genus. We identified hundreds of genes that show high genetic differentiation between high and low elevation, as well as several multi-copy gene families that showed copy number differences associated with elevation. Interestingly, the strongest copy number correlation with elevation involved a cluster of three co-amplified gene genes that appear to have been recently acquired on the *Phyllotis* Y chromosome. Two of these genes have autosomal paralogs associated with sperm head development, while the third was associated with cellular responses to hypoxia. We predict that these Y-linked genes have co-amplified to increase production or performance of sperm at high elevation. Ongoing work is focused on comparing copy number and testis gene expression variation of these genes with reproductive phenotypes from laboratory individuals exposed to hypoxic conditions to understand if Y-linked copy number evolution contributes to adaptive spermatogenic responses to extreme elevation. Together, this work reveals that rapid Y chromosome gene family evolution may underlie adaptive benefits to male reproduction that are critical for population persistence at extreme elevations.

53 PRDM9-Directed Meiotic Breaks are Hotspots for Structural Variants Formation Across Mouse Strains

Hilda Opoku Frempong[1], Beth Dumont[2] [1]The Jackson Laboratory and The University of Maine, [2]The Jackson laboratory

Meiotic recombination is initiated by the programed induction of hundreds of double-stranded breaks (DSBs), which must be properly repaired to ensure stable transmission of an intact genome. In mice and humans, the location of meiotic DSBs is directed by the site-specific DNA binding of PR domain containing protein 9 (PRDM9), a histone methyltransferase that modifies local chromatin and recruits the DSB repair protein machinery. Although PRDM9 is essential for meiotic progression and genome stability, PRDM9-associated DSBs that arise in repetitive genomic regions may also promote structural mutations through homology-driven non-allelic repair. Thus, PRDM9 may serve both as a guardian of genome integrity and as catalyst for large-scale structural variation (SV) and karyotypic innovation in mammalian genomes. To investigate PRDM9's role in SV formation, I quantified SV enrichment at PRDM9-directed DSB sites across a panel of genetically diverse mice (the Nachman strains), each harboring a distinct PRDM9 allele. To map meiotic DSB sites, I adapted the CUT&RUN chromatin profiling method to target DMC1, a single-stranded DNA binding protein that localizes to PRDM9-directed DSBs. This approach enables high-resolution mapping of DSBs on single-stranded DNA, with strong concordance to previously published ChIP-seq DMC1 datasets and high binding specificity. DMC1 profiling across four inbred Nachman strains revealed highly reproducible DSB landscapes within strains, but significant heterogeneity across strains with divergent Prdm9 alleles. Integrating these DSB maps with strain-specific SVs, I uncover a significant enrichment of duplications at PRDM9-directed DSB sites in all surveyed strains. These results demonstrate that PRDM9 not only orchestrates essential meiotic processes but also plays a previously underappreciated role in driving structural genome evolution in mammals. Ongoing work aims to further dissect PRDM9's role in shaping genome architecture and its evolutionary implications

54 The evolution of structural and single nucleotide mutation across haplotype-resolved vertebrate genome assemblies

R. Nicolas Lou[1], Daven Lim[1], Minoli Daigavane[1], Landen Gozashti[1], Nilah M Ioannidis[2], Peter H Sudmant[1] [1]Department of Integrative Biology, University of California, Berkeley, [2]Applied Mathematics, University of California, Santa Cruz

Structural variants (SVs) contribute substantially to genetic variation and play vital roles in adaptation and disease. However, SVs are poorly captured by short read sequencing and thus are understudied, particularly in non-model organisms. Here, taking advantage of recently generated haplotype-resolved genome assemblies from >600 vertebrate species, we present the most comprehensive survey of the diversity of SVs and single nucleotide variants (SNVs) across the vertebrate tree of life to date. By identifying SVs and SNVs that segregate across two representative haplotypes in each genome assembly, we confirm patterns of reduced diversity of both SNVs and SNVs in endangered or threatened taxa. However, we find that while SNV and SV diversity are correlated across species, the proportion of these two forms of genetic diversity is fundamentally distinct across taxa, with fish and amphibians harboring 3 to 6-fold more segregating SVs than amniotes given the same number of segregating SNVs. We show that recent transposable elements (TE) activity is a significant source of SVs across vertebrates, with particularly rapid turnover observed in several mammalian lineages and higher diversity in TE composition in fish, amphibians, and reptiles. Using machine learning models we identify genomic features predictive of structural mutations across taxa with the top features broadly conserved across species, reflecting common bases underlying genomic instability in vertebrates. Lastly, we demonstrate that SVs are more likely to alter protein coding sequences than SNVs. Most of these variants are likely deleterious, and species harboring less genetic diversity tend to have a higher proportion of putatively deleterious variants. However, several genes, many of which are involved in olfactory and immune systems, are repeatedly impacted by SVs in multiple species, hinting at the adaptive roles SVs can play in evolution. Together, this study characterizes the diversity of SNVs and SVs across the vertebrate tree of life and highlights that patterns of segregating genetic variation are distinct across taxa with broad implications for vertebrate genome evolution, selection, and biodiversity conservation.

55 Endogenous Retrovirus Editing Drives Structural Remodeling of Human Acrocentric Chromosomes

Rafael Contreras, Mohammad Waseem, Amit Dutta, Azait Imtiaz, Hudson O'Neill Genetics, University of Alabama at Birmingham

Acrocentric chromosomes are among the most structurally dynamic regions of the human genome, yet the mechanisms underlying their extensive remodeling remain poorly understood. Here, we show that endogenous retrovirus editing centered on the HERV-K–derived element K111 represents a major force shaping acrocentric chromosome architecture in humans. Using telomere-to-telomere assemblies, 53 human pangenome genomes, and comparative primate analyses, we reconstruct the evolutionary history and structural diversification of K111 and related K111-derived variants across acrocentric chromosomes. Phylogenetic analysis of K111 internal genes and long terminal repeats indicates that K111 arose from a single ancestral integration event approximately six million years ago, near the Homo–Pan divergence. Consistent with this timing, K111 is present in humans, chimpanzees, and bonobos, but absent from gorillas and orangutans. In chimpanzees and bonobos, K111 is restricted to three to four full-length proviruses with limited structural diversification and little evidence of solo LTR formation. In contrast, human genomes show extensive diversification of K111 across all five acrocentric chromosomes, including full-length proviruses, solo LTRs, truncated elements, and recombinatorial variants representing distinct K111 species. Solo LTR formation, a host-mediated viral editing mechanism that removes proviral coding regions, is markedly enriched in humans and in some individuals is associated with complete loss of full-length K111, while phylogenetic diversification among solo LTRs is consistent with continued recombination between acrocentric chromosomes. In six of fifty-three individuals, more than one K111 species is present on the same chromatid, a configuration consistent with segmental duplication and recombination between acrocentric arms rather than independent retroviral insertions. Structural variant breakpoints cluster at junctions between K111-derived elements and flanking centromere-associated repeat (CER) sequences, implicating these boundaries as recombination hotspots. Immunofluorescence-FISH demonstrates that K111 localizes to acrocentric short arms at the nucleolar periphery. Together, these data support a model in which human-specific endogenous retrovirus editing repurposes K111 from an ancestral provirus into LTR-rich structural elements that continue to participate in recombination within and between acrocentric chromosomes, driving copy-number diversity and structural heterogeneity in modern human genomes.

56 Strong Negative Selection on Structural Variants of Common Fruit Flies

Jen-Yu Wang, J.J. Emerson Ecology and Evolutionary Biology, University of California Irvine

Long-read sequencing technologies have significantly facilitated genome assembly and the discovery of structural variants (SVs). In this study, we assembled over sixty genomes of *Drosophila melanogaster* using long-read data from the NCBI SRA. We identified SVs, including deletions, insertions, duplications, inversions, and translocations, through an assembly-based pipeline validated by simulation. While mapping-based programs are computationally efficient, they exhibited lower sensitivity compared to assembly-based methods, which effectively captured larger-scale genomic changes.

Our findings indicated that insertions were generally larger than duplications, followed by deletions. Single nucleotide polymorphisms (SNPs) were found to be thousands of times more abundant than SVs. We explored the association between SVs and various types of repeat units. Insertions were primarily linked to retrotransposons, whereas deletions and duplications were more frequently associated with simple repeats and satellites. A detail often overlooked when only inspecting repeat elements in an assembly. Notably, around 30% of SVs contained more than one type of transposable element

Unlike SNPs, SVs, which often span and affect multiple genes, are more likely to influence an organism's phenotype. This suggests stronger selective pressure on SVs, although the extent was previously unknown. In this study, we generated unfolded site frequency spectra for each category of SVs and transposable element superfamilies. Our results show that SVs are subject to much stronger negative selection compared to non-synonymous SNPs, by several orders of magnitude. Deletions were the most negatively selected, insertions were less selected and duplications were most relaxed from selection. We observed that selection signals varied among different types of transposable elements. Additionally, we investigated genes with specific functions or high allele frequencies to identify SVs that may contribute to population differentiation.

57 Transposable elements shape local mutation landscapes through chromatin remodeling

Yuheng Huang[1], Xu Mark[2], Grace Yuh Chwen Lee[2] [1]University of California Irvine, [2]UC Irvine

Mutation is the ultimate source of genetic variation, yet understanding what drives variation in mutation rates across genomes remains a central question in evolutionary genetics. Both sequence context and chromatin states influence mutation rates and spectra. However, whether and how transposable elements (TEs), selfish genetic elements present in nearly all eukaryotes, shape mutation patterns remains unaddressed. This question is critical because TEs not only change the local sequence context but also alter chromatin states. The latter occurs because host-directed silencing mechanisms enrich repressive epigenetic marks at and around TEs, even within gene-rich euchromatic regions. To investigate how TEs influence mutation rates in flanking unique sequences, we conducted mutation accumulation (MA) experiments using Drosophila strains with varying expression levels of Su(var)3-9, a dosage-dependent modifier of repressive mark spreading. This design enabled comparison of the same TE insertions exhibiting different levels of enrichment of repressive marks across strains, isolating the impact of TE-mediated chromatin changes on mutation rates. Using whole-genome sequencing and newly developed analysis pipelines, we identified de novo point mutations and TE insertions across MA strains. As expected, our analysis revealed that increased enrichment of repressive marks reduced de novo TE insertions. Interestingly, this enrichment also lowered point mutation rates around TEs, with particularly strong effects for TEs transposing through DNA intermediates. To validate these findings through an orthogonal approach, we analyzed wildtype inbred strains established from natural populations and sequenced a decade apart, capturing mutations accumulated over ~200 generations of lab maintenance. Comparing homologous alleles with and without TEs, we again observed the protective effect of TE-mediated enrichment of repressive marks against point mutations. Given the extensive variation in TE insertion sites and composition both within and between species, our work reveals a previously unrecognized mechanism by which TEs actively shape genome-wide mutation landscapes by remodeling the chromatin environment.

58 Selection on telomere homeostasis drives the adaptive evolution of TERT in *Mimulus*

Naseem Samo[1], Linh Nguyen[2], Jae Young Choi[2] [1]EEB, University of kansas, [2]University of Kansas

Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, US

Every round of cellular replication shortens chromosome ends and without any intervention will result in catastrophic genomic instability. Hence, a specialized ribonucleotide protein complex named the telomerase maintains chromosome ends against attrition. Catalytic activity of the telomerase is determined by the *Telomerase Reverse Transcriptase* (*TERT*) protein that uses a noncoding Telomerase RNA (*TR*) as a template and synthesizes short tandem nucleotide repeats to form the telomere. Recently we discovered a monkeyflower species (*Mimulus lewisii*) with a duplicated *TR* gene and sequence divergence between the paralogs resulted in mutations in the templating domain. Due to this *M. lewisii* has a non-canonical sequence heterogeneous telomere that consists of a 6 bp and 7 bp tandem repeat sequence. Here, we investigated the molecular evolution of the *M. lewisii TERT*, which is the binding partner of *TR*. Initially, we conducted a phylogenomic analysis using Nanopore sequencing to assemble 5 new *Mimulus* species genome and conducted a comparative genetic analysis of the *Mimulus TERT*. A clade-specific PAML analysis detected significant evidence of positive selection for *TERT* but only in lineages sharing the *TR* duplication. We interrogated the functional basis of the positive selection, first using yeast three hybrid assay and testing the physical binding between *TERT* and *TR*. Results showed *M. lewisii TERT* binds the ancestral *TR* (*TR1*) and recently duplicated *TR* (*TR2*) paralog, but in the sister species without the duplicated *TR* its *TERT* did not bind *TR2*. Using inter-species domain swap experiments we narrowed the casual region of binding both *TR* duplicates to two amino acids. Three-dimensional modeling with Alphafold indicated the residues were located within a potential intrinsically disorganized domain of the *TERT* protein. We investigated the catalytic activity between the *M. lewisii TERT* and *TR* paralogs by combining next-generation sequencing with the Telomeric Repeat Amplification Protocol ($TRAP_{NGS}$). Results showed *M. lewisii TERT* synthesized more nucleotide products using *TR2* but each product had significantly lower number of telomere repeats, suggesting selection on the *M. lewisii TERT* may have involved *TR* paralog preference and telomerase processivity. Finally, we discovered *M. lewisii* genotypes with and without the *TR2* paralog have significantly different telomere length, indicating the selection on *TERT* is ultimately related to telomere homeostasis.

59 Tempo and mode of transposon mobilization in the *Drosophila* male germline

Peiwei Chen, Alexei A Aravin, Andrew G Clark, Cedric Feschotte Cornell University

Transposons are selfish genetic elements that generate copies of themselves at new genomic loci, repeatedly producing mutations that fuel evolution. While transposon mobilization may be unequivocal, it remains nontrivial to systematically characterize the tempo and mode of transposition. In addition, it is often unclear whether activation of transposon expression in the germline actually leads to mobilization. Here, we develop a pedigreed mutation accumulation approach in *Drosophila melanogaster*, using *rhino* mutations to disrupt piRNA production and derepress transposon expression. Leveraging artificial selection and the lack of recombination in male flies, we followed a full complement of nonrecombining haploid genome in the male germline through 45 generations, accumulating transposon insertions on whole chromosomes. We performed single-fly Nanopore sequencing and identified a dozen families that transposed after piRNA pathway perturbation, all of which showed elevated expression in *rhino* mutant testes. Among them, *copia* was the most prolific and completed 63 jumps, with new insertions scattered across each chromosome arms. Using genomic DNA extracted at each of the 45 generations, we verified and dated every single *copia* insertion, revealing that *copia* transposed in bursts, departing from a Poisson process. Meanwhile, another transposon family, *roo*, transposed with high frequency irrespective of piRNA pathway perturbation, likely by exploiting a unique temporal niche in early embryonic development before the piRNA pathway is established. Collectively, by experimentally minimizing selection, recombination, and piRNA-mediated control, this work defines the tempo and mode of transposon mobilization in the male germline with unprecedented resolution. Given the ubiquity of transposons across the tree of life, these insights into their mobilization have broad implications for understanding a fundamental mutational process that has had a profound impact in evolution.

60 Notes on a model for evolvability at the molecular scale: A reflection on Weinreich et al. 2006

Brandon Ogbunu[1], Brandon Ogbunu[2] [1]Ecology & Evolutionary Biology, Yale University, [2]Yale University

In 2006, Weinreich and colleagues published a landmark empirical fitness landscape of TEM β-lactamase, demonstrating that Darwinian adaptation can proceed along only a small subset of mutational trajectories connecting low- and high-resistance genotypes. Twenty years later, this system remains one of the most tractable experimental models for interrogating the mechanistic basis of evolvability. Here, I revisit the Weinreich et al. landscape as a canonical case study for connecting population-genetic theory with experimentally measurable genotype–fitness maps. I synthesize core findings from the original work—including the pruning of mutational pathways by sign epistasis under strong-selection, weak-mutation dynamics—with theoretical and empirical developments that followed. These include advances in quantifying higher-order epistasis, global epistasis, and landscape ruggedness; extensions to fluctuating environments and fitness “seascapes”; and new metrics for trajectory structure and the speed of molecular evolution. I argue that the enduring impact of the Weinreich landscape lies not only in its biomedical relevance, but in its role as a conceptual bridge linking adaptive dynamics, contingency, and navigability to experimentally resolvable molecular systems. More broadly, I highlight how small, combinatorially complete fitness landscapes continue to shape modern thinking about predictability, constraint, and the geometry of adaptive evolution.

61 Module-selection balance in the evolution of modular organisms

Mark Kim[1,2], Sarah M Ardell[1], Sergey Kryazhimskiy[1] [1]Department of Ecology, Behavior and Evolution, University of California San Diego, [2]Department of Computational Biology, Cornell University

The structure of genotype-phenotype-fitness maps (GPFMs) plays a central role in shaping evolutionary dynamics. One salient feature of biological GPFMs is that they are variationally modular, meaning that each mutation affects only a small subset of functional traits. By restricting trait combinations that are mutationally accessible to the organism, variational modularity imposes some constrains on evolutionary dynamics, but what these constraints are is not well understood.

Here, we investigate how variational modularity constrains trait evolution using a series of minimal GPFM models based on Fisher's geometric model with two quantitative traits. We find that the evolutionary dynamics in the trait space that emerge on GPFMs with universal pleiotropy—where most mutations affect both traits simultaneously—differ qualitatively from those emerging on variationally modular GPFMs—where each mutation affects a single trait. On a pleiotropic GPFM, natural selection guides a typical population along the fitness gradient whereby the trait under stronger selection approaches the optimum exponentially faster than the trait under weaker selection. In contrast, on variationally modular GPFMs, a typical population approaches a quasi-steady state where both traits improve at the same rate and their ratio remains constant. We refer to this quasi-steady state as a "module-selection balance". We argue that this balance is an inherent feature of adaptation on variationally modular GPFMs, although the specific steady-state trait ratio as well as the rate of convergence to it depend on the details of the GPFM and on the mode of adaptation, e.g., the supply of adaptive mutations and recombination rate.

One consequence of our theory is that adaptive substitutions accumulating in populations at the module-selection balance should be distributed broadly across modules. To test this prediction, we re-analyzed published genomic data from Lenski's LTEE (Good et al, *Nature* 2017). We find that adaptive mutations acquired by these populations early on—when they are likely to be far from the module-selection balance—are concentrated in a few genes. In contrast, adaptive mutations acquired at later stages of adaptation—when the populations are expected to be closer to the module-selection balance—are distributed broadly across the genome, consistent with our theory.

Overall, our theory reveals the module-selection balance as an important constraint on trait evolution imposed by variational modularity. This finding may have implications for our understanding of evolution at both genetic and phenotypic levels.

62 Massively parallel interrogation of the fitness of natural variants in ancient signaling pathways reveals pervasive local adaptation

Jose Aguilar-Rodriguez[1,2], Jean C. C. Vila[1], Shi-An A. Chen[1,3], Manuel Razo-Mejia[1,3], Olivia M. Ghosh[1,4], Daniel F. Jarosz[2,5], Hunter B. Fraser[1], Dmitri A. Petrov[1,6] [1]Department of Biology, Stanford University, [2]Department of Chemical and Systems Biology, Stanford University School of Medicine, [3]Present address: Altos Labs, Bay Area Institute of Science, [4]Department of Physics, Stanford University, [5]Department of Developmental Biology, Stanford University School of Medicine, [6]Chan Zuckerberg Biohub

The nature of standing genetic variation remains a central debate in population genetics, with differing perspectives on whether common variants are almost always neutral as suggested by neutral and nearly neutral theories or whether they can commonly have large functional and fitness effects as proposed by the balance theory. We address this question by mapping the fitness effects of over 9,000 natural variants in the Ras/PKA and TOR/Sch9 pathways—key regulators of cell proliferation in eukaryotes—across four conditions in *Saccharomyces cerevisiae*. While most variants are neutral in our assay, ~3,500 exhibited significant fitness effects. These non-neutral variants tend to be missense and to affect conserved, more densely packed, and less solvent-exposed protein regions. While some of these non-neutral variants are younger and rarer, and more often found in heterozygous states—consistent with purifying selection—a substantial fraction is present at high frequencies in the population, which is expected under balancing selection. Indeed, we find that variants with a positive fitness effect in our laboratory measurement show strong signs of local adaptation as they tend to be found specifically in domesticated strains isolated from human-made environments. Our findings support the view that while many common variants might be effectively neutral, a significant proportion have locally adaptive functional consequences and are driven into a subset of the population by local positive selection. This study highlights the potential to combine high-throughput precision genome editing with fitness measurements to explore natural genetic variation on a pathway-wide scale, thereby bridging the gap between population genetics and functional genomics to understand the nature of evolutionary forces in the wild.

63 The emergence of global epistasis from the geometry of the genotype-phenotype-fitness map.

Jhelam Nitin Deshpande, Sergey Kryazhimskiy Department of Ecology, Behavior and Evolution, University of California, San Diego

Global epistasis, or a linear, usually negative, dependence of the fitness effects of mutations on the fitness of the background strain into which the mutations are introduced has been found in a variety of organisms and environments. Previous statistical models by Lyons et al (Nat Ecol Evol 2020) and Reddy and Desai (eLife 2021) explain global epistasis as a "regression to the mean", which arises when the genetic architecture of the organism has many uncorrelated genetic interactions. However, the extent to which such regression to the mean explanation is consistent with biological mechanisms, for instance, gene regulation or metabolism remains unclear. Another class of models that could potentially explain global epistasis are those in which fitness is a non-linear transformation of an underlying phenotype, such as the Fisher's geometric model (FGM). Such models lend themselves to incorporate biological mechanisms more naturally, but the extent to which they can capture all empirically observed properties of global epistasis is unknown. To address these problems and potentially reconcile the statistical and mechanistic explanations for global epistasis,, we take a twofold approach. First, we examine geometric constraints in the FGM on directions of mutations such that they exhibit realistic global epistasis. We find that for most mutations to exhibit global epistasis with negative slopes, it is necessary to break the symmetry of the canonical FGM and ensure that the directions of mutations are distributed non-uniformly. We then ask whether and how such non-uniformity in the directions of mutations can emerge from a genotype-to-phenotype map. To this end, we examine a simple model of a linear metabolic pathway with a proteome allocation constraint. We find that, in this model, mutations that affect gene expression are strongly constrained in their directions and therefore can readily exhibit global epistasis., Overall, this work demonstrates a so far missing link between global epistasis, biological mechanisms and the geometry of the genotype-phenotype-fitness map.

64 Nucleotide and amino acid-level effects combine to shape the navigability of protein fitness landscapes

Bryan Gitschlag, Carlos Martí-Gómez, David M McCandlish Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory

Fitness landscapes are a fundamental concept in evolutionary genetics that help us understand how evolving populations navigate sequence space. Due to technological advancements, such as high-throughput combinatorial mutagenesis assays, it is now possible to measure the structure of empirical fitness landscapes consisting of hundreds of thousands of amino acid sequences. Analyzing multiple such protein fitness landscapes, we show that evolutionary barriers to the accessibility of high-fitness sequences often arise not only due to interactions between amino acids but also the interplay of these interactions with nucleotide-level biases in mutation and the structure of the genetic code. To gain an intuitive biological understanding of these large datasets and the complex evolutionary dynamics they induce, we combine recent advances in machine learning with computational techniques that provide low-dimensional visualizations of high-dimensional fitness landscapes. Applying this workflow to several functional elements, including enzyme active sites and protein-binding motifs, we show that high-fitness regions of sequence space are often defined by highly specific genetic interactions. Although such amino acid-level analysis helps elucidate the biophysical basis of the large-scale geometry of fitness landscapes, evolution ultimately depends on nucleotide-level events, which introduce biases due to heterogeneity in the rates of different kinds of mutations, as well as biases in the mutational connectivity between amino acids due to the structure of the genetic code. By projecting our inferred protein fitness landscapes into nucleotide sequence space, we show how selection on amino acid sequence and nucleotide-level effects jointly determine the course of evolution of protein functional motifs. Specifically, we find that while observed wildtype sequences are sometimes separated from fitter genotypes by amino acid-level fitness valleys, such separation can also arise from amino acid changes that are mutationally unfavorable at the nucleotide level, inaccessible due to the genetic code, or a combination of these factors. Overall, this work provides an integrated perspective on how the structure of the genotype-phenotype map combines with mutational biases and the structure of the genetic code to determine the qualitative features of molecular evolution on high-dimensional fitness landscapes, and helps to explain observed patterns of amino acid sequence variation.

65 ARGuing with data: tree sequences in real plant genomes

Jeffrey Ross-Ibarra Evolution and Ecology, University of California

Ancestral recombination graphs (ARGs) are powerful tools for reconstructing the evolutionary history of a sample. ARGs are growing increasingly popular in empirical evolutionary genetics, used to answer questions ranging from geographic spread to the timing of mutations, selection on quantitative traits, or the identification of admixture. Most applications of ARGs, however, have been limited to human genomics, and the few uses in other species have mostly taken ARG estimation from real data as a given. Here, I highlight some of the challenges in applying ARGs to real plant genomes, from repeats to indels and ancestral states, and the merits of short reads and whole-genome alignments. I then present several different use cases: testing the origin of a key domestication allele, identifying selection on transposable elements, and highlighting the demographic impact of European colonization on American crops.

66W Genetic and temperature variation in anteroposterior (AP) axial patterning of *Drosophila*

Lossie (Elle) Rooney[1], Gregory T Reeves[2] [1]Biological Sciences (Genetics), NC State University, [2]Chemical Engineering, Texas A&M University

Axial patterning is an early step in determining the body plan by establishing the axes along which a developing organism is organized. In *Drosophila*, the anteroposterior (AP) or head-to-tail axis is determined by a large number of genes acting in a gene regulatory network (GRN). Null mutations in canonical AP patterning genes result in overt changes in AP patterning, ranging from missing body segments to death; however, AP patterning is generally robust to perturbations due to genetic and environmental variation. This robustness exposes the existence of molecular mechanisms that buffer the effects of variation.

We present data from *Drosophila* embryos of different genetic backgrounds and temperature conditions. We used the *Drosophila melanogaster* Genetic Reference Panel (DGRP), a collection of 200+ wild-derived inbred lines that are fully sequenced and suitable for association mapping, to identify novel components of the AP patterning system. Using RNAseq and confocal imaging, we aim to demonstrate molecular changes in AP patterning associated with genetic and environmental variation. Preliminary analysis of spatial expression patterns of the AP patterning genes Krüppel (Kr) and evenskipped (eve) where embryos aged at two temperatures (18 and 25 C) have been pooled suggests involvement of the dorsoventral (DV) patterning genes in AP patterning. Here, we analyze the two temperatures separately.

67W A Simple Developmental Trait as a Genetic Entry Point into Complex Behavioral Variation in Drosophila

Sarah N Ruckman, Anthony D Long Ecology and Evolutionary Biology, University of California, Irvine

Behavioral and psychiatric disorders impose staggering societal costs, exceeding $282 billion annually in the United States through healthcare expenditures, lost productivity, and related social impacts. Yet many frontline medications are only modestly effective and can cause severe side effects, so treatment often relies on trial-and-error prescribing rather than mechanistic prediction. This trial-and-error approach reflects two fundamental gaps: we lack a clear understanding of how genetic variation produces behavioral differences, and behavioral traits themselves are complex, plastic, and highly sensitive to environmental context. One potential solution is to leverage simple, scalable, and highly heritable developmental traits that share genetic architecture with more complex behaviors, providing a tractable entry point into otherwise intractable phenotypes. Here, using Drosophila melanogaster, we test whether a simple developmental decision, larval pupation site choice, measured as pupal height, shares genetic control with adult boldness, exploration, and aggression. We phenotyped 60,000+ individuals from the Drosophila Synthetic Population Resource and used an ExtremeQTL (XQTL) design, pooling the top and bottom 5% of the pupal height distribution and comparing allele frequencies between pools. This approach identified five major genomic regions, ranging in size from 0.42 to 1.14 Mb, underlying pupation height variation, collectively spanning 606 genes, and a single generation of selection yielded an estimated narrow-sense heritability of ~24%. To begin pinpointing causal loci, we are functionally testing the top 10 candidate genes using panneuronal elavGAL4/UAS-mediated manipulation. Behaviorally, flies that pupate higher show increased exploration and reduced wall-hugging in an open-field assay, whereas lowpupating flies exhibit reduced exploration and stronger wall-hugging, consistent with shared genetic control between development and adult behavior. Ongoing work will test whether the same genes jointly influence pupation height, boldness, and aggression. By revealing how shared genetic architecture between development and behavior constrains evolutionary responses to selection, this work advances basic mechanistic insight that is essential for ultimately developing more precise, mechanism-based tools to predict who is at risk for adverse effects before a prescription is written and to mitigate the enormous economic and public health burden of behavioral disorders.

68W Exposing the hidden load: Recessive deleterious variation and inbreeding depression in an outcrossing nematode.

Tom Parée, Jose S Salome-Correa, Derin Caglar, Jacqueline Jackson, Matthew rockman New York University

Recessive deleterious variants play a central role in shaping genetic load, adaptive dynamics, the evolution of sex, and species extinction risk. Despite their importance, these variants remain poorly characterized. Natural selection generally maintains recessive deleterious alleles at low frequencies, where they are rarely expressed and contribute only a small fraction of the observable genetic variance. Their study is further complicated by the fact that, although numerous, individual variants typically have small effects that are difficult to detect in isolation. Cumulatively, however, they can have strongly deleterious consequences for fitness, particularly in inbred individuals, where increased homozygosity exposes recessive effects. Here, we investigate inbreeding depression and the contribution of recessive deleterious variants in the outcrossing nematode *Caenorhabditis becei*. The moderate level of inbreeding depression in *C. becei* enabled the construction of a large panel of recombinant inbred lines (RILs); the first to date in an outcrossing nematode species. This panel was derived from six founding haplotypes sampled from a natural population and has several features that make it uniquely suited to the study of recessive variation and inbreeding depression. The limited number of founders ensures that no variant is not rare, while the crossing design generates RILs with varying degrees of relatedness. This design allows crosses among RILs to produce F1 individuals spanning the full range of inbreeding coefficients, thereby experimentally manipulating homozygosity to expose recessive variants. Quantitative genetic analyses reveal weak inbreeding depression in females but strong inbreeding depression in males, accounting for a substantially larger fraction of the genetic variance in fitness in the latter. We further assess which classes of genetic variants and functional annotations are most strongly associated with inbreeding depression. Finally, transcriptomic profiling enables us to test for the contribution of gene expression dysregulation to inbreeding depression and to test whether the recessive effects of variants are rooted in non-additive patterns of gene expression.

69W Natural Genetic Variation Shapes Metabolic Outcomes in an Insulin-Deficient *Drosophila* System

Yuyan Chen, Xuan Zhuang University of Arkansas Fayettville

Insulin deficiency is a hallmark of diabetes and a key driver of metabolic dysregulation, yet individuals with insulin deficiency can display strikingly different metabolic outcomes. This variability indicates that genetic background plays a critical role in shaping disease expression, but the genetic modifiers that influence insulin-deficient phenotypes remain poorly defined. Here, we used *Drosophila melanogaster* as a genetically tractable model and leveraged the *Drosophila* Genetic Reference Panel (DGRP) to dissect natural genetic variation in an insulin-deficient state. We generated an adult-onset insulin-deficient model by inducing targeted ablation of insulin-producing cells (IPCs) through expression of the pro-apoptotic gene *reaper*. This perturbation is introduced into diverse DGRP backgrounds via genetic crosses, producing F1 progeny with distinct genetic architectures. In these F1 flies, we quantified metabolic and physiological traits, including glucose, glycogen, and triglyceride levels and measured *dilp2* expression as a molecular readout of insulin signaling. We will use these phenotypic data to perform genome-wide association studies (GWAS) to identify genetic loci and pathways that modify insulin-deficient phenotypes. This project is expected to identify genetic modifiers that enhance or suppress insulin-deficiency–associated metabolic phenotypes, advancing our understanding of the genetic architecture underlying susceptibility to metabolic disease.

70W Genetic architecture of variation in locomotor senescence in *Drosophila melanogaster*

Bibhu Simkhada[1], Olivia R Hamilton[2], Patrick S Freymuth[3], Vijay Shankar[2], Robert Anholt[2], Trudy Mackay[2] [1]Institute for Human Genetics and Department of Genetics and Biochemistry, Clemson University, [2]Institute for Human Genetics, Clemson University, [3]Clemson University

Limited lifespan and age-related decline in survival and fitness are universal to nearly all eukaryotic species. Age-related increase in diseases and functional decline across multiple organ systems are hallmarks of senescence. A decline in mobility is among the most frequent concerns reported by the elderly. However, identifying the genetic factors that give rise to variation in age-related decline in mobility in human populations is challenging due to uncontrolled genetic backgrounds and confounding environmental factors such as the history of health, diet and exercise. *Drosophila melanogaster* presents a powerful model system to study the genetic basis of variation in locomotor senescence. We performed negative geotaxis assays on 400 lines from the recently expanded Drosophila Genetic Reference Panel of inbred wild-derived fully sequenced lines at two ages (week 0 and week 2) for a total of 230,198 flies to quantify negative geotaxis behavior at different ages, and locomotor senescence, defined as the difference in locomotor behavior between young and aged flies. We observed significant variation in negative geotaxis across both ages and sexes with broad-sense heritability estimates of ~0.40. We also observed significant variation in locomotor senescence and in sexual dimorphism for locomotor senescence. Genome-wide association analyses have identified sex- and age-specific genetic variants that contribute to variation in both negative geotaxis behavior and locomotor senescence. Notably, the genes implicated in locomotor senescence are relevant for developmental processes, including the development of visual system and nervous system, among others. Our findings reinforce the antagonistic pleiotropy theory, indicating that genes governing early developmental processes also regulate senescence later in life. Evolutionarily conserved genetic pathways underlying locomotor senescence may enable translation of these findings to humans.

71W FlyPaths: A pangenomic analytic architecture for complex trait mapping in *Drosophila*

Logan K. Cohen, Laura K. Reed Biology, University of Alabama

FlyPaths is a pangenomic genome browser that supports complex trait analysis in *Drosophila* by enabling direct comparison between the *Drosophila* Synthetic Population Resource (DSPR) and the *Drosophila* Genetic Reference Panel (DGRP). These populations are complementary, with DSPR representing multiparent haplotype mosaics analyzed through Quantitative Trait Locus (QTL) mapping and DGRP representing wild-derived inbred lines analyzed through variant-level Genome-Wide Association Studies (GWAS). Both QTL mapping and GWAS are susceptible to signal loss and effect-size inflation from significance-based selection, leading to missed associations and overestimated effects.

A genotypic census of five traits (weight, triglycerides, trehalose, development time, and survival) across DSPR and DGRP allows us to leverage the complementary strengths of GWAS and QTL mapping, enabling direct comparison of association signals derived from these fundamentally different population designs. While sample size can help mitigate the multiple-testing burden, population-specific analyses disagree in the position and scale of effects by phenotype. For some traits, this disagreement is strongly asymmetric; for example, male weight exhibits a highly polygenic architecture in DSPR QTL analyses, while corresponding GWAS in DGRP detect few loci. This discordance raises the question of whether observed differences reflect true population-specific variations in effect or are artifacts introduced by analytic multiplicity and stringent significance thresholds.

We address this by mapping DSPR QTL intervals and DGRP GWAS variant-level signals onto shared pangenomic paths in FlyPaths, rendering their outputs as queryable tracks on a pangenomic graph. This shared coordinate framework reduces reference bias and coordinate mismatch, enabling direct cross-population comparison of effect location and scale by shared genomic sequence. FlyPaths enables a biologically informed reassessment of subthreshold associations by linking path-specific association signals with known annotations and biological evidence. These analyses support traditional downstream analyses and provide a scalable foundation for Course-based Undergraduate Research Experiences in complex trait mapping.

72W Kinship as a modifier of the transcriptome, epigenetic age predictions and physiological outcomes

Hippokratis Kiaris, Kim-Tuyen Huynh-Dam, Celia Jaeger, Ioulia Chatzistamou University of South Carolina

Increased parental relatedness is common in small communities and populations that are isolated for geographical or cultural reasons. We have explored the impact of parental kinship in the epigenome and brain transcriptome by using as model closed colonies of deer mice (*Peromyscus*) that are maintained for several decades in captivity. Global analysis of DNA methylation involving ~37,000 CpGs dispersed throughout the genome showed that parental kinship, calculated by pedigree analyses, imposes epigenetic signatures in the offspring that can predict kinship relationships. Moreover, co-analysis of kinship with epigenetic age estimators showed that increased kinship delays epigenetic aging, with effects that are estimated to be at the range of about 13%. Genome-wide association studies identify specific loci that are associated with the epigenetic age acceleration, pointing to genes that modulate epigenetic aging in *Peromyscus*. Males were more sensitive than the females in the effects of parental relatedness which induces more profoundly CpG hypermethylation than demethylation. The impact of kinship in the epigenome predicted effects in gene regulation and nervous system development which were confirmed and extended by transcriptomic analyses in the brain, suggeting the operation of anatomic location and sex-dependent effects in gene expression due to parental relatedness. These effects of kinship were further confirmed by proteomic studies in the plasma of F1 and F2 hybrids between different *Peromyscus maniculatus*stocks and were corroborated by experimental studies that evaluated responses to stress. Collectively, these findings underscore the impact of parental kinship in transcriptomic, proteomic and epigenetic profiles, and identify an underappreciated modifier of physiological outcomes that is particularly pertinent in smaller communities and isolated populations.

73W Evolution towards monomorphy leads to a loss of plasticity in wing polymorphic crickets

Lourenco M.A. Martins[1], Lily J Donzeiser[2], Olivia M Ruffins[2], Jacqueline E Lebenzon[3], Lisa A Treidel[4], Abigail G Keller[2], Kristi L Montooth[5], Colin D Meiklejohn[5], Caroline M Williams[2] [1]Integrative Biology, University of California, Berkeley, [2]University of California, Berkeley, [3]University of Calgary, [4]College of William and Mary, [5]University of Nebraska, Lincoln

Threshold traits (e.g., insect-wing dimorphic, temperature-dependent sex determination, Type 2 diabetes) are widespread dimorphic traits, wherein an unobservable continuous liability is mapped to a discrete, dimorphic phenotype via a threshold. These traits are shaped by multiple genes and environmental factors, complicating predictions of how they respond to joint variation in both. The threshold trait model predicts that phenotypic plasticity only arises when the population distribution of liability spans the threshold and is therefore expected to be latent if genotypes have liability far from the threshold. We tested this prediction in the wing polymorphic cricket *Gryllus lineaticeps*, in which wing length is a heritable and plastic threshold trait with an endocrine basis. First, we confirmed the sensitive period for wing-morph determination during the last juvenile stage and validated general esterase activity as physiological proxy for liability. We then generated genetic variation in liability through quantitative genetic crosses and artificial selection and measured plasticity in wing-morph frequency and general esterase activity between split-brood individuals reared under differing environmental conditions. Artificially selected populations showed reduced wing-morph plasticity, despite maintaining plasticity in general esterase activity. This finding is consistent with a model where the underlying liability remained plastic but no longer spanned the threshold in selected populations, supporting the ability of the threshold trait model to predict phenotypic outcomes under joint genetic and environmental variation. This work provides a general framework for studying the evolutionary basis of plasticity in other threshold traits, including those relevant to human health.

74W Wild Barley Cytoplasm Confers Environmental Stability and Modulates the Phenotypic Output of the *ari-e* Semi-Dwarfing Locus

Schewach Bodenheimer[1,2], Avital Beery[1], Abelina Fuentes[3], Christine Diepenbrock[3], Eyal Fridman[4] [1]Plant Sciences Institute, [2]The Robert H. Smith Faculty of Agriculture, [3]Plant Sciences, University of California Davis, [4]Plant Sciences Institute, ARO

The interaction between nuclear and cytoplasmic genomes (cytonuclear interactions, CNIs) is a key determinant of phenotypic variation. However, harnessing cytoplasmic diversity remains a major hurdle in crop improvement. We utilize intra- and interspecific populations of wild and cultivated barley (*Hordeum vulgare* and *ssp. spontaneum*) to examine how CNIs influence the plasticity and stability of adaptive traits. Previously, we showed that wild cytoplasm contributes to trait stability. Recently, using the Cytonuclear Multi-Parent Population (CMPP), we identified marker-trait associations with cytoplasm-dependent effects. We validated these signals in new reciprocal populations, uncovering a significant three-way interaction hotspot on chromosome 5H that colocalizes with the ari-e semi-dwarfing locus (HvDep1). We demonstrate that the phenotypic expression of this major agronomic gene, particularly relating to grain protein content and germination, is not autonomous but depends on the cytoplasmic background. Additionally, we show that the wild cytoplasm provides environmental buffering, maintaining grain weight stability across different environments where cultivated cytoplasms fail. Lastly, we highlight how a genomic prediction strategy that includes these cytonuclear terms achieves cross-validation accuracies better than standard models, paving the way for cytonuclear genomic selection (cnGS) to develop climate-resilient crops.

75W The *TaGSNE*, a *WRKY* transcription factor population genetic analysis indicates adaptive variation in relation to yield contributing traits in the global wheat germplasm

Nadia Khan, Muhammad Sabir, Rifat Zubair Ahmed Genetics, University of Karachi

WRKY transcriptional regulators are important regulatory factors in plant growth and stress adaptation, so they are useful in understanding adaptive evolution in crops. In this research, we examined genetic variation at two locus on *TaGSNE*gene in different wheat (*Triticum aestivum* L.) germplasm such as Pakistani and CIMMYT wheat germplasm to gain insight into the population differentiation and adaptation. The Kompetitive Allele-Specific PCR (KASP) markers were deployed to genotype allelic variants of WRKY gene and marker-trait association was conducted on important traits related to grain yield. The *WRKY* alleles were found to be significantly associated with various yield-relevant traits including grain per spike (GPS), which showed the applicability of natural genetic variation in this locus. Genetic studies of the population showed that the allele frequency between the local lines of Pakistani germplasm and CIMMYT varied and this was attributed to the operation of selection in the process of adapting to the regions and the evolution of the breeds in the current times. The patterns that have been observed point to the role played by stable genetic variation and specific selection at regulatory genes to improve wheat. The findings that we have provided show that by combining population genetics with trait association studies one can discover adaptive alleles behind agronomically relevant traits and also gain insight into the evolutionary dynamics that may drive crop genomes under varying environmental and mating conditions.

76W Cellular Phenotypes That Mediate Context-Dependent Fitness Effects of Mutation

Kerry Geiler-Samerotte Arizona State University

Predicting evolutionary trajectories requires more than mapping the effects of mutation directly to fitness. It requires identifying the cellular phenotypes that mediate fitness and understanding how their effects change with environmental context. Importantly, many mutations impose fitness penalties that are unrelated to the biological role of the protein in which they arise. These penalties reflect “collateral” effects, which are consequences of how aberrant proteins engage with cellular systems, yet the cellular origins of these costs remain largely unresolved. Using deep mutational scans across diverse proteins and organisms, our recent work indicates that such collateral fitness costs cannot generally be attributed to protein misfolding alone. This finding shifts the focus to a broader question: which cellular processes translate molecular perturbations into fitness consequences?

We address this question using two complementary strategies. First, we apply single-cell transcriptomic profiling to yeast strains expressing mutant proteins that lack biological function, ensuring that any observed fitness effects arise from collateral interactions. At the single-cell level, we observe pronounced heterogeneity: some cells activate canonical stress-response programs, whereas others exhibit elevated transcription of transposable elements. Aberrant proteins may overwhelm proteostasis systems, indirectly permitting transposable element activation. Together, these results suggest that stress-response induction and transposable element expression represent alternative cellular outcomes of collateral protein stress, both of which may contribute to reduced fitness.

In parallel, we are performing large-scale, pooled G×E fitness assays on mutations known to exhibit collateral effects, spanning environments that place distinct demands on cellular homeostasis, such as chemical inhibition of chaperones or proteasomes. By grouping mutations according to how their fitness effects vary across environments, we infer shared cellular constraints without presupposing specific mechanisms. This approach identifies sets of mutations whose fitness costs are mitigated, exacerbated, or revealed under particular conditions, illuminating the cellular phenotypes that shape selection.

Together, this work outlines a general framework for inferring the cellular bases of mutational fitness effects from high-dimensional G×E data, advancing our ability to predict which mutations are likely to persist or be eliminated over evolutionary time, a central objective of evolutionary and quantitative genetics.

77W The Influence of Demographic History and Genetic Architecture on Complex Traits via Runs of Homozygosity

Mingzuyu Pan[1], Zachary Szpiech[2] [1]Biology, Pennsylvania State University, [2]Pennsylvania State University

Runs of homozygosity (ROH) are contiguous genomic regions where all sites are homozygous, inherited from identical haplotypes due to shared ancestry. The number and length of ROH in individuals varies based on population history and sociocultural behaviors. Although often discussed in the context of inbreeding, ROH are ubiquitous in putatively outbred human populations, and their prevalence are associated with multiple complex traits, including height and measures of lung function. Importantly, ROH have been shown to be enriched for deleterious alleles, suggesting a mechanism by which ROH prevalence can influence traits. Here we employ realistic forward-in-time population genetic simulations and a flexible quantitative model of a generic complex phenotype to explore how population history and genetic architecture influence ROH associations with a generic quantitative phenotype. We show that ROH are important for all simulated demographic histories and genetic architectures but especially when phenotypes have a recessive component. This is even more prominent when the rare-allele contribution to the phenotype is upweighted and in high-diversity populations (e.g. African). For a fully recessive phenotype, ROH can account for 25-45% of an individual’s total phenotype score, depending on demographic history and rare-allele weight. Our results emphasize the utility of ROH in helping to explain phenotype variation across different population histories and genetic architectures.

78W High-resolution, meiosis-free mapping of genetic variation with CRI-SPA-Map

Megan Lawler[1], Samuel Amidon[1], Sheila Lutz[2], Frank Albert[2] [1]University of Minnesota, [2]Genetics, Cell Biology, and Development, University of Minnesota

Genome sequence variation contributes to phenotypic trait differences, but the exact genes and variants influencing trait values remain largely unknown. Linkage limits the resolution of traditional genetic mapping approaches. We have developed CRI-SPA-Map, a strategy that combines CRISPR, selective ploidy ablation (SPA), and high-throughput phenotyping to increase the resolution of genetic mapping without relying on meiotic recombination.

CRI-SPA-Map utilizes a donor Saccharomyces cerevisiae W303 strain with a SPA cassette near each centromere and a plasmid encoding CRISPR machinery targeting the commonly-used KanMX antibiotic resistance cassette. When this W303 donor strain is mated to a BY strain of the S. cerevisiae Yeast Knockout (YKO) collection in which a given gene is replaced with KanMX, the resulting diploid repairs a double-stranded break created in KanMX using the homologous chromosome from the donor strain. SPA is then induced to remove the donor genome and generate haploid BY strains that carry small stretches of W303 DNA at the location previously occupied by the KanMX cassette.

Whole-genome sequencing of 552 isolates derived from YKO strains harboring gene deletions on the left arm of chromosome XIV showed successful replacement of KanMX with W303 DNA. Repair tracts typically spanned between 4,878 and 14,082 base pairs with variable lengths observed even across isolates derived from the same YKO strain. Using phenotyping data from over 1,000 isolates, we mapped increased growth rate in nutrient-rich media to a region containing the MKT1 and SAL1 genes. This region was four times smaller than yeast growth quantitative trait loci identified with traditional genetic mapping approaches. We further dissected this region to two causal variants in MKT1 and SAL1 and revealed that these variants show epistatic interactions that depend on the growth environment.

To broaden the applicability of CRI-SPA-Map to strain backgrounds beyond the YKO collection, we used the PiggyBac transposon to create a library of strains with a single KanMX insertion. We randomly selected 9 insertion strains to use in the CRI-SPA-Map procedure. Whole-genome sequencing of the resulting CRI-SPA-Map isolates confirmed the PiggyBac insertion was seamlessly replaced with tracts of donor DNA. Therefore, the combination of transposon mutagenesis and CRI-SPA-Map creates new opportunities to fine map genetic variation at high resolution without meiotic recombination.

79W Epistasis between gene expression noise and functional mutations shapes cellular fitness

Wei-Han Lin, Chun-Hsin Yang, Chi-Yen Wei, Jun-Yi Leu Institute of Molecular Biology, Academia Sinica

Epistasis is the phenomenon in which the phenotypic effect of a mutation depends on the genetic background. Beyond interactions among coding mutations, noncoding mutations—such as those affecting promoter activity—can also contribute to epistatic effects. Notably, the phenotypic consequence of a coding mutation can be deleterious, neutral, or beneficial depending on its expression level, indicating epistasis between regulatory and coding changes. Even within genetically identical populations, the abundance of any given protein varies among individual cells, a phenomenon known as gene expression noise. Such variability can enable rapid responses to fluctuating environments and facilitate phenotypic evolution. We therefore hypothesize that epistasis between expression noise levels and functional mutations can arise within clonal populations. To test this idea, we performed a fitness screen of mutant clones generated by random mutagenesis in budding yeast using a genetic background that allows tunable control of noise levels. Under high-noise conditions, the mutant library exhibited increased phenotypic diversity, suggesting that expression noise could enhance phenotypic heterogeneity. We then identified mutant clones whose fitness differed across noise levels. Intriguingly, one clone displayed a fitness defect that was strongly potentiated by increased noise. Using quantitative trait locus sequencing (QTL-seq), we pinpointed a mutation in *RNT1*, the gene encoding the sole RNase III in budding yeast, as the causal variant. Taken together, our findings reveal epistasis between gene expression noise and functional mutations and suggest that expression noise can uncover cryptic fitness effects, thereby contributing to phenotypic diversity during evolution.

80W Genetic control of phenology and plasticity in sunflower across six diverse field environments

Peter Innes[1], James McNellie[2], Eliza I Clark[1,3], Miles A Moore[1], Brent Hulke[1,2], Nolan C Kane[1] [1]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, [2]Sunflower Improvement Research Unit, USDA ARS, [3]Institute for Arctic and Alpine Research, University of Colorado, Boulder

Organismal responses to climate change often involve changes in phenology, particularly reproductive timing, which might be due to either genetic evolution, plasticity, or a combination. Genetic changes in mean flowering time could be beneficial for optimizing performance in specific environments, while plasticity in flowering time may support resilience across diverse environments. Here we examine the genetic basis of flowering timing and its plasticity using a diversity panel of 287 domesticated sunflower lines replicated across six diverse field environments ranging from Western Canada to the southeastern United States. We found limited overlap in genetic architecture across trials, which could be attributable to the diverse environments and the polygenic nature of flowering time regulation. Variation in plasticity also showed distinct loci compared to mean flowering time within environments, suggesting independent genetic control of average flowering time versus plasticity. The top locus associated with flowering time plasticity was a putative regulatory SNP near the sunflower homolog of EIF4A1, which was recently shown to control translation of flowering genes such as FT and FLD in Arabidopsis, suggesting that flowering time plasticity may be partly controlled via tuning of circadian clock pathways. Top candidate genes associated with variation in mean flowering time included MAF2, a flowering time repressor in the MADS-box transcription factor gene family, and DCL3, an siRNA-processing ribonuclease known to promote flowering in Arabidopsis via repression of FLC. This new understanding of the genomic regions underlying natural phenological variation in a key crop improves our understanding of plant responses to their environment and identifies targets for sunflower breeding that will enable development of more climate-resilient varieties. Notably, the separate genomic basis for mean and plasticity of phenology could enable breeders to select for these traits independently. We encourage further study of natural variation, to complement the extensive existing work on induced mutant lines and test expectations based on controlled growth chamber or greenhouse environments.

81T Leveraging Runs of Homozygosity to Identify Recessive Complex Trait Architecture

Tessa Ferrari[1], Arun Durvasula[2], Jazlyn Mooney[1] [1]Quantiative and Computational Biology, University of Southern California, [2]Population and Public Health Sciences, University of Southern California

Runs of homozygosity (ROH) are continuous stretches of homozygous genotypes created when haplotypes are inherited identically-by-descent from both parents. Rare variants within an ROH are inherited in their homozygous form, which unmasks potentially deleterious recessive variation that would otherwise be obscured in heterozygotes. Longer ROH are associated with more recent inbreeding and consanguinity, and variants within them likely have not experienced selection. Thus, ROH are enriched for deleterious variation and could be associated with complex diseases and traits. Previous empirical studies have found associations between genomic ROH burden (F_{ROH}) and complex phenotypes as evidence for recessive complex trait architecture. Reliably identifying recessive traits is an important first step to closing a long-standing gap in typical genome-wide association (GWA) methodology, since additive GWA models are underpowered for detecting rare recessive variation. However, the mechanism by which these F_{ROH}-trait associations occur in structured populations is not yet fully understood. There is no consensus on an appropriate model to reliably capture the signal while simultaneously correcting for structure. Here, we simulated traits under various architectures using multiple analytical and population genetics simulation frameworks and conducted F_{ROH}-trait association tests. We then evaluated their ability to accurately capture signals of trait recessivity in order to elucidate the nuanced relationship between ROH, complex traits, and population dynamics. As expected, we find that as trait narrow-sense heritability increases, the power to detect recessive traits increases. However, under simulation frameworks using more realistic population dynamics, we find decreased power and a false discovery rate that deviates from expectations, especially in more structured populations. Finally, we implement F_{ROH}-trait association tests in 3 empirical populations from the UK BioBank across 22 health-associated complex traits. In a White British population and South Asian population, we detect 6 shared trait associations (e.g. BMI, Vitamin D, etc.) as well as 6 population-specific ones (e.g. LDL in White British, Hemoglobin in South Asian). We detect no associations in a Black Caribbean population, perhaps due to sample size or complex structure from recent admixture.

82T Genetic and environmental interactions outweigh mitonuclear coevolution for complex traits in *Drosophila*

David M Rand[1], Leah J Darwin[2], Faye A Lemieux[1], Yevgeniy Raynes[1] [1]EEOB, Brown University, [2]Center for Computational Molecular BiologyBiology, Brown University

The interdependent relationship between mitochondrial and nuclear genomes is a powerful model for under-standing how epistasis shapes the architecture and evolution of complex traits. Once considered a neutral marker, mitochondrial DNA variation is now recognized as critical to phenotypic evolution because of its epistatic interactions and history of coevolution with the nuclear genome. A central challenge in evolutionary genetics is to quantify the relative importance of stabilizing and directional selection shaping complex trait distributions within and among species. Both can act on interacting and/or co-evolving genes contributing to quantitative traits, but resolving their relative roles is complicated by the complex architecture of most traits. Here, we use a panel of 90 Drosophila mitonuclear genotypes to quantify the relative contributions of mitochondrial, nuclear, and environmental variation and their interactions to four metabolically demanding complex traits. We sample both within-species and between-species mitochondrial variation and observe stronger interaction effects attributable to within-species variation, consistent with stabilizing selection maintaining mitonuclear function. Additionally, culturing the flies on a mitochondrial Complex I inhibitor, rotenone, reveals significant genotype x environment (GxE and GxGxE) interaction effects, providing insight into how genetic variation can be maintained across changing environments. Our results have broader implications in medicine, where mitochondrial DNA donors with longer purifying selection histories may be safer for mitochondrial replacement therapies.

83T Additive genetic control of gene expression underlying wood chemistry in *Populus trichocarpa*

Chanaka R Abeyratne[1], David Macaya-Sanz[2], Brian Stanton[3], Gerald A Tuskan[4], Kerrie W Barry[5], Gancho T Slavov[6], Stephen P DiFazio[7], Daniel A Jacobson[8] [1]Oak Ridge National laboratory, [2]3Department of Forest Ecology & Genetics, CIFOR-INIA, [3]Poplar Innovations, [4]Bioscience Division & Center for Bioenergy Innovation, Oak Ridge National laboratory, [5]Joint Genome Institute Lawrence Berkeley National Laboratory, [6]6Department of Forest Engineering, Resources, and Management, Oregon State University, [7]Department of Biology, West Virginia University, [8]Biosciences Division, Oak Ridge National Laboratory

Understanding the genetic architecture of complex traits such as wood chemistry remains challenging, particularly when regulatory relationships are inferred using expression-based networks that do not reflect additive inheritance. In long-lived, undomesticated species such as *P. trichocarpa*, extensive standing genetic variation, weak linkage disequilibrium, and high polygenicity limit population scale dissection of genetic architecture for adaptive traits. Here, we replace traditional co-expression networks with additive genetic correlation (r_G)-based networks to identify regulatory modules consistent with quantitative genetic inheritance, using a 7×7 full-factorial multi-parental cross that captures population-scale trait variation. Additionally, we use multiplex network modelling to identify key master regulators of wood chemical composition (lignin syringyl/guaiacyl (S/G) ratio). Module structure and trait variation were dominated by additive genetic effects, with minimal dominance, consistent with quantitative genetic expectations and accurate mid-parent prediction. To link additive modules to regulatory mechanisms, we constructed gene co-regulatory networks using predicted additive genetic values rather than co-expression. These, r_G-based regulatory modules were smaller, more connected, and more functionally coherent than conventional co-expression modules, and showed stronger associations with S/G. In parallel, we mapped cis- and trans-eQTLs while accounting for relatedness among offspring and organized significant associations into a bipartite SNP–gene network. Community detection within this network revealed densely connected regulatory neighborhoods enriched for S/G-associated gene modules, highlighting shared regulatory control. Integrating evidence from (i) r_G-based regulatory modules (ii) eQTL network topology and (iii) r_G between expression and S/G, we identified two putative master regulator transcription factors (TFs) *PtrSND1* and *PtrMYB018* and their downstream counterparts. Embedding these candidates within an *Arabidopsis* multi-omic multiplex network revealed a conserved regulatory core linking multiple SND and MYB TFs integrated with metabolic and signaling pathways that integrate abiotic and biotic stress response. These findings demonstrate that S/G variation arises from a largely additive and hierarchically organized regulatory architecture and establish r_G-based network inference as a general framework for identifying additive regulatory modules in undomesticated species.

84T ESCALATOR: an efficient bioinformatics pipeline to harmonize and calculate polygenic scores in large-scale datasets

Meng Lin[1,2], Matthew Fisher[2], Christopher R Gignoux[3] [1]Department of Biomedical Informatics, University of Colorado Anschutz, [2]Colorado Center for Personalized Medicine, University of Colorado Anschutz, [3]University of Colorado Anschutz

Applying weights to calculate polygenic risk scores (PRSs) is increasingly common in biomedical research, with the purpose to test, evaluate, or potentially deploy externally developed PRS models. Although the process is considered straightforward, there are practical challenges from data heterogeneity between the test and the external training dataset, requiring systematic harmonization of weights before score calculation. The discrepancies may include strand mismatches, non-biallelic codes at overlapping genomic positions, and genome build differences, all of which can arise from heterogeneous dataset generation workflows and frameworks. Here we present ESCALATOR, a convenient, open-source pipeline that efficiently harmonizes and applies external PRS weight files to a test dataset of interest, and leverages existing PLINK software for score calculations. The pipeline is memory-efficient when applied to large biobank-scale datasets, and containerized versions are available with the widely-used Docker and Singularity platforms. ESCALATOR aims to streamline and accelerate risk score calculations for large datasets processed on most computational cluster environments, including ongoing deployment into the PRIMED consortium's ANVIL platform for broader applications. We benchmark ESCALATOR and show its efficiency on genome-wide datasets from the tens to hundreds of thousands of individuals.

85T Inferring selection on complex traits in British individuals using marginal coalescent trees

Obadiah Mulder, Michael Edge Quantitative and Computational Biology, University of Southern California

Selection on complex traits is difficult to detect using classic methods. Variants affecting highly polygenic traits generally have small effect sizes and selection at any given locus is usually weak. Several recent methods have addressed this problem by using genome-wide association study (GWAS) information to construct polygenic scores for traits of interest and then testing these polygenic scores for evidence of selection. One set of methods in this family, due to Edge and Coop (2019), uses marginal coalescent trees to estimate the historical time course of population-mean polygenic scores, giving information about both the timing and strength of selection. Further work by Peng and colleagues (2025) demonstrated that the marginal coalescent trees estimated by SINGER perform very well at this task. However, there have been few efforts to assess selection in empirical data using these tools. Here, we use family-based allelic effect size estimates to form population-mean polygenic scores for several complex traits, and we test them for evidence of directional selection among the ancestors of the British (GBR) subset of 1000 Genomes Project participants.

86T Genetic architecture of PFAS toxicity revealed by natural variation in *C. elegans*

Tess Leuthner[1], Jameson Blount[2], Abby Fernandez[2], Jenna Valley[2], Ryan Baugh[3] [1]The O'Neill School of Public and Environmental Affairs, Indiana University, [2]Duke University, [3]Biology, Duke University

Individuals in a population vary in their susceptibility to environmental challenges, yet we lack a mechanistic understanding of how genetic variation shapes responses to toxicants. Per- and polyfluoroalkyl substances (PFAS) are persistent environmental contaminants associated with adverse health effects, but the genetic basis of differential susceptibility remains poorly understood. Here, we leverage natural genetic variation in *Caenorhabditis elegans* to characterize the genetic architecture underlying responses to ten structurally diverse PFAS. We used a pooled-population sequencing approach to quantify strain frequencies among 192 wild *C. elegans* isolates following 96 hours of PFAS exposure. Changes in strain frequency relative to control conditions revealed sensitive and resistant genotypes for nine of the ten chemicals tested. Narrow-sense heritability ranged from 0.01 to 0.42, indicating substantial variation in the genetic architecture underlying PFAS responses. Two emerging PFAS, PFEESA and PFBS, exhibited largely additive genetic variance (h^2 = 0.42 and 0.34, respectively), whereas PFBA showed low additive variance (h^2 = 0.03) but high broad-sense heritability (H^2 = 0.86), suggesting a strong contribution of non-additive or epistatic interactions. Genome-wide association mapping identified 23 quantitative trait loci (QTL) across the ten PFAS, most of which were chemical-specific. Notably, two loci were shared among PFAS with similar structural features, pointing to shared mechanisms of action and potential structure-toxicity relationships. To investigate molecular responses to PFAS exposure, we performed RNA-sequencing in the N2 reference strain. Differential expression analysis revealed transcriptional responses involving xenobiotic metabolism, stress-response pathways, and lipid homeostasis. Several genes within GWAS loci were also transcriptionally responsive to PFAS exposure, suggesting candidate pathways that may mediate genetic susceptibility. Together, these results demonstrate that PFAS responses are genotype-dependent and that distinct chemical structures operate through different mechanisms of action. This work establishes a scalable framework to link chemical structure to genetic susceptibility and molecular mechanism in a naturally diverse *C. elegans* population.

87T Complex traits: genetic effects from social partners and from the microbiome

Helene Tonnele[1], Francesco Paolo Casale[2], Abraham A Palmer[3], Amelie Baud[1] [1]Centre for Genomic Regulation, [2]Helmholtz Munich, [3]University of California San Diego

A major goal of modern biology is to understand how genetic variation gives rise to differences among individuals in complex traits and disease risk. Most studies focus on the effects of an individual's own genotype ("direct genetic effects", DGE), yet individuals live and evolve in rich social and microbial contexts. Our research expands the genetic architecture of complex traits to include **genetic effects arising from the genomes of social partners ("indirect genetic effects", IGE) and from the gut microbiome.**

Using multiparental populations of laboratory rodents as models, we combine genome, phenome, microbiome and sociome data with advanced quantitative genetics models and systems biology approaches. A key objective of our work is to uncover the mechanisms of IGE—specifically, which genes and traits of social partners give rise to these effects.

I will first present the results of a phenome-wide genetic framework that identifies proxy phenotypes for the heritable traits of social partners mediating IGE by estimating genetic correlations between IGE on focal phenotypes and DGE on measured traits. Applying this approach to two large, outbred mouse datasets comprising hundreds of behavioural and physiological phenotypes, we find that behavioural traits are neither more affected by IGE nor better proxies for the traits mediating them, challenging the prevailing behavioural-centric view of social effects. Instead, immune, metabolic and growth phenotypes are both affected by IGE and informative proxies for their underlying mechanisms. We hypothesize that these signals reflect the social transmission of gut microbes. I will present a second line of evidence supporting this hypothesis: significant indirect genetic effects on gut microbiome phenotypes detected in a rat multiparental population.

This work highlights new opportunities for systems biology approaches to bridge genetic and environmental research.

88T *The center for genetics, genomics, and epigenetics of substance use disorders in outbred rats*

Abraham A. Palmer Psychiatry, UCSD

For over 12 years our center has been collecting genotype and phenotype data in outbred heterogeneous stock (HS) rats. HS rats were created in 1984 by interbreeding 8 inbred laboratory strains and have been maintained as outbred for ~110 generations since then. In the past 12 years we have genotyped and phenotyped ~25,000 HS rats. Genotyping is accomplished by light whole genome sequencing followed by imputation which yields millions of SNPs with an accuracy >99.75% compared to 30x whole genome sequencing. We can also express genotypes as imputed founder haplotype dosages. Currently we are developing methods to impute non-SNP features such as indels, tandem repeats and structural variants.

Phenotyping is focused on biomedically relevant traits, especially behavioral traits, but a wide range of other traits have been examined including body weight, body length, BMI, optical pressure, blood glucose, bone density, leg muscle weight, skull morphology, organ weights, microbiome, and metabolome.

We have also collected gene expression data from almost 3,000 samples, including brain and other tissues. Gene expression data is hosted at RatGTEx.org and can be explored online or downloaded for bulk analysis. We have used Pantry (Pan-transcriptomic phenotyping) to explore multiple expression modalities: gene expression, isoform ratios, splice junction usage, alternative TSS/polyA usage, and RNA stability. Finally, we have used Fusion to perform TWAS across all phenotypes using all of the Pantry expression traits.

Finally, we have introduced RATTACA, which uses polygenic prediction to identify HS rats that are likely to show extreme phenotypes, allowing, for example, examination of treatment naive HS rats that are genetically predicted to self-administer large or small amounts of cocaine. This can be useful for examining putatively correlated traits e.g. traits for which exposure to cocaine could alter the second trait.

The center is intended to be a national resource and can provide technical advice, data, tissues and live animals to support novel quantitative genetic analyses by members of this community.

89T Evolved differences in mitochondrial function between temperate and tropical house mice

Mal Ballinger Ecology & Evolutionary Biology, Cornell University

Mitochondria play a central role in linking cellular metabolism to environmental conditions, yet their contribution to rapid environmental adaptation remains poorly understood. House mice provide a powerful system to address this question because they have recently colonized and adapted to a wide range of climatic environments. Here, we use an integrative approach combining mitochondrial bioenergetic assays, liver RNA-seq, and mitochondrial DNA (mtDNA) analyses to investigate how mitochondrial function varies across populations of house mice. We show that mice from tropical environments exhibit greater mitochondrial respiratory capacity than mice from temperate environments. We further integrate mitochondrial physiology with gene expression and mtDNA copy number to link evolved differences in mitochondrial performance with genomic and regulatory variation. Together, these results suggest that evolutionary changes in mitochondrial function may represent an important mechanism enabling rapid adaptation to novel environments.

90T Elucidating the genomic architecture governing chloroplast-nuclear stoichiometric balance

Evita Chee, Joel Sharbrough University of California, Santa Barbara

Plant genomes are partitioned into three subcellular compartments: the nucleus, mitochondria, and chloroplasts, which together encode multi-subunit enzyme complexes that carry out respiration and photosynthesis. Interactions between nuclear-encoded and cytoplasmically encoded genes and gene products (i.e., cytonuclear interactions) are critical for producing cellular energy, and perturbations to one genome have profound consequences for the other(s). For example, plants respond to whole-genome duplications by increasing the number of chloroplast and mitochondrial genomes per cell, indicating that cytonuclear interactions are dosage sensitive. The mechanism whereby plants sense and regulate cytonuclear gene dosage balance, and the resulting impacts on bioenergetics, are entirely unknown. To characterize the genomic architecture of cytonuclear gene dosage balance, we estimated chloroplast and mitochondrial genome copy number from the 1001 *Arabidopsis* genomes project sequencing data and carried out a genome-wide association analysis (GWAS). Our GWAS identified 34 SNPs that were significantly associated with chloroplast genome copy number, spread across 17 genomic regions. We annotated these SNPs finding two that are within 1kb of previously known cytonuclear interacting genes. We also described the genes and conserved non-coding regions harboring the remaining 32 SNPs. In parallel, we are also performing a quantitative trait loci (QTL) mapping experiment, in which genetically and geographically diverse *Arabidopsis* accessions that exhibit high vs. low chloroplast genome copy number were crossed. The resulting F_2 progeny were reared in a common garden, assayed for photosynthetic traits using a MultispeQ, plastid genome copy number was estimated using qPCR, and then sequenced for mapping analysis. Together, our GWAS and QTL mapping results will provide a set of candidate gene regions for functional molecular characterization. More broadly, our work will provide important advances to our understanding of cytonuclear epistasis, which is critical for the long-term goal of improving photosynthetic efficiency in crop plants.

91T The Limits of PCA: The Origin and Propagation of Residual Confounding in GWAS

Walid Mawass[1], Jennifer Blanc[2], Ahmed Selim[3], Jeremy J Berg[1] [1]Human Genetics, University of Chicago, [2]Data Science Institute, University of Chicago, [3]Graduate Program in Biophysical Sciences, University of Chicago

Principal component analysis (PCA) is the standard method for correcting population stratification in genome-wide association studies (GWAS). Yet residual confounding often persists after adjustment, biasing polygenic scores and evolutionary inferences. Here, we develop a theoretical framework to quantify the limits of PCA-based correction, showing that residual bias arises from geometric misalignment between sample principal components and latent ancestry axes. This misalignment is most severe in a transition regime where population structure is statistically detectable but too weakly estimated to be precisely removed. Consequently, residual bias peaks near the detection threshold rather than decreasing monotonically with signal strength, demonstrating that a principal component's statistical significance is a poor proxy for its adequacy as a covariate. By exploring various confounding architectures, we show that stratification is particularly difficult to mitigate when environmental variation aligns with many minor ancestry axes, as these subtler dimensions remain poorly estimated. Finally, we establish that systematic error in meta-analysis is constrained by the estimation accuracy of its constituent cohorts. Because meta-analysis relies on local PCA, it preserves the high residual bias of smaller contributors rather than leveraging the aggregate sample size to improve ancestry estimation. These preserved residuals are amplified by the consortium's large effective sample size, providing a mechanistic explanation for the persistent inflation of LDSC intercepts. These results establish the theoretical bounds of confounding control in massive-scale genomics.

92T Thrifty Gene Signatures Underlying Diet-Dependent Variation in Drosophila Development

Yulin Bai, Sumaira Shabbir, Yuyan Chen, Mutheshree Rajesh, Xuan Zhuang University of Arkansas

Nutritional and developmental plasticity are essential for organisms adapting to fluctuating environments, and diet-induced phenotypic variation has long been a major target of natural selection. However, the genetic basis of diet-induced developmental plasticity remains unclear. The thrifty gene hypothesis proposes that alleles once advantageous for energy storage now increase disease risk in nutrient-rich environments. In this study, we quantified key metabolic, physiological, and life-history traits in a subset of lines from the Drosophila Genetic Reference Panel (DGRP), measured in larvae and adults raised on diets with differing nutritional content. We estimated trait heritabilities and correlations across diets and life stages, identifying development time as the trait with the highest heritability and the strongest diet effect. We therefore prioritized development time for genetic mapping, complementing the DGRP with a Drosophila recombinant outbred population to increase power and sensitivity. Genome-wide association analyses within and across diets revealed candidate loci enriched for conserved metabolic regulators, including several with human orthologs associated with metabolic diseases. Consistent with predictions of the thrifty gene hypothesis, we also detected diet-specific shifts in development time, suggesting the presence of "thrifty" variants that respond sensitively to nutritional availability. These findings advance our understanding of the genetic architecture of diet-induced developmental plasticity and highlight the value of combining natural panels with outbred mapping populations to dissect gene-by-environment interactions.

93T Search for durable wheat stem sawfly resistance QTL in spring wheat

Isha Isha, Jason Cook, Hwa Young, Nancy Blake Plant Sciences and Plant Pathology, Montana State University

Wheat stem sawfly (WSS) *Cephus cinctus* Norton is a native insect pest in North America, that causes major yield loss in wheat. The lifecycle of this opportunistic pest has evolved to coincide with the lifecycle of wheat. The WSS lays eggs in wheat stems that produce larvae that feed within the stem during the growing season and eventually cuts the stem at plant maturity, making it susceptible to lodging. In recent years damage caused by WSS has expanded to Colorado, Nebraska, Kansas and Wyoming. Even though numerous control measures have been evaluated, there are limited options that are economical and practical for controlling this pest. The primary means of controlling WSS is through the use of plant host resistance that is provided by the solid stem trait. Presence of solid stems reduces the oviposition of eggs by female WSS and acts as a barrier to WSS larvae movement. Unfortunately, solid stem varieties have a yield drag compared to hollow stem varieties, and the solid stem trait is not stable across environments making them prone to WSS damage. Large effect quantitative trait loci (QTL) have been identified for the solid stem trait including *Qss-msub-3BL* and a QTL on chromosome 1B. In this study, we aim to identify additional sources of WSS resistance by performing QTL analysis on bi-parental recombinant inbred line (RIL) populations. The RIL populations are made by crossing a susceptible spring wheat variety, Hi-Line, and resistant landraces/cultivars that showed resistance to WSS in the field and/or had solid stems. The RIL populations have been genotyped and phenotyped in the field for solid stem score and stem cutting. We will show preliminary results from our QTL analysis with the ultimate goal of improving the durability of wheat stem sawfly resistance.

94T Predicting the body composition change in response to an aerobic exercising intervention on young adults, using genomic DNA, phenotypes and blood metabolomic profile before the intervention.

Yifei Li[1], Guanqi Lu[1], Molly S Bray[2], Ana I Vazquez[1] [1]Epidemiology and Biostatistics, Michigan State University, [2]Nutritional Sciences, The University of Texas at Austin

This study investigated the ability of plasma metabolomic and genomic information to predict inter-individual changes in body composition, together with initial physical status, following a structured aerobic exercise intervention. We analyzed data from 820 participants enrolled in the Training Interventions and Genetics of Exercise Response (TIGER) Study. TIGER recruited sedentary young participants (aged 18 to 35 y) to a 15-week supervised endurance training program and includes multiple measures of body compositions at baseline (before the program) and after the program.

In this study we used genetic data, metabolomic data at baseline, and other phenotypes to predict body change following exercise training. The untargeted metabolomic assay identified 5,723 metabolites measured at baseline. The genomic data includes 129246 SNPs after QC. Body size measurements, and body composition at baseline was also utilized for prediction, including Body Mass Index (BMI), waist circumference, height, fat mass, lean mass, bone mass, and percent body fat. The physical response to exercise was measured as heart rate responses, body size, and body composition change. Change was defined as the difference between the measurement before and after the intervention divided by the measurement before the intervention.

We developed a novel bidirectional transformer-based deep learning framework to capture complex nonlinear relationships between high-dimensional metabolite responses and downstream physiological remodeling. The model incorporated rotary positional encoding and Query and Key normalization to enhance training stability and generalization. Predictive performance will be evaluated using rigorous cross-validation against established methodologies, including penalized regression and tree-based methods. Model performance is evaluated with predictive correlation for change in BMI, %body fat, %trunk fat, and waist-hip ratio change due to exercise.

The transformer improved predictive accuracy across a broad range of variables measuring the change in body composition as response to the exercise intervention. When using baseline metabolomic data, the transformer model achieves an average predictive correlation of 0.336, significantly outperforming conventional Lasso approach (r=0.224). In this study we combine genomic, metabolomic and baseline phenotype to predict response to exercise. Individual phenotypes demonstrated particularly high predictability, including increases in lean mass change and in an increase of bone mass, highlighting substantial metabolic determinism in exercise adaptation.

Keywords: metabolomics, deep learning, transformer, body composition change.

95T Genome-wide association studies of puberty and fertility traits in US beef heifers

Garrett L Franklin[1], Dotun L Olaoye[1], Jordan L Thomas[2], Jared L Decker[1] [1]Animal Science / Genetics Area Program, University of Missouri, [2]Animal Science, University of Missouri

Reproductive failure is an important fitness trait and contributes to significant losses annually within the beef cattle industry. The lowly-heritable nature of reproductive traits has delayed development of genetic predictions, thus genetic progress has been limited. Selection tools, such as the Heifer Pregnancy expected progeny difference (EPD), have been implemented to identify females more likely to become pregnant within a normal breeding season. However, there is a clear need for the development of selection tools based on quantitative measures of traits directly measuring puberty and fertility, rather than relying on low-information binary measures. Further, dissecting the genetic underpinnings of complex traits related to female reproduction would enable the industry to more accurately predict and select for reproductive success. To identify quantitative trait loci (QTL) influencing puberty and fertility, a series of genome-wide association studies were conducted using multiple SNP array densities and imputed 800k genotypes from more than 9,000 of Angus, Red Angus, Hereford, and Bos indicus–cross descent. Phenotypic records include reproductive tract score (RTS), pelvic width, pelvic height, and days to conception, defined as the number of days required for a heifer to become pregnant during a normal breeding season. Breed-specific analyses were performed in GCTA using linear mixed models, while multivariate GWAS incorporating all traits were conducted in GEMMA to improve power for QTL detection. In addition, information across breeds was integrated into a pooled dataset to fit both univariate and multivariate mixed models. Differences in phenotypic variance among genotype classes will also be evaluated to identify variance QTL (vQTL), which may reflect genotype-by-environment or genotype-by-genotype interactions. Finally, a robust meta-analysis combining p-values from breed-specific GWAS further increases power to detect QTL. Collectively, these analyses advance our understanding of the genetic basis of puberty and fertility and support the development of more effective selection strategies for improving reproductive performance in beef cattle.

96T Epistatic interactions detected by GWAS in thousands of F1 families establishes the molecular identities of phenotypically impactful natural variants

Brian Dilkes[1,2], Amanpreet Kaur[1], Rajdeep S Khangura[3] [1]Biochemistry, Purdue University, [2]Center for Plant Biology, Purdue University, [3]Plant and Agroecosystem Sciences, University of Wisconsin-Madison

Mutant phenotypes are valuable resources for studying genetic mechanisms that regulate plant physiology, growth, and development. They serve as links between genotype and phenotype, allowing the examination of genetic variations and their effects on the phenotype. The establishment of molecular identities for natural variants, and even the delimitation of reasonable candidates, is greatly enhanced when the pathways affected by natural variants are known. This can be established by mapping modifiers that are epistatically determined by a mutant allele. This is most easily achieved with dominant mutants where F1 families can be phenotyped directly. In maize, mutants can be crossed to diverse maize lines to generate an F1 association mapping population (FOAM). Each F1 family segregates 1:1 for wild type and mutant plants, permitting the observation of background effects in F1 hybrid siblings that differed at the mutant locus. GWAS on the F1 families permits the discovery of epistatic interactions conditioned on the mutant genotype. This can help expand the membership of genes contributing to the pathway delimited by mutant, and annotate the natural variant by pathway, when the identity of the mutant gene is known. The extensive genetic diversity captured in the FOAM populations allows for high-resolution mapping of genetic modifiers that affect the mutant phenotype. Integrating these data with the effects of genetic variation on gene expression provides insights into molecular mechanisms underlying the altered phenotype. We have employed this approach on multiple dwarf, lesion-formation, leaf patterning and polarity, and metabolic mutants. This has identified natural variation in multiple core metabolic and developmental pathways that contribute to maize growth. Via this approach we have established the molecular mechanisms for multiple natural variants as cis-acting expression polymorphisms, identified alleles at known protein-protein interacting partners, and annotated natural variants in core development regulators via the detection of epistatic interactions with a subset of mutants in F1 GWAS populations. This demonstrates the value of both traditional developmental and biochemical genetics and identifies natural alleles of phenotypic consequence in key developmental regulators by GWAS.

97W Representation in genetic studies affects inference about genetic architecture

Jared Cole, Shane Rybacki, Samuel P Smith, Olivia S Smith, Arbel Harpak University of Texas at Austin

Knowledge of a trait's "genetic architecture," namely the joint distribution of allele frequencies of causal variants and the direction and magnitude of their effects, is essential to understanding its evolution and underlying biology. Inferences about genetic architecture are based on data collected in heterogeneous ways in cohorts recruited through heterogeneous mechanisms. This heterogeneity, in turn, results in differences in genotype, environment and trait distributions across cohorts. For instance, the UK Biobank (UKB) aimed for broad population representation, whereas FinnGen drew heavily from clinical registries enriched for diagnosed health conditions. Here, we asked whether representation in genetic studies influences inferences about genetic architectures. Using GWAS data from the UKB, FinnGen, and All of Us (AoU), we find that some summaries of a trait's genetic architecture, such as effective polygenicity, vary little across biobanks. Others, like SNP heritability, are systematically lower in one biobank (AoU) than in another (UKB) across almost all traits examined, even when matching samples such that they have similar genetic ancestry compositions. This result aligns with other recent evidence of lower heritabilities in biobanks with disease-enriched recruitment than in population-based biobanks, which may consequently reduce power to detect genotype–phenotype associations in the former. We highlight a third case, where a summary of genetic architecture varies considerably but not systematically across traits and biobanks. Such is the case for the mean direction of allelic effects ("sign bias"). For example, 72% of rare minor alleles affecting type 2 diabetes risk are inferred to be risk-increasing based on AoU data, while nearly 100% are inferred to be risk-increasing based on UKB data. We hypothesize that the inferred sign bias is heavily influenced by the skewness of the trait distribution in the study and is otherwise independent of other study or trait characteristics, including whether the trait is binary or quantitative. We provide strong support for this hypothesis through simulations and data from the three biobanks: the variation in inferred sign bias across traits and biobanks is explained remarkably well (81% and 97% of variance explained for trait-associated and for a random set of SNPs, respectively) solely by the trait's skewness in the study cohort, with residual biobank-specificity explaining little (incremental adjusted R^2=2.4%). These results illustrate that inferences about the map between genetic and phenotypic variation can depend on representation in genetic studies in surprising ways.

98W Improving polygenic score prediction for non-European ancestry groups through transfer learning.

Hao Wu[1], Paulino Pérez-Rodríguez[2], Michael Boehnke[3], Yuehua Cui[1], Xiaoyu Liang[1], Ana Vazquez[1], Gustavo de los Campos[1] [1]Michigan State University, [2]Colegio de Postgraduados, [3]University of Michigan

The advent of large biobanks has substantially increased the accuracy of polygenic scores (PGS). However, most existing PGSs were derived from European-ancestry data and often exhibit reduced predictive performance when applied to individuals of non-European ancestries. Transfer Learning offers a promising strategy to address this limitation by leveraging information learned in one population to improve prediction in another. Here, we introduce GPTL, an R package that implements three Transfer Learning based approaches for developing PGS: (1) gradient descent with early stopping, (2) a penalized regression model that shrinks variant-effect estimates toward prior values, and (3) a Bayesian method with a finite-mixture prior that enables integration of multiple prior sources of information. Using both simulated data and real data from the UK-Biobank and All of Us, we demonstrate that PGS generated with GPTL's Transfer Learning algorithms consistently outperform single-ancestry PGS and, in many settings, match or exceed the performance of multi-ancestry ensemble-based PGS. Our software can be used with either individual genotype-phenotype data or summary statistics from genome-wide association studies.

99W The site-frequency spectrum under selection and time-varying demography

Anastasia Lyulina[1,2], Benjamin Good[1,2,3] [1]Department of Biology, Stanford University, [2]Department of Applied Physics, Stanford University, [3]Biohub

Demographic history and natural selection both influence the site-frequency spectrum of new mutations, but their joint effects remain difficult to disentangle — especially in populations far from equilibrium. In this work, we derive an analytical expression for the frequency spectrum of rare alleles for arbitrary time-dependent population sizes and deleterious fitness effects. This forward-time approach allows us to trace the trajectories of mutations contributing to different parts of the spectrum. We find that rapid population growth can produce an abundance of older deleterious variants at intermediate frequencies relative to the neutral expectation, resulting in a non-monotonic ratio of nonsynonymous-to-synonymous mutations. By applying these results to recent demographic reconstructions of European human history, we show that these nonequilibrium effects are likely to play an important role in shaping the observed distribution of deleterious variants. These results provide a new theoretical framework for interpreting the site-frequency spectrum in populations with complex demographic histories, and highlight scenarios where selection and demography interact in non-intuitive ways.

100W Biobank-scale visualization and interactive exploration of ancestral recombination graphs with Lorax

Pratik Katte, Russ Corbett-Detig Biomolecular Engineering and Bioinformatics, University of California Santa Cruz

We present Lorax, a GPU-accelerated, browser-based platform for real-time, interactive visualization of population-scale Ancestral Recombination Graphs (ARGs). ARGs capture the full history of coalescence and recombination underlying observed genetic variation and are increasingly central to analyses of demography, natural selection, admixture, and disease association. As inference methods now scale to biobank-sized cohorts, existing visualization tools remain limited to small sample sizes and static representations, leaving a critical gap between inference and biological interpretation.

Lorax addresses this gap with a unified interactive environment integrating genome-wide navigation, coalescent time, local tree topology, and sample metadata. Users can traverse genomic coordinates, inspect local genealogies at recombination breakpoints, and dynamically filter or color samples by population labels, phenotypes, or other annotations. Using GPU-accelerated rendering and a streaming server architecture, Lorax scales to ARGs with over one million samples while maintaining interactive performance. By making complex ancestry structures directly navigable, Lorax lowers the barrier to ARG-based analysis and supports exploratory investigation of population history, evolutionary processes, and disease-relevant variation. Lorax is open-source and freely available at lorax.ucsc.edu.

101W Modeling Local Ancestry Covariance to Infer the Timing of Denisovan Admixture Events

Sarah A Johnson[1], Nick Patterson[2], Priya Moorjani[3] [1]Computational Biology, University of California, Berkeley, [2]Broad Institute, [3]University of California, Berkeley

Gene flow from our extinct hominin cousins, Denisovans, has shaped the landscape of the modern human genome and contributed to adaptive variation in present-day populations. Yet our understanding of the timing and geographic context of Denisovan admixture remains poorly resolved. Existing approaches for dating admixture events typically rely either on linkage disequilibrium (LD) among a limited set of ascertained archaic-informative markers or on local ancestry inference, which has reduced power to detect short introgressed fragments. These limitations are particularly acute for Denisovan ancestry, which is rare (~0.1–4%) in many populations and unevenly distributed across the globe. Here we introduce a composite method that uses LD in local ancestry across the genome. Our method models the decay of covariance in local ancestry across genomic distances, leveraging the recombination clock. By using genome-wide local ancestry information rather than sparse Denisovan informative positions, our method accesses orders of magnitude more data with an effective gain in power of nearly 1,000-fold relative to SNP-based methods. Additionally, by computing pairwise covariance across multiple individuals rather than directly modeling inferred archaic fragments, our method robustly captures Denisovan ancestry decay without relying on precise local ancestry inference in any single individual. We perform simulations under a range of demographic scenarios and demonstrate that our method works reliably to date both Neanderthal and Denisovan admixture, even in cases of low admixture proportions (< 0.5%). We apply this method to a dataset of over 4,700 published whole genome sequences from across Asia and Oceania to infer the timing and structure of Denisovan admixture events in modern humans. Together, our study provides a powerful new method for dating admixture, and sheds light into the complex history of modern human-Denisovan interactions and human dispersal following the Out-of-Africa migration.

102W Pangenome architecture of quantitative trait loci in Darwin's finches

Jiayang Sun[1], Matteo Sebastianelli[2], Leif Andersson[2], Carlos A Valle[3], Erik Enbody[1] [1]Cornell University, [2]Uppsala University, [3]Universidad San Francisco de Quito

Pangenomes provide a powerful framework for representing genetic variation beyond a single linear reference, capturing complex and structural variants that are poorly represented in SNP-based analyses. While pangenomic methods have reshaped analyses in human and other model systems, their potential to advance quantitative genetics, particularly in non-model organisms, remains largely unexplored. This gap is especially relevant for quantitative trait loci (QTLs), where causal variants are often unknown and may lie outside canonical SNP classes. Darwin's finches in the Galápagos Islands provide an exceptional natural system for addressing this challenge: decades of ecological, phenotypic, and genomic work have mapped variation in beak morphology and body size to a small number of well-defined QTLs, yet the underlying causal variants within these regions may be complex and remain unresolved.

We constructed a graph-based pangenome for Darwin's finches to resolve the genetic architecture of these QTLs beyond SNP-based representations. Using high-quality PacBio HiFi assemblies from six representative finch species (four ground finches, one tree finch, and a continental outgroup), we generated a pangenome that resolves shared and species-specific variation across the radiation. We identify approximately 28 million variants genome-wide, of which nearly 20% fall outside canonical SNP classes, representing a substantial component of genetic variation that has been largely inaccessible to reference-based analyses.

Across all previously identified QTLs associated with beak morphology and body size, we observe extensive non-SNP variation, with individual loci harboring thousands to tens of thousands of variants, including numerous large structural variants (≥50 bp). Notably, the G03 locus, which was previously shown to have a large effect size of ~0.3 on beak size, contains more than 60 large structural variants alone. These patterns reveal extensive standing structural variation at key QTLs and suggest that the genetic architecture underlying quantitative traits in this system is far more structurally complex than previously appreciated. Ongoing analyses leverage population-scale data from over 4,000 individuals to estimate allele frequencies and assess the evolutionary and phenotypic relevance of pangenome-resolved variants at these loci.

103W Genome-wide linkage disequilibrium estimation reveals extensive correlations among unlinked sites

Anthony Greenberg[1], Diane Wang[2] [1]Bayesic Research, [2]Agronomy, Purdue University

Levels of linkage disequilibrium (LD) among unlinked sites reflect the interplay of population genetic forces shaping the history of a species. For example, theory suggests that LD is more sensitive to population stratification than sequence divergence at individual sites. However, enumerating pairs of unlinked correlated loci normally requires the examination of all pairwise relationships in a genome. This task becomes computationally prohibitive with modern data sets comprising millions of polymorphisms. We have developed an approximate hash table based method to identify groups of loci in LD across the whole genome that does not require computing associations between uncorrelated pairs. We used this method to study LD among unlinked sites in three major domesticated rice populations: indica, aus, and tropical japonica. These populations vary in the extent of local linkage disequilibrium, with indica the lowest and tropical japonica the highest. We applied our method to accessions from the three populations that were sequenced as part of the 3K genome genome project. We recapitulate the previously-described ranking of these populations by overall levels of LD. We also find more linkage disequilibrium among unlinked sites than expected by chance. We further attempt to tease apart the influences of selection and demography by identifying highly correlated locus pairs that are common among populations. Linkage disequilibrium among unlinked sites can make interpretation of genome-wide association signals more difficult.

104W Predicting Theoretical Linkage Disequilibrium for Arbitrary Genome Sizes

Daniel Liu University of California, Berkeley

Linkage disequilibrium (LD) reflects the extent to which recombination constrains the transmission of genetic information along chromosomes and plays a central role in population genetics and genome-wide association studies. Existing LD measures are defined empirically from genotype or haplotype frequencies and therefore depend on population data and evolutionary history. Here, we develop a first-principles theoretical framework that predicts intrinsic LD structure directly from meiotic crossover dynamics and genomic architecture, without requiring genetic variation data. Modeling crossovers as a spatial point process along chromosomes, we derive the distribution of linkage disequilibrium block lengths and show how its moments depend on the underlying recombination landscape. From this distribution, we introduce a dimensionless LD dispersion statistic that quantifies deviations from Poisson recombination and captures the effects of heterogeneous or constrained crossover placement. The statistic provides a natural baseline expectation for linkage determined solely by recombination, enabling principled comparison with empirically inferred LD patterns. Through analytical results and simulations, we demonstrate how recombination hotspots, deserts, and crossover interference shape LD variability across genomes of arbitrary size. This framework establishes a mechanistic reference for interpreting observed linkage disequilibrium and offers a unified way to compare recombination-driven linkage structure across species, chromosomes, and evolutionary contexts.

105W GhostBuster: A Multi-Test Framework for Detecting Ghost Introgression

Margaret Wanjiku, Margaret Wanjiku, Arun Sethuraman Biology, San Diego State University

Gene flow from extinct or unsampled ghost populations is being continually discovered across species, but continues to be challenging without genomes from donor populations. Signatures of ghost introgression can also mimic other demographic events, like bottlenecks, population structure, or migration among sampled genomes. We introduce GhostBuster, a multi-step framework that combines (1) population structure inference under an admixture model, (2) analyses of coalescent time distributions across loci, and (3) goodness-of-fit and likelihood-based tests under the isolation-with-migration (IM) model to detect ghost introgression from population genomic data. GhostBuster utilizes agreement across multiple independent tests to capture different effects of ghost introgression: hidden ancestry components, variation in genealogical histories across the genome, and better fit of demographic models that include an unsampled source. Using extensive simulations under the IM model with varying ghost migration rates and divergence times, GhostBuster analyses show that ghost introgression leaves clear genome-wide patterns of modality in TMRCA distributions. In particular, TMRCAs often stop looking like they come from a single history to indicate multimodality and step-like patterns in the ECDF, even in cases where median coalescent times are similar across scenarios. Likelihood comparisons under IM models unanimously support the presence of an unsampled lineage, but estimating ghost-mediated migration rates itself is harder and depends on the strength of admixture and divergence times. We also apply GhostBuster to human genomic data comparing Central Europeans (CEU) and Han Chinese (CHS) from the 1000 Genomes Project, where TMRCAs show strong heterogeneity across tiles (62,748 TMRCAs; median 42,471 generations; median KS D=0.271; Dip test rejects unimodality) and likelihood-based clustering favors K=2 subpopulations, with weaker support for K=3. IMa3 model comparison also favors a three-population IM model including an unsampled lineage over a two-population model for CEU–CHS and allowing gene flow improves fit relative to models that constrain ghost migration to zero; estimated ghost-mediated migration rates are low but non-zero under the best-supported model. Overall, these results support using GhostBuster as a practical first screening step to flag hidden ancestry and reduce avoidable errors before doing full demographic inference, especially in non-model systems. GhostBuster is available as an open-source Snakemake pipeline at https://github.com/Megmugure/GhostBuster.

106W Panmap: Scalable alignment, genotyping, and phylogenetic placement on pangenomes at sample and read resolution

Alan Zhang[1], Alex Kramer[1], Bianca De Sanctis[2], Sumit Walia[3], Yatish Turakhia[3], Russell Corbett-Detig[1] [1]Biomolecular Engineering and Bioinformatics, University of California, Santa Cruz, [2]Genetics, University of Cambridge, [3]University of California, San Diego

Graph-based pangenome methods face scalability challenges, requiring extensive memory and time for read alignment and variant calling, particularly for large microbial pangenomes containing thousands or millions of genomes. We developed Panmap, a tool that leverages phylogenetic structure to enable ultrafast k-mer-based phylogenetic placement at both sample and read resolution. Panmap uses reference pangenomes encoded as mutation-annotated phylopangenomic trees (PanMAN format) and generates a phylogenetically compressed k-mer index that stores only the differences between sequences along each branch, achieving up to 500-fold smaller size than alternatives and building in a fraction of the time. In single-sample mode, Panmap uses k-mer-based placement to identify the closest reference to the sample, then performs alignment using already-indexed k-mers and generates a consensus assembly. In metagenomic mode, Panmap scores each read against all nodes to place each read to its closest reference, identify likely haplotypes present, and estimate their relative proportions in the sample. We demonstrate Panmap's effectiveness for genome assembly of various viral species, such as RSV and SARS-CoV-2. We further demonstrate accurate abundance estimation of viral species in simulated data and accurate lineage abundance estimation in real wastewater data, rivaling and in some cases outperforming state-of-the-art alternatives. Finally, we built a vertebrate mitochondrial PanMAN and demonstrate fast and accurate phylogenetic placement of ancient eDNA onto the mitochondrial tree, using a competitive-mapping filtering strategy. Application to previously published data revealed clear assignment hotspots concordant with previous results, with reads clustering at their expected species nodes. Panmap provides a scalable and accurate framework for phylogenetically-anchored analysis across genomic, metagenomic, and paleogenomic applications.

107W Balancing and Positive Selection Shapes the Spot Patterning Phenotype in the Masai Giraffe

Joy A Love, Zachary A Szpiech, Douglas R. Cavener Biology, Penn State University

Masai giraffes have intricate spot patterning with highly variable shape and size. Spot patterning is heritable, and variation in spot patterning influences survivorship in both calves and adults. This suggests that natural selection may influence spot pattern variation in the Masai giraffes. Here, we sampled and sequenced 23 Masai giraffes from the Serengeti National Park in Tanzania, Africa, and 12 captive Masai giraffes to study how selection shapes genetic variation in these animals. After calling genetic variants, we characterized population structure among four wild populations and the captive individuals. We then performed multiple allele-frequency and haplotype-based selection scans to detect signals of both positive selection and balancing selection. These analyses, combined with functional enrichment analysis, identified several candidate genes involved in melanin production that may play key roles in spot patterning. One gene, putatively under positive selection, is essential for the development, survival, and function of melanocytes. Another gene, putatively under balancing selection, is involved in the degradation and trafficking of melanin. These results suggest that giraffe spot patterns are not merely aesthetic traits but are shaped by selection. Furthermore, spot patterning may have important implications for understanding unusual phenotypes and informing conservation strategies.

108W Downstream biases introduced by identity-by-descent inference

Aydin Loid C Karatas, Jazlyn Mooney Quantitative and Computational Biology, University of Southern California

Inference of recent demographic history leverages the breakdown of ancestral haplotypes, such as those reflected in identity-by-descent (IBD) segments. While IBD-based methods are widely applied to study relatedness and population structure, the extent to which IBD calling error propagates to downstream inference is unknown. Here, we address this gap by simulating multiple demographic histories using coalescent simulations and inferring IBD segments with two widely used methods, hap-IBD and RaPID. Leveraging tree-sequence data, we quantify each method's ability to recover true IBD segments, estimate population-level IBD sharing, and infer kinship statistics. Across genetic map choices and filtering schemes, we first find that genetic map misspecification can substantially bias IBD inference. Furthermore, genomic regions with low sequence complexity or poorly resolved recombination structure introduce systematic errors. Second, we find that IBD inference often performs best in admixed populations, supporting the utility of IBD-based approaches for resolving recent, complex demographic histories. Third, method-specific trade-offs are evident. RaPID achieves higher accuracy in recovering IBD segments and population-level IBD sharing statistics when genetic maps are correctly specified and spurious segments are rigorously filtered. hap-IBD is more robust to map misspecification and less stringent filtering. Altogether, our results demonstrate that IBD calling accuracy depends on the interplay of demographic history, genetic map specification, algorithmic design, and genomic context. We highlight considerations for the use of IBD-based methods in relatedness and demographic inference, particularly in non-model systems where genetic maps and genome annotations are limited.

109W Using Ancestry Switches to Infer Demography in Admixed Populations

Shengmiao Huang, Jazlyn Mooney, Shirin Nataneli Quantitative and Computational Biology, University of Southern California

Demographic inference is a central problem in population genetics. Many downstream analyses—such as tests of selection, inference of population structure, and admixture mapping—depend on accurate assumptions about population size history. While there are well-established methods for inferring ancient demographic history, few approaches focus on recent demographic history. Further, those that do often rely on large sample sizes, are computationally expensive, and do not allow for streamlined inference in admixed populations.

In this project, we propose a new approach for inferring recent effective population size (Ne) by using ancestry switch accumulation in admixed genomes. Recent theoretical work has shown that the expected number of ancestry switches can be expressed as a function of recombination rate, admixture timing, ancestry proportions, and effective population size. We incorporate these theoretical expectations into a likelihood-based framework, to infer recent demographic history under simple, parameterized models of time-varying Ne.

We evaluate the behavior of the method using simulations, and assess the time periods through which bottlenecks or population growth can be reliably recovered under realistic demographic scenarios. We then apply the method to African American and Latino populations from the 1000 Genomes Project, which have high-quality data, and benchmark our method’s performance against existing methods based on linkage disequilibrium decay and identity-by-descent patterns. We find that more balanced ancestry proportions, higher recombination rates, and more years since admixture events would result in greater ancestry switch accumulation in admixed genomes, thus increase the information available for inferring recent changes in Ne.

In sum, this project harnesses ancestry switches as a summary statistic for recent demographic history and utilizes them to learn about recent changes in effective population size within an admixed population. Thus expanding our ability to conduct demographic inference in admixed populations.

110W How should we report genetic matches following an investigative genetic genealogy search?

Michael Edge[1], Dan Krane[2], Feriel Ouerghi[1] [1]University of Southern California, [2]Wright State University

Traditional forensic genetics practice in the United States has historically relied on short tandem repeat (STR) markers for DNA profiling, as implemented in the Combined DNA Index System (CODIS). One widely used statistic for summarizing the evidence provided by a match between STR genotypes is the random match probability (RMP), the estimated probability that a random person would produce an STR profile matching a forensic sample. In recent years, a new approach called forensic investigative genetic genealogy (FIGG) has been instrumental in solving hundreds of criminal cases. FIGG differs from standard forensic-genetic practice in that it is based on inferences of genealogical relatedness from genome-wide SNPs rather than matching at STR loci. Currently, suspects identified through FIGG are genotyped at CODIS STR loci, and a summary of the evidence from the match—such as an RMP—is subsequently reported in court. However, because FIGG conditions suspect identification on shared ancestry, it enriches the suspect pool for genealogical relatives of the crime-scene profile and for individuals from more narrowly defined subpopulations. Using mathematical models and population-genetic simulations, we show that this enrichment leads to excess allele sharing at STR loci and inflates the apparent strength of STR-based match statistics. We further explore potential adjustments to the RMP using the Balding–Nichols model for match probabilities rather than the unconditional genotype probabilities used in current forensic applications.

111W Estimating uncertainty in sibling-based genotype-phenotype association

Xinyi Miao[1], Michael D. Edge[2], Arbel Harpak[1,3] [1]Department of Integrative Biology, University of Texas at Austin, [2]Department of Computational Biology, University of Southern California, [3]Department of Population Health, University of Texas at Austin

Standard genome-wide association studies (GWASs) are vulnerable to confounding factors, including stratification, assortative mating, and dynastic (indirect genetic) effects. Family studies such as sibling-based GWAS (sib-GWAS) mitigate such confounding and are becoming the tool of choice for teasing apart direct genetic effects–causal effects of one's genotype on their phenotype–from other factors. However, due to their smaller sample sizes, there is substantially larger uncertainty for the sib-GWAS allelic effect. The quantification of this uncertainty is essential for many uses of sib-GWAS, including polygenic scoring, causal inference (e.g., Mendelian randomization), disentangling direct from indirect familial effects, or measuring assortative mating. Here, we use a combination of theory, simulations, and data analysis to understand sources of uncertainty in sib-GWAS allelic effect estimators and biases of three uncertainty measurement methods, including two that are commonly used and one new, resampling-based approach we propose. Our analysis reveals that heterogeneity in allelic effects or heteroskedasticity across families (e.g., due to variation in genetic backgrounds or environments) can bias existing methods, and that this effect is more severe for small samples and rare variants. In contrast, the resampling-based approach we propose is asymptotically unbiased under all scenarios we considered. We validate our theoretical predictions, as well as the importance of effect heterogeneity and heteroskedasticity, using simulations and empirical analysis in the UK Biobank. In sum, our results help understand the sources of uncertainty in family-based genotype-phenotype association and provide a new, accurate, and robust way to estimate empirical uncertainty.

112W Sex differences in crossover interference in house mice

Andrew P Morgan Duke University Hospital

Meiotic recombination ensures the fidelity of chromosome segregation in most organisms with sexual reproduction. The distribution of crossovers along chromosomes is governed in part by interference, which prevents multiple crossovers from occurring in close proximity, though not all crossovers are subject to interference. Neither the factors that control the strength of interference, nor the extent to which they vary within and between species, are well understood. Here, I confirm that crossover interference is stronger in male than in female meiosis in house mice (*Mus musculus*), provide the first estimate of the proportion of non-interfering crossovers in female mice, and show that this proportion is lower than in males. Interference is stronger on shorter chromosomes in both sexes, but the frequency of non-interfering crossovers is similar across the range of chromosome size. Together with evidence that interference varies across strains and subspecies, my results provide a foundation for studying the evolution and sexual dimorphism in this important feature of meiosis in mice.

113W *SaVor* - A Reproducible Structural Variant Calling and Benchmarking Platform from Short-Read Data

Trevor Mugoya, Arun Sethuraman Biology, San Diego State University

Structural variations (SVs) are differences in genomic regions that are larger than 1 kilobase-pair (Kbp) between individuals, and can arise from errant DNA repair mechanisms, whole genome duplications, and transposable element activity across the genome. Recent advances, optimizations, and cost reductions in next generation sequencing technologies have facilitated the exponential increase in the amount of available short read genomic data. Here we present *SaVor*, a flexible, reproducible SV calling workflow that accepts single or multi-lane short-read paired-end Illumina sequence data, or BAM files as input to generate a consensus SV call-set based on user-provided merge parameters. We tested *SaVor* on 1,165 *Arabidopsis thaliana* whole genome sequences and benchmarked its performance on a set of SVs derived from the same accessions using *Lumpy*. Intersection calls i.e. SVs supported by 3 SV callers showed the highest precision (>0.91) while union calls supported by at least 1 caller showed the highest recall (>0.88). We found that the former suffers from decreased recall (<0.51) and the latter decreased precision (<0.57). Depending on the merge strategy, trade-offs in recall and precision need to be considered for downstream analyses of SV call-sets from short-read data. *SaVor* is a Snakemake pipeline and is available on GitHub at https://github.com/ChabbyTMD/SaVor

114W **Mutation Rate Heterogeneity Shapes Population-Level Genetic Variation**

Mariele E Lensink, Grey Monroe, Daniel Kliebenstein Plant Sciences, University of California Davis

Mutation rates vary across genomes due to heterogeneity in DNA damage, chromatin context, and repair activity, with important consequences for genome evolution. Across many taxa, mutation rates are lower in genic regions than in intergenic sequences. In Arabidopsis thaliana, experimental work has linked this pattern to spatial biases in DNA repair pathways. These findings suggest that mutation rate heterogeneity itself may shape genomic patterns of variation and may be subject to evolutionary forces. However, in population-genomic data, the effects of heterogeneous mutation rates are often difficult to distinguish from those of natural selection and demography.

Here, we develop a population-genetic modeling framework to investigate the consequences of intragenomic mutation rate heterogeneity using forward evolutionary simulations in SLiM. We explicitly model genic and intergenic regions with distinct mutation rates and selection regimes, across multiple demographies. Using approximate Bayesian computation, we jointly evaluate the contributions of mutation rate variation and selection to patterns of genetic variation comparable to those observed in the Arabidopsis 1,001 Genomes dataset.

Through this inference framework, we recover genic and intergenic mutation-rate estimates that recapitulate empirically measured rates from Arabidopsis thaliana mutation accumulation experiments. These results demonstrate that mutation rate heterogeneity is a major contributor to patterns of genomic variation, rather than a secondary effect that can be absorbed into selection or demographic parameters. Our findings highlight the importance of explicitly modeling heterogeneous mutation processes in population-genetic analyses, as mutation rate variation can meaningfully influence inferred evolutionary patterns.

Placed in the context of mutation accumulation experiments and experimentally characterized DNA repair mechanisms in plants, these results frame mutation rate heterogeneity as a central component of genome evolution and motivate integrative approaches that connect molecular mechanisms of mutation and repair with population-genetic patterns of diversity across time and space.

115W Inferring the Breeding Sex Ratio by Comparing Variation on the X Chromosome and the Autosomes

William J Spurley[1], Ryan Gutenkunst[2], Bret Payseur[1] [1]Genetics, University of Wisconsin-Madison, [2]University of Arizona

In many populations, unequal numbers of females and males reproduce each generation. This imbalance in the breeding sex ratio (BSR) shapes patterns of genetic variation on the sex chromosomes and the autosomes in distinct ways that are influenced by demographic history. In population genetic analyses, the BSR is routinely assumed to be 1:1, which mischaracterizes the relationship of the X chromosome and the autosomes in natural populations that depart from this assumption. To address this challenge, we introduce two methods to jointly estimate the BSR and demographic history by comparing variation on the X chromosome and the autosomes. Our first method focuses on the site frequency spectrum of polymorphisms, expanding the repertoire of the popular Diffusion Approximations for Demographic Inference (dadi) framework. Our second method leverages approximate Bayesian computation (ABC), combining levels of variation, the site frequency spectrum, and linkage disequilibrium. Each method draws inferences from observed patterns of X-linked and autosomal variation simultaneously to return estimates of the BSR along with population size change. Evaluation of method performance via coalescent simulations demonstrates high power to detect departures from a 1:1 BSR in the context of non-equilibrium demographic history. We examine determinants of method performance, including assumptions about sex-biased mutation rates. Our findings support the feasibility of estimating the BSR across dynamic populations with a range of mating behaviors.

116W A rarefaction approach for identifying local introgression

Thomas Q Smith[1], Zachary A Szpiech[2] [1]Biology, The Pennsylvania, [2]Biology, The Pennsylvania State University

Patterson's *D* statistic, also known as the *ABBA-BABA* statistic, is widely used to detect the presence of archaic genome-wide introgression between two non-sister taxa. Requiring only a single lineage from each of the taxon where one taxon acts as an outgroup to determine the ancestral allele, Patterson's *D*, counts the imbalance between the number of biallelic sites where either the first and third taxa or the second and third taxa share the derived allele. A discordance between counts suggests the presence of introgression. Patterson's *D* is limited to the detection of genome-wide introgression and exhibits a high false-positive rate when applied to smaller genomic segments. Here, we present a new method, D STatistic with Allelic Rarefaction (*D**), to address these limitations. *D** calculates the imbalance between the number of alleles found only in the second and third taxa and the number of alleles found only in the first and third taxa within non-overlapping genomic segments. *D** employs a rarefaction technique for sample size correction, is not restricted to biallelic sites, and does not require an outgroup. We use simulations to show *D** has better precision and recall when compared to similar methods (D and D^+) under a wide variety of model parameters. We also test *D**'s performance in the presence of technical artifacts common to ancient DNA analyses. In addition, we show *D**'s ability to recover possibly introgressed ancient DNA in modern day humans.

117W Extracting quantitative genetic parameters from neural networks using equivalent linear mappings

George Sandler, Ryan York, James Golden Arcadia Science

Deep learning models capture complex non-linear relationships in genomic data but sacrifice interpretability for prediction accuracy. Mechanistic interpretability approaches can bridge this gap by extracting meaningful features from otherwise black box models. Here we describe one such approach referred to as Equivalent Linear Mappings (ELM) which allows us to decompose neural networks into equivalent systems of linear equations. Using theory and simulations we first show how ELM enables us to extract familiar quantitative genetic parameters such as additive and pairwise epistatic effects from neural networks trained on genomic prediction tasks. We then apply ELM to models trained on an existing yeast QTL mapping dataset, demonstrating equivalence to linear regression parameters. Finally, we demonstrate how the rich, sample-specific information provided by ELM enables insights beyond traditional population-averaged approaches. Per-individual feature importance values reveal systematic fitness-dependent measurement artifacts that the neural network implicitly learns to compensate for, illustrating a novel application of ELM for data quality assessment.

118W Rethinking admixture mapping in terms of marginal coalescent trees

Junjian J. Liu, Michael D. Edge Quantitative and computational biology, University of Southern California

Admixture mapping uses local ancestry to identify trait-associated genetic regions. Admixture mapping was adopted enthusiastically in the early 2000s because it required fewer markers than GWAS given the large extent of admixture linkage disequilibrium in recently admixed populations. There has been speculation that admixture mapping is poised for a return as multi-ancestry biobank resources become more common. However, admixture mapping requires assignment of haplotypes to discrete source populations. This conception of discrete populations has been argued to be a poor representation of much human population structure. Here, we re-cast admixture mapping in a continuous form by framing it in terms of variation in local relatedness. This conception does not require postulated discrete admixture groups. We show that a recent method for performing trait mapping using estimated marginal coalescent trees captures the local-ancestry associations that power admixture-mapping signals while also capturing loci that are more easily detected by GWAS than by admixture mapping. Thus, tree-based locus mapping has the potential to be a powerful overarching framework in the era of multi-ancestry studies of complex traits, one that captures the benefits of both admixture mapping and GWAS without requiring discrete ancestry groupings.

119W Supervised machine learning jointly predicts geographic location and climate of origin from genotypes

Jordan Rodriguez[1], Andrew Kern[2], Richard Cronn[3] [1]Biology, University of Oregon, [2]University of Oregon, [3]United States Forest Servic

Spatial genetic variation reflects the combined effects of dispersal, drift, and spatially heterogeneous selection, encoding information about geographic location and the environments to which individuals are best adapted. As climates shift rapidly, a central challenge is identifying environments in which existing genotypes are most likely to persist. Building on our prior work predicting geographic location from genomic data, we introduce **Ecolocator**, a supervised deep neural network that uses multi-task learning to jointly predict location and climate-of-origin (historical climate at sampling locations) from genome-wide variation. Trained on georeferenced samples paired with historical climate variables, Ecolocator integrates information across all variants simultaneously, enabling it to capture polygenic signals of local adaptation while accounting for neutral spatial structure. Using forward-in-time simulations across heterogeneous landscapes, we characterize the evolutionary scenarios under which Ecolocator performs best. We find that predictive accuracy increases with the strength of local adaptation; strong stabilizing selection and limited dispersal generate clear environmental signals, allowing Ecolocator to recover climate-of-origin even in the absence of spatial autocorrelation. Increased gene flow reduces geographic predictability, while reduced selection weakens genotype–environment associations without degrading location prediction. Feature attribution analyses further show that Ecolocator more accurately prioritizes adaptive variants than standard genotype–environment association methods. Lastly, we apply Ecolocator empirically to coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*), a globally important timber species with strong local adaptation, and recover information about both location and climate-of-origin across its range. Together, these results establish Ecolocator as a promising method for linking genomic variation to environmental suitability while separating adaptive and neutral spatial signal.

120W Quantifying Factors That Influence Structural Variant Detection and Characterization

Diana Tataru[1], Patrik Nosil[2], Kira Vasquez-Kapit[3], Zachariah Gompert[1] [1]Biology, Utah State University, [2]Station d'Écologie Théorique et Expérimentale, CNRS, [3]Colorado State University

Structural variants (SVs) can have large effects on evolutionary processes and genomic diversity by creating jointly inherited regions through recombination suppression and by altering gene expression and the 3D organization of DNA. Recent improvements in sequencing technologies have led to more abundant and higher resolution genomic data, and subsequently increased research on SVs. However, researchers must balance trade-offs between scale, resolution, and computational intensity when detecting and characterizing SVs, and comprehensive evaluations of tradeoffs remain limited. Here, we inform these decisions by using new pangenomic approaches to evaluate SV detection and characterization in (i) *de novo* genome assemblies of 8 phased genomes, and (ii) population level genotype-by-sequencing (GBS) data of 216 *Timema cristinae* stick insects. We quantify the effects of data types, analytical approach, and SV type, size and frequency on our ability to detect and characterize SVs, with a focus on inversions and translocations. We emphasize these SV types because they more substantially reorganize chromosome structure and suppress recombination than other types. Additionally, existing large effect inversions and translocations have been previously identified in the *T. cristinae* system, and preliminary comparisons of the phased genomes and population-level GBS data suggest substantial differences in detection for these SV types. We evaluate comparative genomic approaches by constructing a pangenome from the alignment of our 8 *de novo* genome assemblies and calling SVs directly from it, as well as population genomic approaches that rely on patterns of local population structure along chromosomes inferred from GBS data aligned to the pangenome. Our results suggest that different approaches recover partially distinct subsets of SVs and differ in their capacity to characterize SVs, especially complex SVs, which population genomic methods capture less effectively. While this work focuses on one biological system, it serves as a model for comparable studies in a burgeoning research field.

121W Leveraging Factor Analytic Linear Mixed Models for Prediction of Genotype-by-Environment Dynamics in Large-Scale Wheat Breeding Populations

Mitchell Eglinton[1], Lee Hickey[1], Ben Hayes[1], Eric Dinglasan[1], Karine Chenu[1], Bertus Jacobs[2], Alison Kelly[1] [1]Centre for Crop Science, Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, [2]LongReach Plant Breeders

The integration of historical phenotypic data into contemporary genomic selection pipelines presents a significant computational and statistical challenge, particularly in the presence of intense selection and unbalanced experimental designs. We assembled a large, unbalanced dataset from a national bread wheat (*Triticum aestivum* L.) breeding program in Australia, spanning ten years (2016–2025) and implemented a factor analytic (FA) linear mixed model framework for the analysis of multi-environment selection trials. This research leverages the historic yield performance of breeding populations to serve as a high-density training set for the genomic prediction of parental breeding values and the partitioning of environment-dependent combining ability.

Central to this study is the meta-analysis of a national network of early-generation yield trials. Our approach effectively models complex genotype-by-environment (G×E) interactions by integrating high-density parental genomic data with progeny pedigree relationship matrices. To characterize the environmental drivers of these interactions, we utilize latent factors derived from the FA model in conjunction with explicit meteorological covariates. This multi-layered integration establishes genetic linkage across disparate years and regions, significantly refining the estimation of additive genetic effects and improving the predictive accuracy for cross-performance in unobserved environments.

The statistical architecture specifically addresses the challenges inherent to early-generation breeding data, characterized by partial replication, high selection intensity, and limited connectivity between annual cycles. The model addresses the Bulmer effect through the modelling of heterogeneous genetic variances, providing a framework that quantifies the iterative depletion of genetic diversity over selection stages.

Preliminary results demonstrate the model's utility in identifying ancestral lineages that drive G×E sensitivity. Our findings suggest a complex relationship between intra-population genetic variance and environmental plasticity, where specific ancestral lineages, parents and genomic regions appear to govern the trade-off between yield stability and environmental plasticity in their progeny. These results provide a scalable statistical framework for utilizing legacy breeding data to optimize future cross prediction and selection strategies.

122W One Nation under a Groove: Ensemble Learning for Demographic Inference

Ananya Kapoor, Nathaniel S Pope, Andrew D Kern Biology: Ecology and Evolution, University of Oregon

Inferring demographic models from population genetic data is challenging due to parameter identifiability, local optima, and model misspecification, particularly in the presence of background selection (BGS). Model-based inference typically fits a small set of summary statistics to genome-wide expectations and is therefore vulnerable to method-specific biases and reduced robustness when genomic processes deviate from model assumptions. One general strategy for improving robustness in the presence of method-specific biases is ensemble learning, which combines multiple predictors to reduce variance and bias. To address these limitations, we present a framework that integrates multiple demographic inference methods to improve parameter estimation accuracy and robustness. Treating inference outputs as noisy observations, we train a suite of linear and nonlinear models to learn systematic biases using simulations generated under both neutral and background-selection regimes. We evaluate performance across a range of demographic models, including symmetric and asymmetric isolation-with-migration models, bottleneck models, and a *Drosophila* three-epoch model. Across simulations, ensemble predictions consistently outperform individual inference methods. Finally, we demonstrate the utility of the approach with an application to human genomic data from the 1000 Genomes Out-of-Africa populations.

123W Information gain as a framework for evolutionary data collection

Ryan York, Prachee Avasthi Arcadia Science

Biological data have accumulated at an unprecedented scale over the past 25 years, enabling the development of machine learning (ML) models trained directly on evolutionary variation. Public datasets, however, have grown opportunistically rather than prospectively. Source bias, taxonomic imbalance, and pseudoreplication are often present. Recent work has shown that biological ML models can exploit these features to memorize rather than generalize, raising questions about how much biological information they genuinely learn and how additional data should be prioritized. We argue that the next era of biology requires a more prospective, evolutionarily-informed approach. We present a Bayesian framework that treats learning as an iterative process and optimizes for an entropy-based measure of information gained (IG) from new observations. Applying this framework to the AlphaFold database, we quantify the IG contributed by individual proteomes and demonstrate how these values can guide data acquisition across biological contexts. We find that IG is highly context-dependent: efforts to maximize overall novelty benefit most from sampling eukaryotic proteomes, particularly plants, whereas prokaryotic taxa provide greater efficiency per protein sampled. Accounting for this context dependence, we show how to generate evolutionarily informed priors and estimate task-conditional lower bounds on the data required for desired objectives. Future progress will likely depend less on indiscriminate data accumulation and more on principled strategies for sampling evolutionary diversity. An inference-driven and evolutionarily grounded approach to data collection offers a path toward more efficient learning of general biological principles.

124W Model-based inference of regional African contributions in African-American genealogies using Transatlantic Slave Trade voyage records

Kennedy Agwamba Computational Biology, Stanford University

The Transatlantic Slave Trade was transformative in shaping the genetic landscape of modern North America. Today, millions of people in the United States who identify as Black or African-American descend from forcibly transported captives arriving from multiple coastal regions of Africa. However, the systematic erasure of genealogical records under slavery has obscured links between modern descendants and their African ancestry. Here, we integrate data from the Transatlantic Slave Trade Database with a mechanistic model of African-American demographic history to estimate the regional composition of transported African genealogical ancestors. Using time-resolved records of embarkation from eight major African coastal regions, we model how ancestry is transmitted across generations to quantify the proportional contributions of each region to the pedigrees of mid-20th century African-American descendants. Our results indicate that ancestry reflects substantial contributions from multiple regions, with particularly large fractions tracing to the Bight of Biafra, Senegambia, and West Central Africa. Notably, the relative genealogical impact of the various regions does not strictly mirror the total numbers of captives transported. Differences in arrival timing and subsequent population growth substantially shape long-term ancestry patterns, amplifying some regions' genealogical contributions relative to others. These findings provide quantitative insight into the structured African ancestry embedded within African-American family trees and demonstrate how integrating historical records with mechanistic admixture models can recover genealogical patterns. More broadly, this framework offers a generalizable approach for reconstructing structured ancestral contributions in admixed populations.

125W A neural model of linkage disequilibrium decay for recent effective population size inference

Chris C Smith Indiana University

A bioinformatics tool is presented for estimating recent effective population size that uses a neural network to automatically compute linkage disequilibrium-related features conditioned on genomic distance between polymorphisms. The method is evaluated using simulations and demonstrated on empirical data from harbor porpoises. The new method outperforms existing deep learning and summary statistic-based approaches using relatively few sequenced individuals and variant sites, making it particularly valuable for molecular ecology applications with sparse, unphased data.

126T Correcting for hematopoietic variation in whole blood sequencing isolates heritable aging signals in multigenerational pedigrees

Alexis C Garretson[1], Goldberg E Michael[2], Camila Goclowski[2], Thomas A Sasani[2], Hannah C Happ[2], Julia Ostrander[2], Isabel Serrano[2], Richard Cawthon[2], Lynn B Jorde[2], Deborah W Neklason[2], Aaron R Quinlan[2] [1]Genetics, University of Utah, [2]Human Genetics, University of Utah

Many genomic and molecular traits can be estimated via whole-genome sequencing vary predictably with age and are linked to disease. Examples of such somatic traits include telomere lengths, mutation burden, larger mosaic chromosomal alterations, as well as non-nuclear traits like mitochondrial copy number (mtCN) and heteroplasmy. Within blood cells, there are also sex differences in these traits, not limited to but perhaps most dramatically in the mosaic loss of X (mLOX) and Y (mLOY), a common event amongst aging males. Studies using large biobanks have shown many of these traits to be heritable. However, they are also strongly linked to blood composition, which is also heritable. Therefore, accounting for the impacts of blood cell fractions may change the estimated heritability and associated genetic signals, but is not routinely considered in models.

The large, multigenerational CEPH/Utah pedigrees are an ideal framework to understand the covariation, heritability, and signatures of aging because of their three and four generation structure, and the wealth of available health, laboratory, and sequencing data under relatively homogenous environments. Using a Bayesian approach, we control for blood composition while estimating heritability and covariance. Initial estimates suggest significant heritability of blood traits amongst our population in both males and females (0.1 – 0.9) and uniformly higher heritability for the blood traits in males. Correcting for blood composition reduces the heritability of telomere and mtCN by approximately half; for example, telomere length heritability drops from 0.76 to 0.43 and mtCN from 0.98 to 0.5. We also find genetic covariation between mtCN and both telomere length (0.33) and mLOY (-0.44) in males, but only between mtCN and mLOX in females (-0.42).

These results underscore that both cell composition and sex are core determinants of how aging-related genomic traits are measured and interpreted. Moreover, this approach provides a framework for disentangling aging signals from hematopoietic variation in examining genetic architecture. Future extensions involve estimating heritability across specific age brackets to determine how the genetic architecture of these traits varies across the lifespan. Importantly, this framework is extensible to traits in other tissues that may vary with cell type composition, provided that cell composition is either inferrable or directly measurable alongside focal traits of interest.

127T A Comparison of Phylogenetic Genotype to Phenotype Models

Adam W. Warlen, Jared E. Decker University of Missouri

Genotype-to-phenotype mapping is essential in multiple contexts, including medical relevance for human health, economic and food security incentives for agriculture, and understanding evolution of traits. Methods exist to investigate relationships between loci and traits through convergent evolution, where changes to loci across a phylogeny are correlated with trait variation. To expand this area of research, we used the Zoonomia primate expanded alignment alongside published mammalian phenotype data for 447 species to compare three models of associating loci with traits. The models we fit were a Pearson's linear correlation implemented with RERconverge, and two Linear Mixed Models (LMMs), those being the terminal model and total model. We began by calculating loci-specific phylogenetic trees with IQ-TREE, where branch lengths were computed based on nucleotide substitutions with the GTR model. We then used the RERconverge package to convert branch lengths of the loci-specific trees to Relative Evolutionary Rates (RERs), a measure of observed evolutionary rate for a loci relative to expected evolutionary rate. With the RER trees, we fit the three models, starting with the built-in Pearson's linear correlation in the RERconverge package for continuous traits. To test if additional model complexity would improve predictions, we fit the LMMs. The terminal model associates terminal branch lengths with traits while adjusting for species relatedness using a phylogenetic relationship matrix. The total model fits two phylogenetic relationship matrices, one for the species phylogeny and a second for the locus-specific phylogenetic relationship matrix. We compared the calibration of the three methods with visualizations and genomic inflation factors, and explored loci associated with various phenotypes.

128T Tandem Repeats and Adaptive Archaic Introgression in HPRC Individuals

Melody Chang, Alicia Bierly, Hannah Siegel, Mary Crowell, Carolina Adam, Rori Rohlfs University of Oregon

Tandem repeats (TRs) are repeating units of DNA of variable length that undergo frequent mutations, making them important sources of genetic variation. TRs have a range of implications, being linked to gene expression regulation and certain diseases. However, the evolutionary impact of TRs, specifically their role in adaptive variation in humans, is underexplored. One study on the MUC19 gene highlights how archaic introgression may introduce adaptive TR alleles. Individuals were found to have a Denisovan-like haplotype with a large expansion in a coding TR compared to the non-archaic haplotype. The results of the MUC19 study led us to investigate how TR variation may drive adaptive archaic introgression. The availability of long-read sequenced population data and high quality identification of introgressed tracts enables us to explore the joint relationship of TRs and archaic tracts in human adaptation. We used hmmix to identify archaic and non-archaic tracts for 20 individuals in the Human Pangenome Reference Consortium. The Tandem Repeat Genotyping Tool was used to genotype TRs in the same individuals. Then, we merged the data to find which TR alleles are on introgressed tracts. We are analyzing how divergence in TR lengths between introgressed and non-introgressed alleles may indicate adaptive introgression. Our results will contribute to our overall understanding of the role of TR variation in human evolution.

129T The effect of missing data on ARG inference in an empirical system

Nicole Adams[1], Alexander L Lewanski[1,2,3], Benjamin Carlson[1], Gideon S Bradburd[1] [1]Ecology and Evolutionary Biology, University of Michigan, [2]Integrative Biology, Michigan State University, [3]W.K. Kellogg Biological Station, Michigan State University

The concept of the ancestral recombination graph (ARG) was developed in the mid-to-late 20th century. However, with methodological, computational, and genomic sequencing advances in the past decade, it has only recently become feasible to infer ARGs in an empirical context. While ARG inference is becoming increasingly frequent in human studies where there are extensive and high-quality genomic resources, its application in non-model systems remains a significant challenge and is therefore still rare. Non-model studies are often hindered by a lack of fine-scale recombination maps, low sequence coverage, and high proportions of missing data, a factor known to bias traditional population genetic metrics. Here, we use a dataset of wild house mice (*Mus musculus*) to examine how increasing missingness affects ARG inference. We systematically filter high-quality genomic data to create a gradient of missingness. We then impute and phase each dataset with BEAGLE and infer the ARG using SINGER. With the set of inferred ARGs, we calculate a variety of metrics (e.g., tree height and number of nodes and edges) and genetic diversity estimates (i.e., pi, Fst, Tajima's D) to explore how increasing missingness affects the inferred ARG and downstream population genetic quantities. Understanding the effect of missing data on ARG inference is vital for developing best practices and identifying limitations when applying ARG-based approaches in non-model systems.

130T Cross-population replication analysis of STR-trait associations in the All of Us dataset

Nichole Ma Owen University of California San Diego, Nichole Ma Owen1, Arya Massarat2,Jonathan Margoliash3 ,Tara Mirmira3, Sara Javadzadeh3, Wilfredo G. Gonzalez Rivera2,4, Helyaneh Ziaei Jam3, Yang Li1,

Kelly A Frazer5,

Matteo D'Antonio4, Melissa Gymrek1,3,5 1Department of Medicine, University of California San Diego, La Jolla, CA, 92093 2Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, 92093 3Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, 92093 4Department of Biomedical Informatics, University of California San Diego, La Jolla, CA, 92093 5Department of Pediatrics, University of California San Diego, La Jolla, CA, 92093 Recent studies have demonstrated a widespread role of short tandem repeats (STRs) in complex traits, but these have largely been limited to individuals of European descent. The All of Us (AoU) cohort enables detailed study of genetic influences across diverse populations. To evaluate feasibility, we performed a replication analysis of 119 STR–blood trait association signals identified in the UK Biobank (UKB).

We developed reproducible WDL pipelines for phenotype preprocessing, STR genotyping, and association testing, deployable across cloud platforms. We applied these pipelines to genotype 59 unique STRs and tested associations with 19 phenotypes with at least 24,000 samples available in AoU, yielding 71 overlapping STR–trait pairs. Overall, 93% of tested STR–trait pairs showed concordant directions of effect between European AoU and UKB individuals, with strong correlation in −log10 p values (Pearson r = 0.87, p = 1.21e−22). Correlation decreased but remained strong in individuals of African ancestry (r = 0.61, p = 4.95e−08). We replicated several strong UKB associations, including a CGG repeat in the promoter of *CBL* associated with platelet count (p = 5.12e−19; n = 92,099).

We also evaluated STR imputation in AoU SNP data using a published SNP–STR reference panel containing 1.09 million genome-wide repeats. STRs were imputed on Chromosome 11, and at the *CBL* repeat we achieved 79.8% concordance between imputed and HipSTR-genotyped repeat lengths, comparable to the 71% concordance observed at this locus in UKB.

Finally, we compared meta-analysis and mega-analysis GWAS strategies for high-density lipoprotein (HDL) traits. Ancestry-specific GWAS were conducted in European (EUR, n = 55,622), African (AFR, n = 14,350), and admixed American (AMR, n = 11,163) cohorts and combined using METAL under a fixed-effects inverse-variance model. Meta-analysis results were concordant with mega-analysis (n = 90,885; Pearson r = 0.91 for signed −log10 p values), but limited AFR and AMR sample sizes reduced power, favoring mega-analysis for signal discovery.

Overall, these results demonstrate the feasibility of STR association testing in AoU and replication of UKB findings in a diverse population.

131T **Population genomics in hyper-polymorphic species: pangenomic approaches using the nematode *Caenorhabditis brenneri***

Anastasia Teterina[1], John H Willis[1], Nathaniel S Pope[1], Angel G Rivera-Colón[1], Peter L Ralph[1,2], Andrew D Kern[1], Patrick C Phillips[1] [1]Institute of Ecology and Evolution, University of Oregon, [2]Data Science, University of Oregon

Genetic variation carries footprints of population history and evolution of the species, yet decoding what has shaped the diversity landscape remains challenging. This task becomes even more complex in species with high levels of genetic variation, including complex and structural variants and a substantial proportion of multi-allelic sites, which are often excluded from standard population genetic pipelines. We simulated various evolutionary scenarios with high levels of polymorphism to explore how reverse mutations, complex variants, and bioinformatic filtering potentially bias standard population genetic approaches. We applied corrected estimators and compared them to variation graph- and k-mer-based diversity measures to assess how each captures patterns of genomic variation in hyper-polymorphic species. We then applied this framework to *Caenorhabditis brenneri*, an outcrossing nematode that is currently known to be one of the most genetically diverse eukaryotes, with nearly one in ten nucleotides being polymorphic (a level of population diversity comparable to many bacteria). We used long-read assemblies from individual nematodes and pangenome graphs to integrate divergent haplotypes, structural variants, multi-allelic sites, and gene content variation into population-level analysis. We analyzed genome-wide diversity and its distribution within genetic elements, comparing how site-based, graph-based, and k-mer-based approaches track population processes in such extreme regimes. This work contributes to a deeper understanding of the population history and diversity of *C. brenneri*, highlights key limitations of current tools, and offers a path forward for studying molecular evolution in hyper-polymorphic species.

132T ARGformer: learning on ancestral recombination graphs with transformers

David Bonet[1,2,3], Cole Shanks[1], Marçal Comajoan Cara[4], Jordi Abante[3], Alexander Ioannidis[1,2] [1]University of California Santa Cruz, [2]Stanford University, [3]Universitat de Barcelona, [4]University of California Berkeley

Recent advances in inference of the ancestral recombination graph (ARG), which describes how segments of chromosomes trace back through recombination and shared lineages, have made it possible to reconstruct genome-wide genealogies for large cohorts, but it remains difficult to summarize and use this information for population genetic analyses. We present ARGformer, an encoder-only transformer that learns context-dependent embeddings with a self-supervised masked-node objective finetuned with contrastive learning for downstream retrieval tasks. We train ARGformer on genealogies from coalescent simulations and on genealogies inferred from ancient and present-day Homo sapiens genomes. Using only these learned embeddings, without access to genotype matrices, ARGformer recovers familiar patterns of global population structure and supports unsupervised ancestry inference through clustering and nearest neighbor retrieval. On genealogies that include archaic hominins, ARGformer can highlight Denisovan-derived segments in Oceanian genomes and reveals Oceanian-like ancestry in South American indigenous populations.

133T DeepNe, A Novel Deep Learning Approach for Inference of Effective Population Size, Provides Precise Reconstruction of Founder Events Across Oceania

Samuel D. Sacco[1], David Bonet-Sole[1], Alexander Mentzer[2], Andres Moreno-Estrada[3], Alexander Ioannidis[1,4] [1]Biomolecular Engineering, UCSC, [2]Center for Human Genetics, University of Oxford, [3]Advanced Genomics Unit, Center for Research and Advanced Studies (Cinvestav), [4]Stanford University

Inferring the past trajectory of a population's effective size (Ne) is a central goal within population genetics. In human populations, reconstructing Ne trajectories from genomic data can be used to uncover demographic patterns in human pre-history while remaining orthogonal to demographic estimates constructed through archaeological or historical approaches. While several methods currently exist to infer the trajectory of effective population size, their inability to accurately time extreme demographic scenarios such as founder events, their reliance on large sample sizes, and their heavy regularization of inferred trajectories limit their applicability and interpretation. Here we present a novel deep learning approach, DeepNe, which after training on thousands of simulated population histories can more accurately infer the trajectory of recent population size under a wide-range of demographic scenarios from small sample cohorts. Using DeepNe to perform an ancestry-specific analysis of population size we provide precise estimates of the timing of human settlement across Oceania, and provide genetic estimates of the sizes of these populations. Our results provide new insights into demographic events in this understudied region, and illustrate the potential of simulation trained deep learning approaches within population genetics.

134T Ancient local ancestry inference with a graph transformer

Cole Shanks[1], David Bonet Sole[2], Marçal Comajoan Cara[3], Alexander Ioannidis[2] [1]University of California Santa Cruz, [2]Biomolecular Engineering, University of California Santa Cruz, [3]Department of Computer Science and Electrical Engineering, University of California Berkeley

Local ancestry inference classifies segments of DNA in admixed individuals by their originating population. However, as the date of admixture becomes older, these segments become shorter and determining their ancestry becomes increasingly difficult. This limits many existing segment-based methods to relatively recent historical admixture events and more highly diverged populations. The rapidly expanding availability of ancient DNA offers a promising opportunity to use these ancient samples as references for local ancestry inference. A recent approach integrates ancient samples into the ancestral recombination graph (ARG) for local ancestry inference. Here, we introduce recent advances in deep learning for graphs into this ARG framework to create ARGMix, a graph transformer that infers local ancestry using the coalescent trees of the inferred ARG. Our approach employs ancient samples as references in the marginal trees to predict local ancestry. We train ARGMix on data reflecting the well-understood ancient European demography and demonstrate improved accuracy and robustness even under demographic misspecification. We then apply ARGMix to an ARG of ancient and present-day European samples for ancestry-specific analyses, finding evidence of continuity between Ötzi the Iceman and present-day individuals from nearby regions.

135T Predicting Puberty and Fertility Traits in Multi-breed Beef Heifer Cattle

Dotun Olaoye[1,2], Garrett Franklin[1,2], Robert Schnabel[1,2], Jordan Thomas[2], Jared Decker[1,2] [1]Genetics Area Program, University of Missouri-Columbia, [2]Animal Science, University of Missouri-Columbia

Puberty and fertility are critical traits with significant impact on profitability in the cattle industry. These traits are complex as they are influenced by several factors and are difficult to predict. However, technological advancement has allowed for the development of powerful tools like genomic prediction – a genetic tool which produces accurate predictions for lowly heritable traits like puberty and fertility in cattle. Genomic prediction produces estimated breeding values (EBV) for selecting genetically superior animals for any desired trait. Unlike in the dairy cattle industry, relatively less attention has been focused on improving puberty and fertility traits in beef cattle using genomic prediction. Existing preliminary data from our research group indicated that there are similar alleles controlling these traits. Moreso, our preliminary predictions showed promising potential for traits like Days Open and Days to Conception in Red Angus. Days Open and Days to Conception are considered novel traits in the U.S. beef industry. Days Open is the days of the breeding season a heifer remained open while Days to Conception is the number of days in the breeding season it took a heifer to get pregnant with missing values assigned to heifers that did not conceive. In this current study, we considered traits such as Days Open, Heifer Pregnancy, Pelvic Height, Pelvic Width and Reproductive Tract Score in a multi-breed (Angus, Red Angus, Hereford, and *Bos indicus* cross) panel of beef cattle. For genetic evaluation and genomic prediction, we aim to leverage a large dataset (13,781 animals, 5 traits, pedigree records and 800k imputed Single Nucleotide Polymorphism [SNP] data) to pursue the following: (1) estimate variance components and correlation between traits, (2) compare prediction accuracies from univariate and multivariate models using the standard single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) method, and (3) compare prediction accuracies from univariate and multivariate models using a SNP prioritization approach. The LR Method will be used in 3-fold cross validation for assessing model accuracies for all the models. We expect low to moderately high heritability across the traits as well as prediction accuracies comparable or better than current heifer pregnancy predictions.

136T Estimation of marginal geneaology using wavelets and graph convolutional neural networks

Dylan Ray Genetics, University of North Carolina at Chapel Hill

Spurred by the increasing availability of population-scale whole-genome datasets, inferring the full genealogical history of a population sample is a major goal for the field of population genetics. The problem is challenging due to the high dimensionality of the target genealogy at each site and the input alignment and by the difficulty in leveraging long-range linkage disequilibrium to properly identify genealogical similarities and differences across the chromosome. We present a method that employs the continuous wavelet transform, a time-frequency analysis tool, over large scales combined with learned graph convolutions applied over the sample at each site. By using pre-chosen filters specified by a wavelet function, we avoid the need to learn sufficiently large kernels. The output of our architecture is an embedding for each chromosome at each site, and the pairwise log distances of the points allows for the accurate reconstruction of the marginal tree sequence using a hierarchical clustering. Our supervised machine learning approach outperforms existing state-of-the art methods for inferring the marginal tree sequence given a population alignment. We demonstrate this through a comparison of the error in inferred coalescence times, topology, and breakpoints for simulated chromosomes across a variety of population genetic scenarios, demographies, and sample sizes.

137T Beyond Glucose: Machine Learning Identifies Genetic Determinants of Mannose Metabolism Across More Than 1,000 Yeast Species

Benjamin Narh-Madey[1], Linda Horianopoulos[2], Antonis Rokas[3], Chris Hittinger[4] [1]Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, J.F Crow Institute for the Study of Evolution, University of Wisconsin-Madison, USA, [2]Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, J.F Crow Institute for the Study of Evolution, DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, USA, Department of Food Science University of Guelph, Guelph, Canada, [3]Department of Biological Sciences and Evolutionary Study Initiative Vanderbilt University, USA, [4]Laboratory of Genetics, Wisconsin Energy Institute, Center for Genomic Science Innovation, J.F Crow Institute for the Study of Evolution, DOE Great Lakes Bioenergy Research Center, University of Wisconsin Madison

Evolution across time and space has long guided our understanding of life's fundamental principles across various taxa. However, the sheer complexity of the genotype-phenotype landscape across macroevolutionary timescales raises questions about the extent and depth of our knowledge. In yeasts, this complexity is partly attributed to their carbon adaptation across species. While much is known about their adaptation to glucose, little is known about their adaptation to mannose, which is a sugar common in fungal cell walls. Mannose, the C2 epimer of glucose, shares transport pathways, core glycolytic intermediates, and downstream metabolic pathways with glucose, yet exhibits distinct regulatory and metabolic properties. Phenotypic profiling revealed significant inter-specific variation in mannose utilization, with many species exhibiting superior growth on mannose relative to glucose, contrary to expectations from canonical model systems. Leveraging the Y1000+ Project dataset, which captures extensive genomic and phenomic diversity across the Saccharomycotina subphylum, we investigated the genetic and evolutionary basis of mannose metabolism of more than 1,000 yeast species spanning more than 400 million years of evolution.To identify the genetic determinants of this variation, we combined machine learning with comparative genomics. Building on our previous use of random forest models to study oxidative stress resistance and galactose metabolism, we applied logistic regression, support vector machines (SVMs), polynomial kernel SVMs, and radial-basis function SVMs to predict mannose growth. Across models, we identified nine gene families strongly associated with mannose utilization, including hexose and siderophore-dependent iron transporters, whose copy number directly influences mannose growth. We identified four genes, *MNN2* and *MNN5*, which encode α-1,2-mannosyltransferases responsible for branching the mannan backbone, as well as *MAN2* and *DSF1*, which promote α-mannosidase activity and metabolic and mitochondrial reprogramming respectively under alternative carbon sources.Together, our findings support the regulatory roles genes play in carbon-specific metabolic adaptations, which culminates in species-specific preferences for alternative carbon sources.

Keywords: Machine learning, mannose metabolism, yeasts genetics, evolution.

138T Larger polygenic effects on health outcomes in groups with lower socioeconomic status

Olivia S Smith[1], Jared M Cole[2], Arbel Harpak[2] [1]Integrative Biology, University of Texas at Austin, [2]University of Texas at Austin

Researchers are increasingly recognizing the importance of gene-by-environment interaction (G×E) on complex traits and health outcomes. Here, we consider one specific mode of G×E – that of amplification, wherein the genetic variants affecting the trait are the same, but their effects vary systemically in magnitude across different contexts. For example, propensity for diabetes has been found to be higher in more deprived populations, even after adjustment for lifestyle factors. Here, we test for amplification of polygenic effects on disease risk and other biomedically-relevant traits across groups of different socioeconomic status (SES). For ancestry-matched samples from the UK Biobank and the All of Us Research Program, we tested whether genetic effects of polygenic scores (PGS) were different in high- versus low-SES groups. Across the majority of 28 biomarkers, diseases and other biomedical traits and using various metrics of SES, we found consistently larger effects of PGS in groups with more deprived socioeconomic status. Two interesting exceptions were triglyceride levels and risk of heart attack, which were inversely correlated with deprivation levels. Overall, our results are consistent with the amplification of genetic effects on disease risk in socioeconomically deprived populations

139T Dating large ARGs, with clinical applications

Nathaniel Pope[1], Samuel Tallman[2], Ben Jeffery[3], Duncan Robertson[3], Savita Karthikeyan[3], Peter L Ralph[4], Jerome Kelleher[3] [1]Institute of Ecology and Evolution, University of Oregon, [2]Genomics England, [3]University of Oxford, [4]Data Science, University of Oregon

Modern methods allow ancestral recombination graphs (ARGs) to be inferred at the biobank scale. Two potentially transformative advantages to using these for analysis are that the ARG (1) situates genetic variation along the time axis; and (2) describes genetic variation without recourse to imprecise notions of discrete populations. To realize these advantages, ARG inference needs to be sufficiently accurate that summaries from inferred ARGs accurately represent the corresponding quantities in the true, unobserved ARG. We have developed a new variational Bayes method for dating inferred ARGs, and applied the method to 95,000 whole genomes from the Genomics England dataset. Accuracy is very high and increases strongly with sample size, as demonstrated by comparing inferred ages to simulation and in real data against aDNA constraints. We also show that inferred ages of ultra-rare mutations can provide evidence against pathogenicity, and so inferred age provides a potential clinical alternative to allele frequency that does not depend on division into discrete populations.

140T Identifying multi-allelic quantitative trait loci using empirical haplotypes

Katelyn A McInerney, Karen L Mohlke, Michael I Love, William Valdar Genetics, University of North Carolina at Chapel Hill

Understanding the genetic basis of complex traits remains a critical challenge in biology and genetics across humans, animals, plants, and agriculture. Despite advances in genome-wide association studies and quantitative trait locus mapping, the pathways from genotype to phenotype remain largely unknown for most complex traits. Current quantitative genetic approaches predominantly rely on SNPs as the input variable for modeling these relationships. Although such SNP-based methods have enabled significant discoveries, they also have limitations. A promising alternative is to use haplotypes, allowing more concise modeling of local multi-SNP combinations without requiring a search for complex, higher-order SNP-SNP interactions. In order to identify useful haplotypes for association mapping, we have developed *IDHaplos,* which uses a hierarchical clustering algorithm on the genotyping matrix to group individuals into allele groups. We use the Hamming distance to construct the distance matrix, and we cut the tree based on a pre-specified dissimilarity threshold. We then model these haplotype alleles in a linear mixed model and compare performance with other haplotype clustering approaches (*BigLD, HaploBlocker, tskit*) and SNP-based approaches in simulated, model organism, and human data.

141T Interrogating the role of gene-environment interactions in polygenic score portability

Roshni A Patel[1], Arun Durvasula[2], Jazlyn Mooney[3] [1]University of Oregon, [2]University of Southern California, Keck School of Medicine, [3]University of Southern California

In response to observations that polygenic scores (PGS) have poor portability across cohorts, a growing body of work is once again focused on identifying “ancestry-specific effects” – that is, genetic variants whose effects are purported to differ across “ancestries”. The concept of “ancestry-specific effects” is both ethically and technically deficient: it recalls essentialist views of human difference and fails to approximate real biological processes. To highlight these shortcomings, we develop generative statistical models of GxE interactions. Through simulations of quantitative complex traits under realistic human demography, we describe how GxE interactions will affect GWAS estimates of variant effects. We further characterize how the genetic architecture of a trait shapes the extent to which GxE interactions decrease PGS portability across cohorts with different environmental exposures. Crucially, our work demonstrates that GxE interactions generate a distribution of variant effects in a population, reiterating that discrete ancestry categories are a poor proxy for real biology.

142T Scalable Local Ancestry Inference Using Genotype Representation Graphs

Gordon Mei, Ziqing Pan, Mariia Tiulchenko, April (Xinzhu) Wei Department of Computational Biology, Cornell University

Local ancestry inference (LAI) predicts the ancestral population of genomic segments in admixed individuals, providing critical insights into population history, natural selection, and genetic disease mapping. Current methods of local ancestry inference operate on tabular genetic data formats and face challenges scaling to datasets containing hundreds of thousands of whole-genome sequences. Here, we present GRG-LAI, a scalable statistical method that infers local ancestry using the genotype representation graph (GRG), a data structure that is related to the ancestral recombination graph (ARG) but more scalable to construct. In a GRG, nodes ancestral to an admixed sample node represent its ancestral haplotypes, and therefore carry population-specific ancestry information. GRG-LAI leverages these ancestral nodes and their population-specific frequencies estimated from reference samples to probabilistically assign local ancestry. Unlike existing methods that run sample-by-sample inference against the full reference panel, GRG-LAI instead infers local ancestry for all admixed samples simultaneously, thereby reducing computational cost and enabling scaling to large datasets with millions of samples. We evaluate GRG-LAI across different simulated recent admixture scenarios, including two-way European-Asian admixture, three-way African-European-Asian admixture, and four-way admixture involving African, European, Chinese and Japanese source populations. Across all settings, GRG-LAI runs significantly faster than other scalable local ancestry methods such as FLARE, while achieving comparable accuracy to FLARE. For instance, it takes around 7 minutes to infer 1,200 admixed individuals with 12,000 reference samples, whereas FLARE takes roughly 45 minutes. This computational efficiency enables local ancestry inference at biobank-scale, making large-scale studies of admixed populations more feasible than with current approaches.

143T Polygenic risk and association beyond linearity

Christophe Thomassin[1], Marc Franquesa Monés[2], David Bonet[3], Peter A. Gerlach[4], Marçal Comajoan Cara[5], Daniel Mas Montserrat[1], Alexander G. Ioannidis[1,3,4] [1]Department of Biomedical Data Science, Stanford University, [2]Centre de Formació Interdisciplinària Superior, Universitat Politècnica de Catalunya, [3]Genomics Institute, University of California, Santa Cruz, [4]Department of Genetics, Stanford University, [5]Department of Electrical Engineering and Computer Science, University of California, Berkeley

Polygenic risk scores assume additive linear effects between variants, often relying on GWAS-derived marginal effects of allelic dosages. Consequently, non-additive dominance and epistatic effects remain unmodeled, limiting the predictive power of current PRS methods. We present GenomEn, a framework for genotype-to-phenotype prediction that is designed to capture gene-gene interactions by fitting nonlinear estimators on selected variant patches and ensembling them at biobank scale. In 337,129 White British UK Biobank participants across 20 traits, GenomEn outperforms snpnet, LDpred2, and PRS-CS on all traits, improving genotype-only prediction by 21.3% on average. Variant importance scores capture known gene-gene effects in gout and LDL cholesterol, helping to explain the observed performance gains, and reveal heterogeneous trait architectures. Unlike many PRS methods, GenomEn includes X-chromosomal variants, boosting performance in androgenetic alopecia by 11.8% and capturing canonical and potentially novel AR-linked associations.

144T Population Genetic Analysis of highly degraded ancient-DNA data

Arev P Sümer[1], Benjamin M Peter[2] [1]Max Planck Institute for Evolutionary Anthropology, [2]University of California, Los Angeles

Ancient DNA studies are largely restricted by DNA preservation; samples from hot or humid climates, or from old layers, frequently only have miniscule amounts of DNA preserved. For Neandertals, studies have focused on the few fossils with extraordinary DNA preservation, but we lack tools for nuclear DNA analyses of contaminated low-coverage DNA, which, as a result, often remain unpublished.

Here, we introduce admixslug, a method designed to analyze ancient DNA data at ultra-low (<0.01x) coverages. admixslug leverages characteristics of individual sequenced molecules, such as fragment length and cytosine deamination patterns, to compute contamination-aware genotype likelihoods. Using these likelihoods, we estimate the conditional site-frequency spectrum of a low-coverage sample relative to a reference panel of high-coverage genomes.

Applying admixslug to published Neandertal data, we show that contamination patterns, the conditional site-frequency spectrum, and F-statistics can be reliably estimated. For example, when we downsample the data from Scladina (Belgium) to 0.001x (5,000 reads overlapping informative sites), we correctly identify it as most-closely related to the high-coverage Vindija Neandertal. We further find that contamination rates vary substantially with read length; (25% contamination for 35bp molecules, >90% contamination for molecules 60bp or longer). Thus, admixslug substantially extends the scope of nuclear DNA analyses for old and highly degraded ancient-DNA samples.

145T Neural networks preferentially learn additive models for genomic prediction

Aaron Kusmec Agronomy, Kansas State University

The first neural network (NN) for genomic prediction (GP) was proposed in 2011, yet over the last 15 years, there has been no “ChatGPT moment” in GP where NNs decisively superseded classical GP models based on Fisher’s additive model. Multiple meta-analyses indicate that NNs typically struggle to match the performance of gBLUP despite advantageous theoretical properties. When NNs do occasionally outperform gBLUP, this performance comes at the cost of extensive and computationally expensive hyperparameter tuning. I address this performance paradox by investigating the geometrical properties of single-locus genotype-phenotype maps and the learning dynamics of simple ReLU NNs with 1-2 neurons in a single hidden layer under perfect information scenarios. Using this model system, I prove constructive universal approximation (UA) theorems for the single-locus map. The theorems show that NNs with the UA property have weights restricted to low-dimensional submanifolds of weight space. These exact solutions are profiled by perturbation analysis and reveal that their basins of attraction have small volumes, rendering the solutions fragile and difficult to learn. Furthermore, I extend work demonstrating a bias toward parsimonious solutions in NNs trained on classification tasks to the regression setting. Analysis of the gradients and nonlinear learning dynamics of the model system demonstrates that the model NN preferentially learns additive solutions to single-locus maps regardless of the degree of dominance. This is accompanied by a further bias toward constant functions induced by the simplicity of the network and the ReLU activation function. Overall, these results (1) establish the difficulty of training NNs with the UA property for the simplest genotype-phenotype maps under perfect information conditions and (2) demonstrate a parsimony bias in NNs for GP problems toward the expectations of the additive model, explaining patterns of NN performance for GP identified in the literature of the past 15 years. Future work will extend the current analysis to multi-locus genotype-phenotype maps with epistasis and correspondingly larger NNs.

146T Functional and evolutionary determinants of protein divergence in Drosophila

Anthony Jacob Joson[1], Rob Kulathinal[2] [1]Temple University, [2]Department of Biology, Temple University

Protein evolutionary rates vary widely across genomes, yet the relative contributions of intrinsic molecular constraints and extrinsic selective pressures remain unresolved. Here, we investigate the functional and evolutionary determinants of protein divergence across eight *Drosophila* species by integrating comparative sequence analyses with phylogenetically informed expression models. Using multiple general linear models, we assess the effects of codon bias, sex bias, tissue specificity, and chromosomal location on protein evolutionary rate and divergence, measured as dN/dS (ω) and dN, respectively. Across all proteins, codon bias emerges as the strongest predictor of evolutionary constraint, with reduced codon bias associated with elevated dN and dS. Sex-biased and tissue-specific genes exhibit significantly higher protein divergence, indicating relaxed constraint and/or increased selective pressures associated with specialized function. Using the EVE model, we further identify widespread lineage-specific shifts in expression variance across species, revealing substantial regulatory divergence among genes exhibiting elevated protein divergence. We additionally identify significant differences in evolutionary dynamics between X-linked and autosomal genes that persist after accounting for codon bias and expression-related variables. Together, these results demonstrate that while intrinsic molecular properties strongly constrain protein evolution, extrinsic factors—including sex bias, regulatory divergence, and chromosomal context—play a central role in shaping patterns of protein divergence across *Drosophila*.

147T A Statistical Framework to Infer the Mutation Model of Tandem Repeat Variants via The Ancestral Recombination Graph

Sebastián Carlos Iturbe[1,2,3], Luis Gerardo Fernández-Luna[1,2], Carolina Lima[1,2], Diego Ortega-Del Vecchyo[3], Rori Rohlfs[1,2] [1]Department of Data Science, University of Oregon, [2]Insititute of Ecology and Evolution, University of Oregon, [3]International Laboratory for Human Genome Research, National Autonomous University of Mexico

Short tandem repeats (STRs), or microsatellites, are repetitive sequences that make up approximately 3% of the human genome. More than 10,000 STR variants are known to influence gene expression and account for 10–15% of its heritability. Additionally, many STRs contribute to various clinical conditions. Modern long-read sequencing technologies now enable us to determine STR genotypes with higher accuracy and investigate their role in complex traits. However, understanding the evolutionary and mutational behavior of STRs is essential to characterize their impact on different phenotypes. To this end, inferring the mutation model that governs STR variation is critical. This task is particularly complicated by the fact that the mutational dynamics vary substantially across loci. To solve this problem we created Tandem Repeat Ancestral recombination graph Mutational Analysis (TRAMA), a tool that infers the mutational process of each STR locus individually. For each STR, TRAMA uses information from the local Ancestral Recombination Graph (ARG), which encodes the complete local genealogical history of the samples. TRAMA uses a novel maximum-likelihood approach conditioning on the local ARG information to infer the most probable STR mutation model and its parameters per locus from observed variation in a set of samples. We demonstrate the method's accuracy through extensive simulations showing that it can infer the known mutational model acting on each loci along with the parameters that explain its evolutionary history across a wide range of scenarios. We further apply TRAMA to data from the Human Pangenome Reference Consortium - Phase 2. This work offers a path toward modeling STR mutational processes more accurately.

148T Integrating GGE-Biplot Analysis with Genome-Wide Association for Multi-Scale Genetic Dissection of Biomass Stability and Adaptation in Switchgrass

Jonathan Concepcion[1], Jason Bonnette[2], Li Zhang[2], Thomas Juenger[2], David Lowry[1] [1]Plant Biology, Michigan State University, [2]Integrative Biology, University of Texas - Austin

Dissecting biomass stability and adaptability across heterogeneous environments requires analytical frameworks that explicitly take into account genotype (G) and genotype-by-environment interaction (GE). A diverse switchgrass panel representing upland and lowland ecotypes was evaluated across six experimental sites in Illinois, Missouri, Texas, and Michigan from 2019 to 2024. Genotype plus genotype-by-environment interaction (GGE) biplot analysis was implemented to assess biomass stability, identify mega-environments, and quantify genotype performance using Average Environment Coordination (AEC) mean and stability projections. GGE “which-won-where” patterns revealed strong crossover interactions and distinct regional adaptation zones, while AEC vectors identified genotypes combining high biomass productivity with cross-environment consistency. Furthermore, ideal genotypes were identified using “means-vs-instability” across spatial and temporal scales. Stability and adaptability indices derived directly from GGE projections were subsequently treated as quantitative phenotypes for genome-wide association mapping. GWAS of GGE-derived stability metrics identified 61 SNPs associated with spatial stability and 59 site-dependent loci governing temporal stability. In addition, 62 SNPs associated with GGE-derived adaptability across spatial scales and 42 SNPs associated with temporal adaptability. The limited overlap among loci controlling stability versus adaptability indicates scale-dependent genetic architectures underlying biomass accumulation dynamics. These findings demonstrate that performance consistency and regional adaptation are genetically distinct, yet highly polygenic traits shaped by complex G×E interactions. By integrating GGE-biplot analytics with genome-wide association mapping, this study establishes a scalable framework for translating multi-environment performance metrics into genomic targets, accelerating the development of resilient, high-yielding switchgrass cultivars optimized for diverse agroecological conditions.

149T How do Recurrent Bottlenecks Shape Selection in Structured Populations?

Kendra D Zwonitzer[1], Cobi M Henry[1], Parul Johri[1,2,3] [1]Biology, University of North Carolina at Chapel Hill, [2]Integrative Program for Biological and Genome Sciences, University of North Carolina at Chapel Hill, [3]Department of Genetics, University of North Carolina at Chapel Hill

It is well-known that neutral processes such as genetic drift can often overpower selection when effective population sizes (N_e) are small, commonly seen during events such as population bottlenecks. Although populations are often modeled as panmictic, most real populations are structured. Both bottleneck events and metapopulation structure are known to increase stochasticity and may cause beneficial alleles to become lost or fix more slowly than in panmictic populations, and in some cases, may cause deleterious alleles to persist. Although the evolutionary effects of bottleneck events and metapopulations have been well-described individually, many populations, particularly pathogenic and microbial species, may best be modeled as metapopulations where demes (subpopulations) undergo recurrent bottlenecks. Examples of these populations are malaria or lice, that live on or within hosts and generally exist in discrete subpopulations that occasionally interact via host-host contact (migration) or invade new hosts (bottleneck). Here, we ask how recurrent bottlenecks affect the time and probability of fixation and extinction of selected alleles in a metapopulation. We employ forward-in-time simulations to extensively evaluate the effect of the size and duration of the bottleneck, number of demes (subpopulations), within-host carrying capacity, migration rate, and selection coefficient. Understanding the effects of selection in such populations will be relevant to a large group of species, including some plants, mitochondria, and insects, that all undergo similar repeated bottleneck events.

150T Can neutral coalescent models capture the quirks of site-frequency spectra in Atlantic cod?

Jack Edwards[1], John Wakeley[1], Einar Arnason[2], Katrin Halldorsdottir[2] [1]Organismic and Evolutionary Biology, Harvard University, [2]University of Iceland

The Kingman coalescent is a robust model for understanding the genetic variation in many types of populations, provided the individuals comprising the population are exchangeable. Many confounding qualities, such as natural selection or population structure, can violate this exchangeability. This makes a standard Kingman coalescent a poor choice to model genetic variation in populations with such confounding factors. Different types of coalescent models have emerged in response to patterns of genetic variation observed in natural populations. A prime example of unusual variation is the case of highly fecund Atlantic cod. The biology of these fish allow for very skewed offspring numbers, with some individuals possibly having numbers of offspring on the order of the population size. These large family events result in a depletion of intermediate allele frequencies and an enrichment of high allele frequencies, making the site frequency spectrum appear U-shaped. A recent study by Arnason et al. demonstrated that neither a Kingman coalescent with a flexible demography nor a Xi-Beta coalescent (a type of coalescent that accounts for skewed offspring numbers) could reproduce the observed site frequency spectrum of these cod. They found that a modified Durrett-Schweinsberg coalescent, which models recurrent selective sweeps throughout the genome, fit the data better. We are interested in investigating an alternative coalescent model with migration but no selection to better capture the high-frequency tail of the data. We examine the effects of multiple merger events and migration on a flexible timescale under selective neutrality. We compared simulation outputs to the full unfolded site frequency spectrum from SNPs across all biallelic loci on the 19 non-inverted chromosomes of 68 Atlantic cod.

151T Quantifying turnover in microbial communities using the traveling salesperson problem

Chloe E Shiff[1], Maike L Morrison[2], Shaili Mathur[1], Noah A Rosenberg[1] [1]Stanford University, [2]Santa Fe Institute

Many microbiome studies yield time-series data with species composition proportions at each of a series of points in time. The temporal turnover in species composition can provide important information about ecological dynamics. In this study, we present a novel statistic to quantify the amount of turnover present in time-series data that records species compositions. Our proposed statistic, Turnover Across Compositional Orderings (TACO), is equal to the sum of pairwise distances between composition vectors that correspond to adjacent timepoints, normalized by the range that the sum can take across all possible temporal reorderings of the vectors. Because the problems of finding the minimum and maximum distances over all possible reorderings are analogous to traveling salesperson problems, to obtain these distances, we use established computational algorithms for such problems. We demonstrate TACO in two applications. First, we use TACO to understand the longitudinal dynamics of barcoded yeast strains in fluctuating experimental environments. Second, we apply TACO to time-series data of human microbiomes that have been perturbed by antibiotics. TACO provides numerical values that are comparable across many settings, and it can be applied to data arranged in ways other than temporally, such as data arranged in a single spatial dimension.

152T A moment projection framework for computing the expected site frequency spectrum under exact discrete Wright-Fisher dynamics

Jeremy J. Berg University of Chicago

The expected site frequency spectrum (SFS) under the discrete-time Wright-Fisher (DTWF) model encodes the combined effects of drift, selection, and demographic history on genetic variation. Computing it exactly has been challenging: the diffusion approximation discards higher-order dynamics that matter for large samples, while direct matrix methods scale with population size. We present a new approach that retains the exact combinatorial structure of the DTWF model while reducing the computation to a small adaptive ODE system.

Building on recent advances in moment-based inference, our method works by tracking the expected polynomial moments of the allele frequency distribution rather than the discrete density itself. The Wright-Fisher transition is a jump process whose generator expands into an exact infinite series—the Kramers-Moyal expansion—where each successive term corresponds to a distinct class of coalescent merger event: pairwise coalescence, triple mergers, simultaneous pairwise mergers, and so on. The standard diffusion truncates this series at second order; we retain it to arbitrary order. We project the generator onto a polynomial moment basis in which the dynamics close into a linear ODE system. The resulting transition matrix decomposes into a small set of universal constant matrices that depend only on the combinatorics of the Wright-Fisher jump process—not on the population size, selection coefficient, or sample size. At runtime, changing demography enters only through a rescaling of these pre-computed matrices. To maintain computational efficiency, two adaptive truncation schedules dynamically restrict the moment hierarchy and generator bandwidth by exploiting the rapid decay of higher-order terms. Finally, the expected SFS for any sample size is recovered via a single linear projection.

This framework reveals several theoretical insights. First, the higher-order terms in the expansion cleanly separate coalescent event topologies by their timescale, making explicit the simultaneous multiple merger structure that is implicit in the DTWF transition matrix. Second, going beyond the diffusion limit requires retaining not only higher-order neutral drift terms but also the corresponding selection corrections to lower-order terms—terms which are traditionally discarded, but which we derive exactly. Third, the approach generalizes beyond the Wright-Fisher model to any exchangeable reproduction mechanism whose frequency transition has polynomial moments, encompassing the forward duals of Lambda- and Xi-coalescents. We discuss the computational implications and the relationship to existing approaches.

153T Eco-evolutionary Inference for Valley Fever

Julian Gottfried Biology, Stanford University

Valley Fever is an emerging disease in the Southwestern USA caused by the spores of the fungus Coccidioides immitis. Despite the threat to human health we remain ignorant of much of its ecology and environmental presence. In particular, the fungus may rely on natural mammalian hosts – desert rodents – for key stages of its life cycle. Detecting the fungus in the wild is challenging, which has slowed the exploration of these questions. My work approaches the issue from a population genetic framework, extending phylodynamic methods to detect the influence of historical mammal population densities on Valley Fever population parameters. My approach sheds new light on the nature of an important disease and presents a new method for inferring tangled population processes from genomic data.

154W Non-invasive genetic monitoring of a Ugandan elephant population using simulation-based spatially explicit close-kin mark-recapture

Gilia Patterson[1,2], Silas Tittes[2], Nelson Ting[2], Andrew Kern[2], Peter Ralph[2] [1]University of Montana, [2]University of Oregon

Close-kin mark-recapture (CKMR) is a promising, recently developed method for estimating population size from genetic data that identifies related pairs of individuals in a sample and estimates population size from these pairs. Because CKMR takes advantage of more information contained in genetic samples than widely used genetic capture-recapture methods, CKMR has the potential to improve genetic monitoring for many threatened populations.

African elephants in Uganda are one population that could benefit from CKMR. Elephants in Uganda have faced intense threats over the past few generations from habitat loss, poaching, and human conflict, and 50 years ago, many populations were thought to be on the brink of extirpation. Elephant populations have since significantly increased; however, we still know very little about many of these populations, especially those in densely forested areas where collecting genetic samples is challenging and labor intensive. However, CKMR has never been applied to elephant populations, and a main reason is because of a lack of appropriate methods. Current, likelihood-based CKMR methods do not account for the complex spatial and social dynamics of elephants, do not work well with low-quality genetic data collected from elephant dung samples, and often assume independence between pairs of individuals, an assumption that does not hold in small populations.

To overcome these limitations, we extend our recently developed simulation-based spatially explicit CKMR method to model matrilineal social dynamics and high uncertainty in kin pair identification due to low quality genetic data. We then apply our method to elephants in Kibale National Park in Uganda. When tested on simulated elephant populations, we find that CKMR estimates using our novel method perform about 12 percent better than genetic capture-recapture, and that with more sequenced loci, CKMR can reduce error in population size estimates by almost 50 percent. Using microsatellites sequenced from elephant dung piles, we estimate that current population size in Kibale is 684 elephants. We find that immigration into Kibale National Park was a significant factor in elephant recovery, with at least 10 elephants per year immigrating into the park over the past 50 years.

155W Ancient Bottleneck events, varied patterns of selection, and increased risk of maladaptation shape the past, present and future population structure of a widespread conifer

Sean M Collins[1], Anna Lincoln[2], Amanda R De La Torre[1] [1]Forestry, Northern Arizona University, [2]Grand Junction Office, Bureau of Land Management

Demographic events and natural selection can have independent and confounding effects on population structure and adaptation capability. This means different species could have had very different paths to their current population structure. The influences of demography and selection along a species' evolutionary timeline can influence responses to current and future stressors. We used a range-wide dataset in Ponderosa pine (Pinus ponderosa) to better understand the influences of demography and selection on current population structure and future maladaptation risk. Samples were sequenced with 55,000 capture-seq probes generating 300-400Mbp per sample; 1,501,748 SNPs were retained for downstream analysis. We examined current population structure, changes in effective population size, evidence for selection using simulations, genome and environment association methods and future maladaptation risk using genomic offset. We identified four genetic clusters within the species; numerous bottleneck events that mirror glacial advances and past droughts; the presence of purifying and balancing selection in response to moisture availability; many geographic areas with high amounts of genomic offset were found under near optimal climate scenarios. Differentiation across varieties has mostly been driven by two processes (1) bottleneck events causing random losses of alleles and decreased connectivity, (2) spatially heterogeneous selection and purifying selection in response to varied moisture levels. Many areas are predicted to be at risk of future maladaptation, which coincide with areas showing increased mortality events and decreased regeneration in previous studies. Previous evolutionary events and rapidly changing climate have shaped future maladaptation risk as drastic changes from current genomic structure are needed to meet a rapidly changing climate. The results of our study can aid in determining most at-risk areas in need of future conservation efforts and where seed sources may be most likely to perform best under future climate scenarios.

156W Isolation in real-time: the demographic and fitness consequences of declining immigration

Jeremy Summers[1], Faye Romero[2], James Schmidt[2], Daniel Seidman[2], Elissa Cosgrove[3], Tori Bakley[4], Sahas Barve[4], Reed Bowman[4], John W Fitzpatrick[3], Nancy Chen[5] [1]Biological Sciences, Virginia Tech, [2]Biology, University of Rochester, [3]Ecology and Evolutionary Biology, Cornell University, [4]Archbold Biological Station, [5]Ecology and Evolutionary Biology, University of California, Los Angeles

Habitat fragmentation threatens populations globally, cutting off the demographic buffering effects of immigration and the benefits of resulting gene flow. However, the consequences of immigration and gene flow can change over time, and measuring this change is crucial for understanding whether populations can persist through isolation and the impact of conservation efforts such as transolations. Studying the temporal effects of immigration remains challenging due to the need to directly track dispersal into populations and measure the reproductive success of immigrants and their descendants. A population of the Federally Threatened Florida Scrub-Jay (Aphelocoma coerulescens) at Archbold Biological Station that has been monitored since 1969, accruing data on immigration rates, individual life histories and fitness data, and a 16-generation population pedigree, offers a unique opportunity to investigate the impacts of immigration over time. The population is undergoing the process of isolation but has remained stable despite declining immigration and increasing inbreeding. Using a path analysis that linked variation in environmental conditions, population density, and levels of inbreeding to demographic rates, I identified how density-dependence and compensatory trends in breeder survival have buffered the population against decline. To address the multigenerational impact of gene flow, I compared the fitness of individuals based on their immigrant ancestry using a 16-generation population-wide pedigree. Despite the history of high immigration into the focal population, immigrant ancestry displayed additive benefits for females, but not males, demonstrating sex-specific fitness impacts of gene flow in a population experiencing the beginning phases of isolation. We found further sex differences in the lifetime reproductive success of F1 breeders and the association between heterozygosity and juvenile survival, emphasizing the importance of social factors in the effectiveness of gene flow. These results elucidate the mechanisms that underlie demographic and genetic rescue and the dynamics of populations becoming isolated.

157W Genetic early-warning signals of habitat fragmentation in dynamic migration networks

Ryan Chaffee[1], Gili Greenbaum[2], Jaehee Kim[3] [1]Molecular Biology and Genetics, Cornell University, [2]Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, [3]Computational Biology, Cornell University

Habitat fragmentation, the process by which habitat patches in a migration network become disconnected, is a leading cause of wildlife population declines, increased extinction risk, and biodiversity loss. Understanding how fragmentation impacts genetic diversity is essential for informed conservation. As anthropogenic fragmentation from road construction, deforestation, agriculture, and climate change increases globally, predicting its effects on genetic variation and population viability becomes vital. Existing population genetic theoretical frameworks that analytically link migration networks to genetic diversity often rely on simplifying assumptions, including symmetric migration rates, constant effective population size, and migration-drift equilibrium. However, fragmentation induces heterogeneous, time-varying connectivity and non-equilibrium demographic dynamics that violate these assumptions.

We represent landscapes as dynamic, directed, weighted migration networks and use forward-time simulations in SLiM to quantify how progressive connectivity loss reshapes genetic diversity and structure through time, across diverse network topologies, migration asymmetry, and fragmentation patterns and rates. Motivated by early-warning theory, we also evaluate and propose genetic early-warning indicators of impending genetic erosion: trend-based statistics and critical-slowing-down indicators in genetic time series, as well as deviations from analytical population-genetic predictions, to identify conditions under which equilibrium-based inference is reliable. Together, these models and analyses provide a general framework for investigating non-equilibrium dynamics in fragmenting populations, supporting early detection of connectivity loss in conservation management via genetic monitoring.

158W Comparative mutation load in the California Conservation Genomics Project

Anne Nakamoto[1], Erik Enbody[2], Russ Corbett-Detig[1] [1]Biomolecular Engineering, University of California, Santa Cruz, [2]Computational Biology, Cornell University

Earth's biodiversity has been generated through the evolution of genomic sequences. An understanding of the evolutionary history of these sequences can be used in the field of conservation biology to identify threats to biodiversity, and respond with informed management decisions. While traditional conservation efforts have focused on the protection and recovery of individual species, a more holistic and forward-looking approach is to assess the population health of many species across the landscape, and identify regions in need of management attention. As part of the California Conservation Genomics Project (CCGP), we have built an extensive dataset containing species of conservation interest sampled across California, allowing a landscape genomics approach to conservation. Population health can be assessed using many metrics, such as genetic diversity, effective population size, homozygosity, inbreeding depression, and deleterious mutation load. Much attention and regulations have focused on genetic diversity, especially in efforts to rescue species of conservation concern. However, recent work suggests that the genetic burden imposed by deleterious mutations, along with the demographic history of a population, might be more relevant metrics to assess extinction risk. In this study, we identify deleterious genomic variants in CCGP species based on evolutionary constraint. We use this metric of deleterious variation, and thus genetic load, to investigate differences in load between species populations. By taking geographic location and phylogenetic relatedness into account, we also make broad comparisons of load across different species and the landscape. The understanding of population health we gain here can be used to inform management decisions for the conservation of California's biodiversity, and our workflow can be applied broadly to regions beyond California.

159W Population genomics of the stony coral *Acropora millepora* and its symbionts across the Great Barrier Reef

Arjun Sai Krishnan[1], Daria Bykova[1], Carla Hoge[2], Ana Pinharanda[1,3], Zachary Fuller[1,4], Josephine Nielsen[5], Veronique Mocellin[5], Line Bay[5], Peter Andolfatto[1], Molly Przeworski[1,6] [1]Biological Sciences, Columbia University, [2]Department of Human Genetics, The University of Chicago, [3]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, [4]Bristol Myers Squibb, [5]Australian Institute of Marine Science, [6]Department of Systems Biology, Columbia University

Reef-building corals support one of the most biodiverse habitats on Earth. Their survival depends on an obligate symbiosis with dinoflagellates, in which the algae provide organic compounds through photosynthesis, while the coral supplies inorganic nutrients in return. This symbiotic relationship is disrupted under heat or chemical stress, leading to the loss of symbionts, a phenomenon termed bleaching. As oceans warm, mass bleaching events have become increasingly frequent and severe, endangering coral survival worldwide. Yet coral susceptibility to bleaching and heat tolerance vary among individual coral colonies, both within and across species, owing to environmental conditions, differences in symbiont community composition, as well as host genetics. The relative contributions of these factors and the adaptive potential of coral populations remain unknown.

To improve our understanding of these questions, we focused on *Acropora millepora*, a common stony coral species found throughout the Indo-Pacific. We collected samples from 1,081 colonies across 40 reefs spanning 1,600 km of the Great Barrier Reef and phenotyped them for traits related to bleaching and heat stress response. We then sequenced whole genomes from 125 samples to high coverage (>15X) as an imputation panel and the remaining 956 samples to lower coverage (~2X). We used these data to infer population structure, demographic history, and signatures of selection in the coral host. We also relied on the whole genome sequencing data to estimate symbiont composition and compared our findings to the amplicon sequencing-based method that is commonly used. As we show, corals that host the *Durusdinium* genus are less bleached than those with *Cladocopium*, but may also be less likely to survive longer term. More generally, these data allow us to combine information about host genetics, symbiont type, and local environmental data in order to predict and better understand the bleaching process and, ultimately, the adaptive potential of this coral species.

160W Population genomic health of an entire community of California wildlife

Erik Enbody[1], Anne Nakamoto[2], Russ Corbett-Detig[2] [1]Cornell University, [2]UC Santa Cruz

Many of the world's wildlife populations are declining at an alarming rate, precipitating an unprecedented biodiversity crisis. Reductions in population size will eventually lead to isolation and loss of genetic diversity, which can reinforce population decline through mutational meltdown and eventual extinction. An outstanding challenge of conservation genetics research is to implement metrics that reliably predict population health and the factors affecting population change. Many measures of genetic diversity reflect demographic and evolutionary processes during recent glacial cycles, which makes their utility in conservation planning uncertain. One promising avenue is to establish baseline metrics for regional taxa that have experienced recent shared environmental pressures. Here, we leverage a large regional population genomic dataset — representing 221 species across 74 orders of California wildlife and nearly 20,000 individuals — to evaluate the factors that may affect population decline at the regional scale. We show that estimators of inbreeding, calculated from long runs of homozygosity, are better predictors of recent population changes than genetic diversity metrics, such as nucleotide diversity, that instead reflect long-term dynamics of population change. We next describe regional hotspots of inbreeding and how range effects influence different genetic diversity metrics. Finally, we summarize which genomic indicators of population health may be useful for informing conservation action and directing management decisions for an entire community of California wildlife.

161W A unified theory of genetic and species diversity

Meixi Lin[1], Jacky Kaiyuan Li[2], Moises Exposito-Alonso[1] [1]Integrative Biology, University of California, Berkeley, [2]Biostatistics Division, University of California, Berkeley

Global change is shifting ecosystems and imposing unprecedented pressure on biodiversity at all scales. The ecosystem-scale ecological changes and population-scale evolutionary processes are traditionally studied in isolation but deeply intertwined in reality. Ecological changes caused by species extinction influence the evolution of co-occurring species, whereas genetic extinction can alter species abundance in a community. To efficiently halt biodiversity loss, we need eco-evolutionary models that incorporate genetic data transcending traditional species boundaries. Here, we present an individual-based eco-evolutionary process model derived from classical macro-ecology and population genetics theories. By modeling the shared processes of drift, migration, speciation, and mutation that underlie both ecological and evolutionary patterns, our model generates simultaneous predictions for community-wide mutation and species abundance. For example, under the neutral assumption that individuals from different species are competitively equivalent, genetic diversity of an ecological community is not only a simple aggregate of intra-species variation, but correlated with the total carrying capacity of the community. Our model is currently being tested against several macro-genetics datasets, such as data generated from the Global Malaise Program. We envision that our model will serve as a powerful null model to harmonize multi-species genomic data and elucidate the biological processes underlying community genetic observations.

162W The Population Genomics of Invasion: Founder effects, local adaptation, and chromosomal inversions shape the rapid evolution of an invasive migratory fish

Azwad R Iqbal[1], Daniel J Hasselman[2], Jennifer K Grenier[3], Pavel V Dimens[1], Jacob B Landis[3], Adrian McNairn[3], Nina O Therkildsen[1] [1]Natural Resources and the Environment, Cornell University, [2]AquaVeritas Research, [3]Institute of Biotechnology, Cornell University

Understanding the genetic dynamics of biological invasions can address key evolutionary questions and inform interventions by managers to conserve native species. Biological invasions also present excellent opportunities to study rapid evolution because they serve as 'natural' experiments where species are exposed to novel environments and selective pressures. Invasions by anadromous fish – where marine adults return to freshwater to reproduce – are especially interesting as successful introductions of these species, while uncommon, can spread rapidly across variable environments. In this study, I leverage whole-genome sequencing to understand the rapid evolution of invasive American shad (*Alosa sapidissima*), an anadromous fish native to the Atlantic coast of North America that was transplanted to the Pacific in the late 19th century. In the 150 years post-introduction, American shad have experienced remarkable success, spreading across multiple Pacific coastal rivers and outnumbering native salmon during annual migrations.

To identify signals of genetic divergence and local adaptation in invasive shad populations, we sampled 420 individuals spanning both the native and invasive ranges. We used Haplotagging, a linked-read sequencing technology, to generate phased whole-genome sequencing data for all individuals. Unlike standard short-read sequencing, Haplotagging preserves long-range haplotype information for individual reads, allowing for enhanced structural variant detection and robust variant phasing.

By leveraging demographic history inference, we found that Pacific populations experienced a sharp reduction in effective population size immediately following introduction that drives contemporary patterns of genome-wide diversity. Admixture and principal component analyses revealed that all Pacific populations diverged from their Atlantic counterparts and could be grouped into northern and southern Pacific clusters.

To identify genomic regions associated with local adaptation, we conducted selection scans to identify loci putatively under positive selection in invasive populations. We used a combination of divergence summary statistics (Fst, Dxy), haplotype-based selection scan metrics (EHH, iHS), and the Ancestral Recombination Graph (ARG) to identify and characterize regions under selection. We found outlier regions overlapping genes associated with neurological function, eye development, metabolism, and immune system development. We also discovered a putative chromosomal inversion segregating between rivers and by inland migration distance. Preliminary evidence suggests this inversion may be associated with migratory capacity, though further work is necessary. Overall, despite strong founder effects, invasive shad exhibit signatures of rapid adaptation to novel environments.

163W Nothing to see here? Population and species level comparisons highlight lack of parallelism in polygenic trait evolution

Claire Schraidt[1], Kerry Reid[2,3], Eric Palkovacs[2], John Carlos Garza[3], David Post[1] [1]Ecology and Evolutionary Biology, Yale Unviersity, [2]Institute of Marine Sciences, University of California Santa Cruz, [3]NOAA Southwest Fisheries Science Center

The transition between marine and freshwater environments is one of the most extreme evolutionary transitions an aquatic organism can undergo. In fishes, this transition drives phenotypic changes related to almost every aspect of life: feeding morphology, osmoregulation, body shape and size, fecundity, and life history. In salmonids and stickleback, this transition has given rise to some seminal studies in parallel evolution, structural variation, and adaptive radiation. Alewife (*Alosa pseudoharengus*) have undergone a similar evolutionary transition from anadromy to freshwater, but less is known about the genomic mechanisms that facilitated this transition. We know from previous work that landlocking has led to independently evolving populations that have undergone similar phenotypic shifts, but the genetic changes driving this transition are unknown. Using whole genome sequencing and NCBI data, we explored the genetic underpinnings of this transition within and across species. First, we compared seven landlocked populations with two anadromous populations to determine the extent of parallelism. We found that, despite phenotypic parallelism, zero outlier loci were shared between all seven populations. Of the ~2000 outliers identified, 75% were unique to a single population and 21% were shared only between two. We broadened this observation to study synteny across multiple species in the evolutionary transition from marine to freshwater. Phylogenetic analyses showed high synteny and 1:1 ortholog ratios, and a low number of species specific gene counts suggesting a high level of genomic stability and conservation across species.

164W Genomic diversity across 200 plant species of variable Red List status

Andrew Fairclough[1], Jules Perez[2], Moisés Expósito Alonso[1] [1]Integrative Biology, University of California, Berkeley, [2]Center for Computational Biology, University of California, Berkeley

Protecting the genetic diversity of threatened species arose as a major conservation priority in recent years. In vertebrates, IUCN Red List status scales reliably with metrics of genetic diversity. However, given the vast differences in life history between plants and animals, it is unclear if this relationship holds for the plant kingdom. This study investigates the degree of contemporary genetic diversity loss in plants through a meta-analysis of 200 species with publicly available genomes and resequencing data and an evaluation on the IUCN Red List. We re-called several measures of diversity (nucleotide diversity, heterozygosity, frequency of runs of homozygosity, and historical effective population size) in species from their original published data and a standardized bioinformatic workflow. We then combined this with data on 15 functional traits—including life form, lifespan, and reproductive mode— for each species to investigate how threat status affects diversity and what species traits potentially attenuate or exaggerate diversity loss. Preliminary results indicate that the mating system explains the most variation in genetic diversity across plant species. These findings suggest that plant extinction risk assessments based on genetic diversity must integrate life-history traits to accurately contextualize genetic health.

165W Local Environmental and Regional Oceanographic Factors Shape Genetic Structure for *Crassadoma gigantea* Across Natural and Anthropogenic Reefs

Hayley Goss[1], Rachel D. Simons[2], Paige Miller[1], Susan F. Zaleski[3], Robert J. Miller[1], Donna M. Schroeder[3], H.M. Page[1] [1]Marine Science Institute, University of California, Santa Barbara, [2]Earth Research Institute, University of California, Santa Barbara, [3]Bureau of Ocean Energy Management, Department of Interior

Understanding population connectivity is essential for assessing ecosystem resilience and informing management decisions. Estimating connectivity in marine systems is challenging, as models often reflect potential rather than realized connectivity. In the Santa Barbara Channel (SBC), more than a dozen oil and gas platforms are currently undergoing various stages of decommissioning, raising key questions about their connectivity with, and effects on, surrounding ecosystems. To address this, we evaluated potential connectivity between platforms and nearby natural reefs and assessed the extent to which biophysical models and environmental factors described genetic connectivity and fine-scale structure in *Crassadoma gigantea* (the purple-hinged rock scallop).

We genotyped 344 individuals from 30 sites in the SBC and Southern California Bight (SC Bight) at 10,545 single-nucleotide polymorphisms (SNPs), identifying 97 outlier loci (0.92%). Analyses revealed low but significant population structure and isolation by oceanographic distance, with biophysical modeling derived distances explaining 17 - 35.7% of pairwise genetic differentiation. Redundancy analysis identified 25 outlier SNPs significantly associated with environmental variables, 7 of which aligned to known functional genes involved in metabolism, catabolism, regulation, and gene expression, suggesting a potential role in local adaptation. Potential connectivity and genetic analyses indicate that platforms are well integrated with surrounding reefs, and no genetic differences were observed by habitat types. This integration of biophysical modeling, environmental data, and genetic analysis refines connectivity assessments between platforms and natural reefs, and it provides insight into how local and regional factors shape fine-scale genetic structure in marine invertebrates with moderate pelagic larval durations (PLDs).

166W Population genomics of ʻōpaeʻula (*Halocaridina rubra*): conservation implications for highly structured populations of the state shrimp of Hawaii

Russell Clark Fitzgerald-Cook, Justin Havird Integrative Biology, The University of Texas at Austin

Current understanding of the population genomics of many wild species still relies on single or multi-locus barcoding studies. Updating these underpowered studies using genome-level datasets allows us to make reliable claims about population structure, genetic diversity and divergence to inform conservation decisions. Initial barcoding in 2005 of mitochondrially encoded COX1 in ʻōpaeʻula (‘little red shrimp’) identified eight divergent populations. However, mitochondrial and nuclear genomes often have starkly different evolutionary trajectories despite pressure to co-evolve. We have generated 2bRAD-seq nuclear SNP data and parallel whole mitochondrial genome sequencing of 163 ʻōpaeʻula individuals collected between 2005 and 2024 from anchialine habitats on Oahu, Maui, and Hawaii. The nuclear dataset (N=248) identifies nine lineages of H. rubra, while the mitochondrial dataset (N=163) identifies ten. Both datasets identify the same general lineages and relationships, but some patterns of discordance exist likely due to introgression or incomplete lineage sorting. These data resolve relationships between lineages that were polytomies in the early COX1 barcoding phylogeny. We identify four singleton genetic lineages that are found in only one habitat, making them extremely vulnerable to extirpation. Three of the four are on Oahu -- the oldest of these three islands, and the most disrupted by tourism. Fst values indicate that ʻōpaeʻula sequentially colonized new habitats as the islands formed. Oahu lineages have Fst values of ~0.3-0.4 when compared to one another – the same range as Fst values for comparisons between islands. Our work highlights the value of whole genome-level data for identifying genetically vulnerable lineages in natural populations.

167W Genomic insights into equids at the edge of extinction

Arielle S Fogel[1], Mathew M Ndunda[2], Julien F Ayroles[3], Daniel R Rubenstein[4], Andrew G Clark[1] [1]Molecular Biology and Genetics, Cornell University, [2]Veterinary and Capture Services, Kenya Wildlife Service, [3]Integrative Biology, University of California, Berkeley, [4]Ecology and Evolutionary Biology, Princeton University

The causes of species declines in modern times are well-documented (e.g., human activities) but the genomic consequences of these declines are less clear. Indeed, genomic data from endangered species suggest that the "genetic costs" of species decline, such as loss of diversity and increased inbreeding and frequency of deleterious mutations, are not inevitable. Long-term population size or life history traits may mitigate the effects of recent declines but testing this requires studying a wider range of species that vary in their evolutionary pasts and life histories.

Here, we focus on one of Africa's most endangered mammals, the Grevy's zebra (*Equus grevyi*). Since the late-1970s, the Grevy's zebra, a keystone herbivore in East Africa, has declined in numbers by >83%. Less than 3,000 wild animals remain today. We generated a haplotype-resolved, chromosome-level reference genome for this species and confirmed previously described, extensive chromosomal rearrangements in the genus: 14 of the 22 Grevy's zebra autosomes correspond to chromosomal fusions relative to the horse (*E. caballus)*. We also generated high coverage, whole genome data (mean=45x) for 62 wild Grevy's zebras and 3 wild individuals from its non-endangered sister species, the plains zebra (*E. quagga*). Consistent with their recent decline, Grevy's zebras are depleted of rare genetic variation (median Tajima's D=1.37) and plains zebras have on average 2.3x higher per site diversity (π) compared to Grevy's zebras. Plains zebras also have 1.2x higher X:autosome diversity (π) relative to Grevy's zebras, in line with species differences in mating systems. However, Grevy's zebras—despite their current effective population size of ~500—are only mildly inbred, with long runs of homozygosity (>1 Mb) spanning on average 9.8% of their genomes (F_{ROH}). As expected, Grevy's zebras have higher levels of F_{ROH} compared to plains zebras, but do not exhibit abnormal stripes which have been reported for inbred plains zebras. Together, our results support the idea that the suite of "genetic costs" predicted for endangered species are not necessarily coupled, at least immediately after a bottleneck. Because we have also biologically sampled ~85% of the species and animals are individually-identifiable using AI-based image recognition, our reference genome and high-quality genotype discovery set will facilitate studying nearly an entire species at individual-level resolution—a long-held dream in conservation biology.

168W Higher genomic offset in planted forests of the UK reveals maladaptation risks under climate change

Guillermo Friis[1], Louise Gathercole[1], Michael Charters[1], Mohammad Vatanparast[1], Rômulo Carleial[1], Nicola Cotterill[2], Tom Jameson[2], Nadia Barsoum[2], Richard Buggs[1], James S. Borrell[1] [1]Plant Health & Adaptation, Royal Botanic Gardens, Kew, [2]Forest Research

The UK aims to dramatically accelerate woodland creation by 2050, yet the consequences of active afforestation for landscape-level genetic diversity and resilience remain unclear. Here, we combine environmental data and whole-genome sequencing to evaluate whether current UK afforestation strategies provide locally suited genotypes for large-scale planting, relative to wild, recently established populations of pedunculate oak and silver birch, two species that together account for approximately 17% of UK tree cover. Based on data from more than 4,500 individuals across over 80 populations, our results show that genetic diversity metrics in planted populations are comparable to those observed in naturally colonised woodlands. However, planted populations exhibit higher within-group coancestry and moderate genetic homogenisation among sites, potentially reducing adaptive differentiation. In contrast, naturally colonised populations show higher inbreeding coefficients in both species, potentially reflecting fragmentation of parental populations. Genotype association analyses reveal divergence between planted and naturally colonised forests at functionally relevant loci, indicating distinct selective pressures associated with afforestation practices. Based on projected future climate data, genomic offset under warming scenarios varies markedly across the landscape, highlighting geographically heterogeneous risks of maladaptation due to climate change. Furthermore, planted populations frequently show higher predicted maladaptation than nearby naturally regenerated cohorts. To address this issue, we identify regions of the UK where forest reproductive material is currently insufficient to support climate-resilient, genetically informed afforestation, based on the availability and climatic representativity of registered seed sources. Together, these results indicate that contemporary planting pipelines may constrain both genetic and climatic diversity in afforested landscapes, increasing vulnerability to future environmental change.

169W Genetic diversity loss in the Anthropocene

Moises Exposito-Alonso University of California Berkeley, Howard Hughes Medical Institute

Anthropogenic habitat loss and climate change are reducing species' geographic ranges, increasing extinction risk and losses of species' genetic diversity. Preserving genetic diversity is key to maintaining species' adaptability, but how much genetic diversity has been lost globally across species and ecosystems? Is it already measurable? How much will be lost in this century? In this talk, I will critically test whether temporal meta-analyses using public genetic data in the last decades can measure losses, and what data may be needed. Because for most species we will never have sufficient empirical data, I introduce a genetic-diversity-area-relationship power law model that predicts genetic diversity loss from habitat area loss in recent decades, which can be used in conservation policy. In combination with a new dynamic population genetic theory framework, I will discuss future expected global losses due to increased genetic drift. Together, this research suggests that more than 10% of genetic diversity may already be lost for many threatened and nonthreatened species, surpassing the United Nations' post-2020 targets for genetic preservation.

170W The genetics of inbreeding depression in a pedigreed wild population of Florida Scrub-Jays

Faye G. Romero[1], Jeremy Summers[2], James Schmidt[1], Daniel Seidman[1], Sahas Barve[3], John W Fitzpatrick[4,5], Nancy Chen[6] [1]Biology, University of Rochester, [2]Biological Sciences, Virginia Polytechnic Institute and State University, [3]Archbold Biological Station, [4]Ecology and Evolutionary Biology, Cornell University, [5]Cornell Lab of Ornithology, [6]Ecology and Evolutionary Biology, University of California, Los Angeles

Inbreeding depression, or the reduced fitness of offspring of related parents, can cause the rapid decline and eventual extinction of threatened populations. Despite numerous examples from the wild, we still know little about the fine-scale temporal variation in inbreeding as populations near extirpation, the number and identity of the loci responsible for inbreeding depression, and the relative contributions of deleterious variation to negative fitness consequences. With increasing habitat fragmentation and loss of genetic variation in species worldwide, it is imperative that we further elucidate the genetic architecture of inbreeding and how it relates to fitness decline over short timescales in natural populations. Our understanding of these questions has been limited by our inability to collect comprehensive genetic and fitness data for threatened populations, except in a few special study systems. A wild population of endangered Florida Scrub-Jays (*Aphelocoma coerulescens*) at Archbold Biological Station has been studied since 1969, resulting in lifetime fitness measures for thousands of individuals on a 16-generation pedigree. Previous work in our study population showed that decreased immigration from smaller peripheral populations resulted in increased levels of inbreeding and reduced fitness via inbreeding depression. To further explore the genetics of inbreeding depression, we performed whole genome sequencing of all individuals in our study population from 1999-2021 (~4,000 jays). We characterized levels of inbreeding and deleterious variation across the genome over time and tested for associations with different fitness components. We found temporal variation in the length distribution of runs of homozygosity and the frequency of different classes of deleterious variants across our 22-year study period, reflecting the history of inbreeding and gene flow. We also found significant associations of both genetic load and inbreeding on individual fitness, further shedding light on the genetic basis of inbreeding depression. This study provides a detailed look at the complexities of inbreeding depression in the wild and highlights how its consequences can differ across individuals and over short timescales. Our results have important implications for conservation management priorities.

171T Realized relatedness in a wild baboon population and its implications for kin discrimination

Carlota Galán-Plana[1], Shuyu He[1,2], Iker Rivas-González[1], Harald Ringbauer[3], Elizabeth A Archie[4], Susan C Alberts[2,5], Jenny Tung[1,2,5,6] [1]Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, [2]Department of Biology, Duke University, [3]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, [4]Department of Biological Sciences, University of Notre Dame, [5]Department of Evolutionary Anthropology, Duke University, [6]Faculty of Life Sciences, University of Leipzig

Kin discrimination—the differential treatment of conspecifics based on relatedness—is thought to evolve because it enables individuals to maximize inclusive fitness and avoid inbreeding. However, selection pressure for kin discrimination in a mating context may be weak, if sex-biased dispersal is very effective at separating opposite-sex kin. To investigate this possibility, we combined behavioral, demographic, and genome resequencing data from 524 wild baboons from a long-term field study in Kenya. We investigated how often dispersal not only separates adult opposite-sex kin, as emphasized in previous work, but also *reunites* adult kin who have no prior familiarity with one another because they were born in different groups. These latter pairs are informative about how often individuals are faced with a kin discrimination problem, in the absence of typical social cues. They also constitute a natural experiment enabling us to test whether kin-biased mating behavior depends on familiarity. We first estimated realized genetic relatedness—i.e., the actual proportion of the genome that two individuals share—by identifying segments identical-by-descent (IBD) in all pairs of individuals. We identified 1431 opposite-sex pairs of adult baboons who were born in different social groups, but co-resided after the male dispersed into the group of his female relative, creating dyads of “unfamiliar” kin. We found that the occurrence of observed mate-guarding was lower for unfamiliar kin than for non-kin, whether familiar or unfamiliar (17.2% of unfamiliar kin dyads mated at least once versus 24.5% of non-kin). Pairs of unfamiliar kin who mated were also 1.37x less likely to ever produce offspring than non-kin who mated. Notably, the offspring of unfamiliar kin who reproduced had lower estimated inbreeding coefficients (mean ± SD = 0.086 ± 0.01) than predicted for the theoretical offspring of unfamiliar kin who mated, but never produced actual offspring (0.114 ± 0.06). Our results indicate that encountering unfamiliar, opposite-sex kin happens rarely but regularly in our population. Because familiarity-based mechanisms can therefore fail, our results suggest that other mechanisms of kin discrimination have evolved as a response.

172T Understanding the genomic architecture of inbreeding in the endangered mountain gorilla

Aditi Prasad Bangalore Venkateshwara[1], Tierra Smiley Evans[2], Benard Ssebide[3], Martha M Robbins[1], Linda Vigilant[1] [1]Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, [2]UC Davis School of Veterinary Medicine, [3]Gorilla Doctors

Individuals in small, isolated populations are likely to inbreed, leading to the accumulation of mutation load and reduced fitness. Puzzlingly, some inbred populations persist through time and show no signs of inbreeding depression. Theory suggests that in small populations, deleterious mutations may be exposed to natural selection in inbred individuals and potentially 'purged' from the population. The efficacy of purging depends on the type of mutation and the population's demographic history, including the intensity of the bottleneck and the time elapsed since it occurred.

Here, we synthesize theoretical expectations and empirical genomic evidence to evaluate how missense and loss-of-function deleterious mutations respond to inbreeding and bottlenecks in two populations of the endangered mountain gorilla (*Gorilla beringei beringei*) in Bwindi Impenetrable National Park, Uganda, and the Virunga Massif of Rwanda, Uganda, and the DRC. Both populations are known to have undergone bottlenecks and possess genomic runs of homozygosity (ROH) longer than 50 kb. We have sequenced 47 new mountain gorilla individuals and used 12 publicly available genomes to integrate measures of mutation load with recent demographic reconstructions. The Virunga mountain gorillas have a higher fraction of the genome in long runs of homozygosity than the Bwindi gorillas, indicating relatively more recent inbreeding in the Virunga mountain gorillas.

We will estimate the impact of purging on inbreeding depression and assess when genetic drift overwhelms selection, constraining the evolutionary potential of small populations. By clarifying how inbreeding and demographic history differentially shape genomic resilience and risk, this work provides the background information necessary to assess the genetic resilience of these highly endangered populations in the future.

173T Population genomic analyses reveal asexual reproduction may facilitate spread in a recently introduced sea anemone

Keira Monuki[1,2], Rachael Bay[3], Eric Sanford[2,3] [1]Evolution and Ecology, University of California, Davis, [2]Bodega Marine Laboratory, [3]University of California, Davis

Global change is driving unprecedented alterations in species distributions. For example, human vectors can lead to range shifts through the introduction of species into new environments. Population genomic tools allow us to track populations across the introduced and endemic ranges to better understand if, and how, gene flow and genetic variation facilitate species introductions. We took a population genomics approach to investigate the spread of a recently introduced sea anemone first reported in Tomales Bay, CA in June 2022. Our morphological and molecular analyses confirm this species is the Small Brown Sea Anemone *Anthopleura hermaphroditica*. This asexually reproducing sea anemone is endemic to Chile, Australia, and New Zealand and, to our knowledge, this is the first record of *A. hermaphroditica* in the Northern Hemisphere. We conducted ddRAD sequencing across populations within Tomales Bay, as well as from populations in the endemic ranges in Chile, Australia and New Zealand. Results show high genetic relatedness and heterozygosity within Tomales Bay populations, indicating high levels of asexual reproduction may facilitate the rapid spread of this species without a reliance on strong propagule pressure. Biogeographic comparisons indicate strong genetic divergence between regions within the endemic range, where introduced populations are more similar to some endemic regions than others. These results highlight the complexity of understanding both contemporary and historical species movement. Overall, our study underscores how population genomics tools can improve our understanding of species distribution changes to better predict how species may continue to spread in the future.

174T Phylogenetic Modeling of the Endangered Pacific Pocket Mouse Using Mitochondrial Genomes

Jackson Wolfe[1], Carter Sherrit[2], Erik Funk[3], Aryn Wilder[3], Arun Sethuraman[2] [1]Biology, San Diego State University, [2]San Diego State University, [3]San Diego Zoo Wildlife Alliance

When populations decline to small sizes, reduction in fitness from genetic drift and inbreeding can drive deleterious feedback loops that threaten populations with extinction. While management strategies such as translocations and reintroductions can be valuable tools to assist population persistence, anthropogenic threats that lead to population bottlenecks and isolation can greatly alter patterns of genetic diversity, making the appropriate management actions unclear. The Pacific pocket mouse (Perognathus longimembris pacificus) is an endangered Heteromyid rodent restricted to three remnant populations along the coast of southern California, USA. Here we generated ~5x coverage whole genomes from historical and current (reintroduced from the San Diego Zoo Wildlife Alliance's breeding program) samples to investigate range-wide patterns of population divergence and connectivity, including two populations from the northern and southern extents of the geographic distribution that went extinct in the mid-1900's using short read sequencing. We assembled de novo mitochondrial genomes (16,293 bp) from these using the GetOrganelle v.1.7.1 toolkit, aligned all 36 mitogenomes using MAFFT v.7, and built a likelihood-based phylogeny using RAxML v.8.2.13. Interestingly, phylogenetic reconstruction indicates that contemporary and historical samples from the same locality do not form clades, instead samples from historical Dana Point formed a clade with a non-sister taxon, and reconciled these inconsistencies through SNP-based demographic analyses. Through this work we highlight the value of historical genomics in assessing risk and guiding conservation decisions, particularly when extirpation, and limited genetic diversity constrain modern management. Although historical baselines often cannot be restored, conservation strategies can leverage genomic insights to enhance future adaptive potential and long-term resilience of threatened species.

175T Recombination and divergent selection shapes the genome architecture of the Hawai'i Island *Metrosideros* speciation

Askhan Shametov[1], John Kelly[1,1], Elizabeth Stacy[2], Jae Young Choi[1] [1]Ecology and Evolutionary Biology, University of Kansas, [2]School of Life Sciences, University of Nevada Las Vegas

A major goal of modern evolutionary biology is to understand the process of adaptive radiation. However, understanding the evolutionary history of species that diverged recently and rapidly remains challenging because of incomplete lineage sorting and hybridization, which can cause discordance among gene trees and blur evolutionary relationships. Certain genomic regions may capture the evolutionary history of recent radiations more so than others, however, and thus may be more useful for understanding speciation. For instance, genomic regions with low recombination rates are more strongly impacted by linked selection, display lower tracts of introgression, and show increased lineage sorting relative to high recombination regions (HRR) of the genome. Alternatively, selection in ancestral population in low-recombination regions (LRR) of the genome might lower effective population size accelerating lineage sorting. Hence, we might expect LRR to better reflect the evolutionary history of recently diverged lineages. Here, we used the landscape-dominant tree species, *Metrosideros polymorpha*, to test the utility of LRR for elucidating the evolutionary relationships among incipient species. Within continuous forest on Hawai'i Island, *M. polymorpha* comprises four ecologically diverged varieties that are partially isolated by reproductive barriers and appear to capture the early stages of speciation. We generated a genetic map of *M. polymorpha*, representing the first for a wild tree species, using an outbred F_1 hybrid crossing scheme involving four genetically distinct parental lines from two varieties. We analyzed 88 F_2 recombinants and 55,680 informative markers to produce a total map length of 591.31 cM. We then used published population genomic data for all four varieties and consistent with expectations, we observed slightly elevated polymorphism in HRR. But this relationship was weak and statistically significant for only one variety (Pearson's $r = 0.08$, $p = 0.019$). Using genome-wide SNPs across the four varieties, a topology-based phylogenetic analysis showed no single dominant topology, which was consistent with the recent and rapid diversification of this group. Subsequent analysis of just the LRR, however, revealed a single dominant topology grouping glabrous populations separately from pubescent populations. Interestingly and unexpectedly, we also detected significant support for the same species tree through analysis of the HRR, which also showed reduced rates of introgression. These findings suggest that high-recombination genomic regions may act as barriers to hybridization at the early stages of speciation. The observed elevated lineage sorting in LRR can be due to reduced effective population size, while strong divergent selection in HRR may be driving force shaping genome-wide variation in the incipient speciation of *Metrosideros.*

176T A genetic characterization of the elusive elephants of the Lisima highlands (Angola)

Carla Hoge[1], Katherine Solari[2], Dmitri Petrov[2], Jordana Meyer[2,3,4], Kerllen Costa[3,4,5], Steve Boyes[3,4], Anna Songhurst[6,7], Daniella Chusyd[8], John Novembre[1] [1]Human Genetics, University of Chicago, [2]Stanford University, [3]National Geographic Okavango Wilderness Project, [4]The Wilderness Project Foundation, [5]Fundação Lisima, [6]EcoExist Trust, [7]Department of Zoology, University of Oxford, [8]Department of Epidemiology and Biostatistics, Indiana University-Bloomington

Characterizing elusive populations remains a technical and logistical challenge, yet is essential for the conservation of endangered species. One such population of African elephants exists in the Angolan "Lisima Lya Mwono" (Lisima) highlands, where seven years of fieldwork have yielded few direct observations. Very little is known about the Lisima elephants, which reside nearly 500km upstream of the nearest large population of elephants in the Botswana Okavango Delta. To characterize the genetic composition, history, and connectivity of the Lisima elephants, we utilized non-invasive fecal sampling and whole genome sequencing (WGS). We applied computational methods adapted from ancient DNA literature to mitigate biases associated with low-coverage fecal DNA and to integrate data from multiple sequencing technologies. We sequenced 24 fecal samples collected in 2024 to 0.1-8X coverage, identifying eight distinct individuals. These individuals share a single mitochondrial haplotype and include several 2nd and 3rd degree relatives. While this suggests a small census size, we find no evidence of long runs of homozygosity or decreased heterozygosity, implying that any isolation or population reduction is likely recent We analyze these data jointly with WGS generated from 53 tissue samples from nearby elephant populations (Botswana and Zambia) and 191 publicly available high coverage WGS samples from across sub-Saharan Africa. Direct estimates of pairwise similarity and models of effective migration rates demonstrate that, while the Lisima elephants are not greatly diverged from elephants from surrounding regions, they are surprisingly more genetically similar to elephants from northwestern Namibia relative to those of Botswana and Zambia (a result supported by genome-wide autosomal markers and mtDNA haplotype sharing). Further, we use identity by descent and linkage disequilibrium based methods to infer recent demographic history. Together, these results characterize a small, cryptic population of an endangered species and highlight an unexpected genetic connectivity with significant implications for conservation efforts.

177T The genomics of urban adaptation in house mice in the northeastern United States

Rene Clark[1], Zade Alafranji[1], Stephen Kupchella[1], Sam Giancarli[1], Adrienne Kasprowicz[1], Nina Mallalieu[1], Logan Lacy[2], Maddie Balman[3], Megan Phifer-Rixey[1] [1]Drexel University, [2]University of California San Diego, [3]Northeastern University

Urbanization is often coupled with increased habitat fragmentation and shifting selective pressures on wildlife. While some species can readily adjust to these man-made habitats, others are less resilient and may experience population declines. As cities continue to expand, understanding how wild populations respond to urban growth will be important. Rapid adaptation (e.g., adaptation on ecological timescales) is one potential response that can lead to long-term population persistence. However, the genomic mechanisms driving rapid adaptive responses remain relatively understudied. This is particularly true in wild populations and vertebrate taxa, as these systems can be challenging to work with, and adaptation may include complex behavioral, morphological, and metabolic traits. Here, we compared populations of wild house mice (*Mus musculus domesticus*) from urban and rural areas in three major metro regions (New York City, NY, Philadelphia, PA, and Richmond, VA) to investigate evolutionary responses to urbanization. House mice are particularly well-suited for this project as they are widespread in both cities and the surrounding areas, and have a wealth of genetic resources because of their regular use as a model organism. We found clear and consistent morphological and genomic differences between our sampled populations, including (1) differences in body size, (2) lower genetic diversity and higher relatedness in rural relative to urban populations, and (3) shared genomic signatures of urban adaptation across cities in functionally important genes. Taken together, these results highlight the potential for urbanization to lead to parallel, repeated evolutionary change across multiple populations and can help us better understand how other species respond to rapid anthropogenic shifts.

178T Assessing the response of giant kelp to marine heatwaves using temporal genomics

Meaghan I Clark[1], Roy L Roberts[1], Rodrigo Beas-Luna[2], Jan Freiwald[3], Daniel Reed[4], Filipe Alberto[5], Malin L Pinsky[1]
[1]University of California, Santa Cruz, [2]Universidad Autónoma de Baja California, [3]Reef Check Foundation, [4]University of California, Santa Barbara, [5]University of Wisconsin, Milwaukee

Climate change forces organisms to face new environmental challenges at an unprecedented rate, threatening global biodiversity. Adaptation is one mechanism by which populations and species can meet these challenges and persist. Demographic recovery due to heritable genetic changes that increase population viability is termed evolutionary rescue. Studying evolutionary rescue in empirical systems is critical to understanding adaptation and its efficacy for mitigating extinction from climate change and other stressors. Here, we focus on giant kelp (*Macrocystis pyrifera*), an ecologically, economically, and culturally important brown alga that experienced declines in response to marine heatwaves. We use preliminary temporal sampling from before and after the major marine heatwave from 2014-2016 to test for loss of genetic diversity associated with bottleneck events and adaptation to ocean temperature. This will provide important information on whether evolutionary rescue can lead to giant kelp populations that are resilient to future warming, and what conservation interventions and management strategies will be most effective to combat population extinction.

179T Uncovering the genetic architecture of herbicide resistance in the agricultural weed *Amaranthus tuberculatus* using popGWAS

William Yaeger, Julia Kreiner Department of Ecology & Evolution, University of Chicago

Herbicide resistance in agricultural weeds is an emerging evolutionary phenomenon that poses substantial challenges to crop production across the world. Understanding the genetic architecture of resistance is important for predicting how this trait evolves in response to selection imposed by agriculture, but for most herbicides, the current understanding of resistance genetics is largely limited to a few large-effect target-site resistance (TSR) mutations affecting herbicide target enzymes. Recent work has demonstrated that for glyphosate resistance, TSR mutations alone often fail to explain the full extent of phenotypic variation observed in the field (Kreiner et al. 2021), suggesting a more complex genetic basis that also encompasses non-target-site resistance (NTSR) mechanisms. This work follows up on this observation with a multi-trait GWAS study dissecting the genetic basis and polygenicity of NTSR to three common agricultural herbicides in *Amaranthus tuberculatus*, a major North American agricultural weed exhibiting widespread resistance to multiple herbicides. To maximize GWAS efficiency and sensitivity, we are implementing a conditional popGWAS approach (Pfenninger 2025), which relates population trait means to genome-wide allele frequencies while accounting for the contributions of known TSR alleles, using a diverse panel of accessions collected from populations across North America. By clarifying the basic genetic architecture of herbicide resistance, this work sets the groundwork for more accurate tracking of how resistance allele frequencies in *A. tuberculatus* populations fluctuate across time and space in response to selection. It additionally paves the way for the development of polygenic scores for assessing the degree and prevalence of herbicide resistance in agricultural settings, a powerful potential management tool for farmers.

180T Using century-old fish collections to study rapid adaptation across space and time in the epicenter of marine diversity

Marianne Dehasque[1], Jason Selwyn[2], Chris Bird[2], Malin Pinsky[3] [1]Ecology and Evolutionary Biology, UC Santa Cruz, [2]Texas A&M University–Corpus Christi, [3]UC Santa Cruz

Much of Earth's biodiversity has been shaped by rapid bursts of adaptive evolution. Rapid evolution through natural selection may be a critical survival mechanism for species facing fast-changing environments (*i.e.,* evolutionary rescue). Yet our understanding of the genomic architecture and origin of adaptive variation remains limited due to a lack of data spanning relevant temporal and spatial scales. In this study, we leverage historical fish collections from the Philippines to test the hypothesis that local adaptation provides the raw substrate for rapid evolution through time. Using historical and modern collections of the delicate round herring (*Spratelloides delicatulus*), a small and widely distributed coastal fish, we will test (1) if local adaptation across geographic regions maintains genomic variation and (2) if this variation has served as the substrate for rapid evolution over the last century. To test these hypotheses, we have generated ten haplotype-resolved assemblies and are now whole-genome sequencing over one thousand historic and contemporary individuals from multiple locations to identify genomic regions under selection across time and space. Overall, this project will uncover the broad patterns and generalizability of rapid adaptation and its importance for population persistence in the face of climate change.

181T Evolution and diversity of major histocompatibility complex II gene *DRB* in yellow-tailed woolly monkeys (*Lagothrix flavicauda*)

Chloe R Leaman[1], Melissa A Zarate[2], Sofia M Weaver[1], Evelyn L Pain[1], Maria Solis Quispe[3], Mrinalini Watsa[4,5], Fanny M Cornejo[6], Sam Shanee[7], Christopher A Schmitt[1,2] [1]Biology, Boston University, [2]Anthropology, Boston University, [3]Conservación Amazonica, Estación Biologica Los Amigos, [4]Field Projects International, [5]San Diego Zoo and Wildlife Alliance, [6]Yunkawasi Peru, [7]Neotropical Primate Conservation Peru

The yellow-tailed woolly monkey (YTWM; *Lagothrix flavicauda*) is a critically endangered platyrrhine primate endemic to the montane cloud forests of Andean Perú. Little is known about the immune capacity of YTWM; however, genetic variation in major histocompatibility complex (MHC) genes are well-known indicators of immune function. The MHC class II region of the genome plays a central role in vertebrate immune defense by enabling antigen-presenting cells to present foreign antigens to T cells, thereby linking innate and adaptive immunity. This may be particularly important to species with small, fragmented populations where reduced diversity can increase disease susceptibility and contribute to an extinction vortex. Here, we present the first data from exon 2 of the hypervariable MHC class II gene *DRB* for YTWM. We used noninvasive fecal samples as a DNA source to successfully amplify exon 2 of the *DRB* gene for 111 YTWM from 6 populations across their known range. We sequenced these amplicons using Oxford Nanopore Technologies (ONT) third-generation sequencing. A preliminary analysis (n = 10), shows evidence of *DRB* duplication within the trimmed and mapped samples. Consensus sequences for each of the samples, aligned to an *Ateles belzebuth* reference sequence, show strong alignment across the 270bp of the *DRB*e2, and potentially contain 22 unique SNPs shared between samples resulting in 3 alleles. BLASTing of consensus sequences show 99.53% similarity to *Brachyteles hypoxanthus*, the sister genus to *Lagothrix*. Next steps include further bioinformatic analyses to identify unique alleles and population-level variation, compare sequence variation to closely related primate species, and to detect signatures of selection and recombination in the gene region. These data will provide a foundation for population-wide analyses of immunogenetic diversity in a critically endangered species, while offering insight into the potential disease resilience of YTWM with important implications for conservation management and comparative primate immunogenetics.

182T From Fragmentation to Near-Complete: A Framework for Managing Genetic Diversity in the Spoon-billed Sandpiper

Rachel L Davis[1], Erik Enbody[1], Fyodor Kondrashov[2] [1]Cornell University, [2]Okinawa Institute of Science and Technology

With only 400–500 adults remaining, the critically endangered spoon-billed sandpiper (*Calidris pygmaea*) continues to decline at 5-10% per year. Genomic resources essential for conservation management of this species are outdated and highly fragmented, with the most recent public assembly released in 2018. Accurate assessment of inbreeding and subtle population structure in a rapidly shrinking population requires a more contiguous reference genome. To address this, we generated a substantially improved genome assembly and conducted the first population genomic analysis of modern spoon-billed sandpipers sampled from the remote breeding population and chick headstarting program of Meinypil'gyno, Russia (n = 44). Using Oxford Nanopore sequencing (44 Gb; ~37× coverage) and the long-read Flye assembler, we produced a 1.18 Gb genome with a contig N50 of 17.8 Mb, outperforming other reference genome candidates. Using this assembly, we detected weak population structure and heterogeneous inbreeding (F_ROH ≈ 0.00–0.16), with long runs of homozygosity revealing recent autozygosity in a subset of birds. This work establishes a transferable framework for data-driven management of genetic diversity in critically endangered species undergoing captive rearing, reintroduction, and non-random mating, an increasingly common challenge in wildlife conservation programs.

183T Decoding the population structure and history of the world's deadliest cat

Victoria Grant Stanford University

Black-footed cats (*Felis nigripes*) are one of Africa's least understood felines as the population dynamics and demographic history of these solitary creatures are not well described. Reports of steady decline of present-day populations resulted in the IUCN Red List categorizing the species as vulnerable to extinction. As populations decline and become isolated from each other, they become susceptible to strong genetic drift and inbreeding which can lead to the accumulation of deleterious alleles and increased extinction risk. However, the IUCN cited data deficiencies across the species range as a limitation in this categorization for black-footed cats. In cases where ecological surveys are lacking, range-wide population genomic surveys can improve our understanding of population dynamics. Here, we sequenced whole genomes of black-footed cats (N=44) from across their distribution in the first genomic study of free-roaming individuals. To do so, we incorporated whole genome sequences generated from modern biological samples and century-old museum specimens. We assembled a highly contiguous reference genome using a combination of PacBio HiFi reads and publicly available Hi-C data and investigated the demographic history, population structure, and genetic diversity of wild black-footed cats. We found evidence of historical effective population sizes estimated in the tens of thousands which is comparatively lower to other felids. Consistent with long-term moderate population size, we found low present-day genetic diversity genome-wide (~ 0.0004), comparable to other species at high risk of extinction. However, despite low genetic diversity, we find that black-footed cats do not appear to harbor long runs of homozygosity. Low genetic diversity seems to result from demographic patterns as recent population size estimates support long-term modest size. PSMC simulations appear to be consistent with a modest historical population size of ~10,000. However, GONE results suggest recent population contraction which can contribute to future genomic erosion. . We compared genomic data in individuals from across the range to evaluate patterns of population structure and found evidence of higher genetic similarity between individuals in closer geographic proximity. Overall, these results provide range-wide information about the demographic history and present-day genetic diversity of an understudied species.

184T Genomic Offset Predicts Allele Frequency Changes but Not Always Population Fitness: Insights from Forward-Time Simulations

Andrew R McCracken, Daniel Saddler, Melissa Pespeni, Joaquin Nunez Biology, University of Vermont

Predicting how populations will respond to rapid environmental change is a pressing challenge in evolutionary biology and conservation. Genetic offset (GO) statistics are gaining popularity in forecasting population maladaptation, yet whether predicted GO values correspond to realized evolutionary changes across diverse ecological scenarios, including future projections, remains largely untested. We used forward-time simulations in SLiM v5.0 to systematically assess how predicted GO values compare (or correlate) to simulated long-term evolutionary changes. Our models of locally adapted populations were used to test the impact of varying trait architectures (oligogenic to highly polygenic), life histories, migration rates, environmental clines, and metapopulation structures. We evaluated model performance by tracking changes in population allele frequencies (F_{ST}) and population fitness, both predicted and measured, over time. We found that GO values and realized change in allele frequencies were highly correlated across the majority of simulated scenarios. However, this association decreased with low migration rates and in edge-of-range populations where the necessary alleles to adaptively track environmental changes were either inaccessible or already near fixation. Additionally, GO values did not always reflect fitness outcomes, particularly in environments beyond the range of conditions used to fit the models. Together, these findings reinforce genomic offset as a robust predictor of allelic turnover under environmental change, while providing a mechanistic framework for understanding when and where predictions may be most or least reliable for assessing maladaptation risk.

185T "The contribution of introgression to local adaptation in North American jackrabbits"

Jeffrey M Good, Jessica D Scales University of Montana

Resolving the origin of adaptive variation is critical to understanding how contemporary populations adapt to local environments. Adaptive variation may arise through multiple sources including new mutation, standing genetic variation, and gene flow (i.e., introgression) from other species. However, teasing apart the contributions of these sources to the process of adaptation remains challenging. Here, we present a large-scale analysis of whole genome resequencing data collected from two widespread jackrabbit species, white-tailed jackrabbits (*Lepus townsendii*) and black-tailed jackrabbits (*Lepus californicus*), with a history of hybridization. Combining field-collected specimens with extensive museum archives, we sequenced >300 low-coverage genomes sampled from across the range of both species. We use these data to characterize patterns of genetic diversity, population structure, and the spatial and temporal patterns of introgression. We then use these landscape-level maps of genomic diversity to evaluate the contribution of introgression to locally adaptive variation in seasonal coat color camouflage in white-tailed jackrabbits. These analyses provide a range-wide assessment of how introgression has shaped adaptive variation of an ecologically important trait underlying evolutionary responses to climate change. These resources will lay the foundation for combining signatures of introgression with genotype–environment associations to dissect the evolutionary origins of climate adaptation between two widespread ecological competitors in North America.

186T Museum genomics clarifies the evolutionary history and taxonomic status of the extinct Lotis blue butterfly (Lycaeides anna lotis)

Alia Donley[1], Zachariah Gompert[1], Matt Forister[2] [1]Department of Biology, Utah State University, [2]University of Nevada

Insight into human-driven impacts on biodiversity is important for conservation. For at-risk or extinct species, museum collections may provide the only remaining window into the evolutionary histories and the biodiversity that has been lost. The Lotis blue butterfly, *Lycaeides anna lotis,* was last observed in the wild in 1983. It was designated as a protected subspecies under the Endangered Species Act in 1976 and has undergone multiple taxonomic revisions since the species was first described in 1879 by Lintner. This study uses population genomic data to resolve the evolutionary history of the Lotis blue and address three key questions: (i) which extant *Lycaeides* taxa was the Lotis blue most closely related to, (ii) how long ago did the Lotis blue diverge from its closest relatives, and (iii) did it represent a distinct subspecies (or species), or was it less genetically differentiated than existing *L. anna* populations? We also reconstruct the demographic history of the Lotis blue relative to other *Lycaeides* populations to determine whether the Lotis blue was already declining prior to anthropogenic habitat degradation or historically persisted as a small population that ultimately failed. To estimate phylogenetic relationships and divergence times among the Lotis blue and other North American *Lycaeides* butterflies, we applied complementary phylogenetic approaches using BEAST and CASTER. Demographic history was evaluated using estimates of effective population size over time. Phylogenetic analyses show that the Lotis blue was most closely related to extant *L. anna* populations from northern California. BEAST analyses indicate that the Lotis blue diverged from these populations about 800,000 years ago. Levels of divergence and genetic differentiation among the Lotis blue and other *Lycaeides* populations suggest that the Lotis blue was as genetically distinct as extant subspecies but not species. Together, these findings place the Lotis blue firmly within *Lycaeides anna*, while highlighting its long and distinctive evolutionary history.

187T Longitudinal population genetics of the cooperatively breeding Acorn Woodpecker (*Melanerpes formicivorus*)

Chi Wei[1,2], Joseph Haydock[3], Eric Walters[2] [1]Biological Sciences, Old Dominion University, [2]Old Dominion University, [3]Gonzaga University

As a facultative cooperative breeder, Acorn Woodpeckers (*Melanerpes formicivorus*) form familial social groups to raise young. In central coastal California, a longitudinal population at Hastings Natural History Reservation has been studied since 1968, providing a rare opportunity to disentangle the effects of dispersal, inbreeding, and temporal trends on the population genetics of a wild population. From 1990 to 2025, 603 breeding individuals of both sexes were genotyped using restriction site-associated DNA sequencing across 62 social groups. Pairwise identity by descent was assessed to evaluate the role of immigrants (i.e., individuals hatched outside the study area) versus residents (i.e., individuals hatched within the study area). Isolation-by-distance was examined to investigate dispersal distances of immigrants. We found that pairwise relatedness among immigrant-immigrant, immigrant-resident, and resident-resident pairs was similar, suggesting that immigrants are primarily recruited via short-distance dispersal from adjacent territories. Additionally, this population showed little genetic structure, with high gene flow likely driven by frequent dispersal among social groups. However, temporal trends in heterozygosity revealed a slow decline over time (Mann-Kendall test = −0.26, p = 0.028). These results suggest that additional ecological constraints may influence the maintenance of genetic diversity in this population.

188W Genetic and environmental basis of Allen's rule in tropical and temperate house mice (*Mus musculus domesticus*)

Sylvia M Durkin[1], Michael W Nachman[2] [1]Integrative Biology, University of California - Berkeley, [2]University of California, Berkeley

Complex phenotypes are governed by a combination of genetic and environmental effects. However, understanding the relative roles of these effects is difficult in systems where experimental manipulation is not feasible. Here, we use a combination of behavioral, molecular, and quantitative genetic approaches to disentangle the genetic and environmental control of adaptive tail length divergence between tropical and temperate house mice. Tail length is longer in tropical house mice reflecting adaptation to warmer temperatures through increased extremity length (Allen's rule). We focus on the maternal environment, as this is particularly important in mammals due to the obligate association between mother and offspring during gestation and lactation. First, through cross-fostering experiments we found that tail length, but not body weight, is strongly influenced by the postnatal maternal environment as a function of a nesting-dependent difference in rearing temperature. Further, analysis of tail vertebrae variation in cross-fostered and pure line pups suggests that the maternal environment influences tail vertebrae length, but not number, which is likely genetically determined. Second, we studied gene expression in the developing tail to uncover specific biological processes and genes associated with the maternal effect on tail length, implicating known bone growth and temperature sensitive signaling pathways. Lastly, through genetic mapping with 449 F3 progeny in a cross between temperate and tropical mice, we discovered a shared QTL between total tail length and tail vertebra number that contains the known tail length regulator *Lin28a*. Increased *Lin28a* expression is known to lead to a greater number of tail vertebra, and we found that *Lin28a* expression is upregulated in tropical house mice, which have an evolved increase in tail vertebra number. Together, these results provide a detailed account of how both genetic differences and environmental variation influence adaptive phenotypic divergence.

189W Early life-stage thermal resilience is determined by climate-linked regulatory variation

Joaquin C. B. Nunez[1], Sumaetee Tangwancharoen[2], Kylie Finnegan[3], Eliza C Bufferd[1], Olin C King[1], Luke C Proud[1], Brent L. Lockwood[1] [1]Biology, University of Vermont, [2]Chulalongkorn University, [3]University of Vermont

Despite decades of research in environmental change, we know relatively little about the genetics of environmentally influenced traits across the life cycle of species with complex life histories. Previously, we reported that natural variation in heat tolerance is life-stage specific in *Drosophila melanogaster*, suggesting that thermal selection predominantly targets the early embryonic life stage. Here, we used advanced introgression and pooled whole-genome resequencing to map the genomic basis of enhanced embryonic heat tolerance in a neotropical line of *D. melanogaster*. We identified two loci on chromosomes 2R and X that were consistently targeted by 16 generations of thermal selection across six replicate introgressions. We compared alleles in these regions to published datasets of natural variation from North America and Europe using the DEST dataset. This analysis revealed that two SNPs associated with embryonic heat tolerance exhibited both clinal and seasonal patterns, with the seasonal variation significantly correlated with environmental variability in average precipitation and temperature variance across space and time. Further, tropical alleles at both loci exhibited enhanced embryonic heat tolerance in the *Drosophila* Genetics Reference Panel (DGRP), demonstrating the genotype-to-phenotype link in an independent set of diverse genetic backgrounds. The two SNPs lie in the putative regulatory regions of the genes SP70 and sog, where allelic differences in gene expression correlate with the heat tolerance phenotype. Overall, our results suggest that regulatory loci that influence embryonic heat tolerance are under selection in nature. Our study extends previous work in developmental genetics of *Drosophila* by characterizing the genomics of an ecologically relevant developmental trait in natural populations.

190W Fitness effects of mitonuclear incompatibilities in Swordtail (Xiphophorus) hybrids

Nemo V Robles, Madison Murata, Maria Jose Rodriguez Barrera, Molly Schumer Biology, Stanford University

As populations diverge, they accumulate genetic variants which may result in negative interactions in hybrid individuals, known as hybrid incompatibilities. Hybrid incompatibilities are often studied in a laboratory environment; as a result, their impact on hybrid fitness under natural conditions is poorly understood. Swordtails (*Xiphophorus*) are freshwater fish endemic to México with a history of extensive hybridization between species. Previous work by our group identified a lethal mitonuclear incompatibility in F2 hybrids between *X. malinche* females and *X. birchmanni* males, involving the *X. malinche* mitochondria and *X. birchmanni* version of the gene *ndufs5* which results in stalled embryonic development. We have further identified two non-lethal hybrid incompatibilities involving the *X. birchmanni* versions of *ndufa13* and *atp5mg* which result in reduced body size. While *ndufa13* and *atp5mg* incompatible individuals survive in the lab, we have not found them in the wild despite extensive sampling. This suggests that their natural environment selects strongly against these hybrid incompatibilities, possibly contributing to the speciation process in this group. We hope to identify how *ndufa13* and *atp5mg* reduce hybrid fitness in different environmental conditions to uncover what selection pressures hybrids face in the wild.

191W There are no unlinked loci: How pedigrees couple neutral genealogies across the genome

Maximillian V Newman Department of Genetic Medicine, University of Chicago

Implicit in current population genetic methods is the assumption that genes far enough apart in the genome, such as on different chromosomes, have independent genealogies. These gene genealogies, however, are subject to the same pedigree, the random graph capturing the reproductive history of the population. I show how the pedigrees of structured populations record macroscopic demographic events, and how these events couple genealogies across the genome. In particular, I will explain how neutral gene genealogies far apart on the genome are only independent in the absence of large migrations and uneven offspring distributions. These results suggest that genome-wide association statistics, which aggregate weak signals across many loci, may be sensitive to pedigree-induced correlations even between unlinked regions of the genome.

192W High-resolution mapping of a rapidly evolving complex trait reveals genotype-phenotype stability and an unpredictable genetic architecture of adaptation

Jessica Smiley-Rhodes[1], Mark C Bitter[1], Skyler Berardi[2], Jack Beltz[2], Dmitri A Petrov[1], Paul Schmidt[2] [1]Genetics, Stanford, [2]University of Pennsylvania

The extent to which adaptation can be predicted, particularly for traits with complex genetic bases, is unknown. We leveraged a model complex trait, model species, and high-powered longitudinal sampling design to test the efficacy of genomic prediction of complex trait variation and evolution in ecologically relevant settings. We monitored genome-wide allele frequencies and pigmentation variation in genetically diverse populations of *Drosophila melanogaster* across seven generations of evolution in both field mesocosms exposed to natural environmental fluctuations, as well as mesocosms housed in a controlled, lab-based setting. At two time points throughout trait evolution, we conducted a high-powered, tail-based mapping of pigmentation, producing a well-resolved genotype-phenotype map that reaffirms canonical pigmentation genes and unveils novel loci. While we were able to use this map to correctly infer the direction of pigmentation evolution in both the field and lab mesocosms, the particular loci responding to selection, and thus architecture of adaptation itself, were largely unpredictable. We suggest this unpredictability to be a result of pleiotropic constraint, which was more pronounced in the field, relative to the lab-based environment. Finally, we quantified a striking stability of the genotype-phenotype map across genetically diverged populations, demonstrating that shifting epistatic landscapes associated with the evolutionary process itself do not alter trait architecture and preclude phenotypic prediction, provided the mapping is sufficiently powered. In concert our results highlight both the promise and limitations of genomic prediction, and exemplify the challenges of applying lab-based studies of complex traits to their evolutionary dynamics in the wild.

193W Hidden in plain sight: How Ks histogram dynamics can reveal and obscure ancient whole genome duplications.

Tamsen Dunn[1,2,3], Arun Sethuraman[3] [1]Ecology and Evolutionary Biology, University of Arizona, [2]San Diego State Univeristy, [3]Evolutionary Biology, San Diego State University

Polyploidy is a significant force in plant and animal evolution. Polyploidy is particularly common in crop species, and thus of great relevance to agriculture. Historically, ancient rounds of WGD have been inferred by the presence or absence of peaks on Ks histograms. However, there are some common simplifications in the interpretation of Ks histograms that can lead to misinterpretations. For example, it is often assumed that the placement of the peak in Ks space corresponds to the timing of WGD, or for autopolyploids may not be present at all. Polyploids are often classified and modeled as either auto or allopolyploids, but real species exist in a shifting continuum, affected by a return to diploidy. In this paper, we address these misconceptions using the novel polyploid simulation engine DemographiKS to demonstrate that the Ks histogram is highly sensitive to evolutionary parameters relating to the mode of origin, as well as demographic parameters such as migration events, population bottlenecks and expansions. Our results show that the location of the Ks peak may correspond to the time of parental divergence (for allopolyploids); might be hidden at Ks=0 (for autopolyploids under a population contraction); may correspond to the mean time of coalescence of the diploid ancestor (for autopolyploids under a population expansion), or the time of most recent migration between the parental species (for allopolyploids). We also fit simulated Ks histograms to polyploid genomes from four well-studied plant lineages that are neither wholly auto or allopolyploid, but exist as points on a multidimensional polyploid continuum (Coffea arabica, Zea mays, Populus trichocarpa, and Saccharum spontaneum), demonstrating that WGD simulations may be used to corroborate inferred demographies. In summary, we demonstrate that Ks histograms are information-rich, computationally tractable, and can be utilized to corroborate detailed evolutionary histories inferred by other methods.

195W Parallel Molecular Evolution Across Replicated *Anolis* Lizard Adaptive Radiations

Aryeh Miller[1,2], Jonathan Losos[1], Anthony Geneva[3] [1]Washington University in St. Louis, [2]Smithsonian Institution, National Museum of Natural History, [3]Rutgers University - Camden

Whether convergent phenotypes arise from similar or distinct genetic underpinnings is a central question in evolutionary biology. In this study, we sequenced the genomes of >99% of all 435 extant anole lizard (*Anolis*) species to dissect the molecular basis of their convergent adaptation and parallel adaptive radiations in the Greater Antilles. We find that anoles exhibit widespread positive selection on loci which regulate developmental processes. Using selection screens, we find these lizards exhibit a genome-wide excess of parallel positive selection, indicating that repeated co-option of loci has shaped these adaptive radiations. In addition, we find that selective regimes at genes underlying limb development—a hallmark of convergent evolution in anoles—co-vary with phenotypic state, with each ecomorph experiencing distinct selection pressures. Lastly, we show that a transcription factor involved in limb development has undergone remarkable parallel evolution in minute and slender twig-specialist anoles, discovering a single missense mutation at a highly conserved position arising independently in three subgenera over approximately 40 million years, mirroring the parallel evolution of twig anole craniofacial and limb morphologies. Together, these findings provide a unified view that natural selection can predictably yield convergence which scales multiple biological hierarchies, from the assembly of ecological communities, to the phenotypes of individual species, to the specific allelic variants potentially underlying such phenotypes.

196W Collateral fitness effects of mutation are not commonly caused by protein misfolding

Natalie Quan, Yuichi Eguchi, Ashley Brown, Kerry Geiler-Samerotte Arizona State Universit

Proteins misfold frequently in all cells, and this process has long been viewed as a universal source of mutational harm. Here we provide a broad, quantitative test of how strongly protein stability predicts cellular fitness. By integrating deep mutational scans spanning eight proteins from both bacteria and yeast, expressed under conditions where their function is either essential or gratuitous, we find that destabilizing mutations predict fitness only when the protein's activity is required for growth. When expression is gratuitous, predicted misfolding fails to explain the collateral fitness costs of mutation, even for variants known to misfold biochemically. These findings suggest that misfolding is not a universal, or even a dominant, driver of mutational harm and that other cellular processes may instead underlie the hidden costs of mutation.

197W Population genomics and shell morphometrics of a coastal dogwhelk identify history of glacial refugia

Emily K. Longman[1], Joaquin C.B. Nunez[1], Eric Sanford[2], Melissa Pespeni[1] [1]Biology, University of Vermont, [2]Evolution and Ecology, University of California, Davis

Contemporary patterns of genomic diversity hold key insights into species' vulnerability and adaptive potential in the face of rapid environmental change. However, for most species, particularly non-model and marine taxa, such knowledge remains elusive due to limited genomic resources and the masking of historical evolutionary patterns by high gene flow. Marine taxa with low dispersal potential may provide a unique window into the evolutionary history of marine ecosystems due to deep evolutionary divergences between populations. The channeled dogwhelk, *Nucella canaliculata*, is a species of predatory gastropod that has crawl-away young and resides on wave-exposed rocky headlands along the west coast of North America. We assembled a high-quality draft genome for this low-dispersing marine species and studied biogeographic patterns of genomic diversity and shell morphometrics in 19 populations distributed along ~1,500 km of the northeastern Pacific coast. We identified strong population structure with a phylogeographic break at Monterey Bay, a pattern which was mirrored by biogeographic variation in shell morphology. Additionally, patterns of genomic diversity and accompanying computer simulations suggested that there were at least two refugial populations during the last glacial maximum that experienced post-glacial expansion and admixture. Furthermore, we identified candidate loci that may underlie the adaptive basis of variation in shell morphology. Overall, these findings highlight that historical events can play a persistent role in structuring the genomic diversity of modern coastal populations, with important consequences for spatial variation in the adaptive capacity and vulnerability of populations to future climatic changes.

198W Origin and evolutionary history of an urban underground mosquito

Yuki Haba[1], Petra Korlević[2], Erica McAlister[3], Mara Lawniczak[2], Molly Schumer[4], Noah Rose[5], Carolyn McBride[6]
[1]Zuckerman Institute, Columbia University, [2]Sanger Institute, [3]Natural History Museum, [4]Stanford University, [5]UCSD, [6]Princeton University

Urbanization is rapidly reshaping landscapes around the world, which poses questions about whether and how quickly animals and plants can adapt. *Culex pipiens* form *molestus*, more commonly known as the "London Underground mosquito," has been held up as a benchmark for the potential speed and complexity of urban adaptation. This intraspecific lineage within *Cx. pipiens*, a major West Nile virus vector, is purported to have evolved human biting and a suite of other human-adaptive behaviors in the subways and cellars of northern Europe within the past 200 years. Form *molestus* features prominently in textbooks as well as scholarly reviews of urban adaptation. However, several lines of evidence are inconsistent with a recent urban origin for this mosquito.

I will present our recent efforts to understand the contentious origin and evolutionary history of the urban, human-biting mosquito. Our synthesis and meta-analysis of rich yet confusing literature show that its London Underground origin is unlikely (Haba and McBride 2022 *Current Biology*). Whole genome resequencing and population genomics of 800+ mosquitoes across ~50 countries again debunk the in situ evolution hypothesis and instead show that molestus first adapted to human environments >1000 years ago in the Mediterranean or Middle East, most likely in ancient Egypt or another early agricultural society (Haba et al. 2025 *Science*). I will outline implications of our results in urban evolutionary biology as well as in public health.

199W Evolutionary adaptation proceeds through a small number of phenotypic modules

Mohammad Hossein Donyavi[1,2], Reza Ghelich[1,2], Kara Schmidlin[1,2], Grant Kinsler[3], Kerry Geiler-Samerotte[1,2] [1]School of Life Sciences, Arizona State University, [2]Biodesign Center for Mechanisms of Evolution, [3]Department of Bioengineering, University of Pennsylvania

Understanding how the myriad molecular impacts of mutation percolate to influence higher-order traits and ultimately fitness requires compressing a many-to-many mapping into something tractable. Decades of theoretical work suggest this may be possible because biological systems are modular: effects of perturbation are often funneled through particular pathways or subsystems rather than propagating freely through the organism. Yet few empirical systems have been able to demonstrate such modularity at scale. Prior low-dimensional models hinted that mutation–fitness relationships can collapse onto a small number of latent axes, but these studies relied on simpler datasets, leaving open whether such structure generalizes to more genetically and environmentally complex systems. Here we show that, even across 774 diverse yeast lineages, fitness variation across 12 drug environments is organized by a strikingly low-dimensional structure defined by only a few inferred phenotypic axes that capture the main patterns of variation. Consistent with these lineages having evolved under strong selection pressure, they reveal a striking asymmetry in the genotype–phenotype map: their mutations exhibit broad pleiotropy, affecting nearly all inferred phenotypic axes, yet fitness in any given drug depends on a much sparser subset of the phenotypic modules these axes reflect. This architecture aligns with central expectations of evolutionary theory. Strong selection often favors mutations with broad physiological effects, whereas long evolutionary history shapes organisms into modular systems in which only certain trait combinations matter to fitness in particular environments. By compressing many-to-many relationships, this low-dimensional framework exposes the modular fitness space that constrains the pleiotropic effects of adaptive mutants. It also lays the groundwork for identifying the key phenotypic modules that matter for fitness across different environments.

200W Rapid genome-wide introgression reveals fitness advantage of immigrant genotypes

Benjamin A Flanagan[1], Arshad Padhiar[1], Foen Peng[2], Saif Quraishi[3], Andrea Roth-Monzón[1], Fahad Gilani[1], Lauren Simonse[1], Megan Maciejewski[1], Noah Reid[4], Milan Malinsky[5], Amanda K Hund[6], Daniel I Bolnick[1] [1]Ecology and Evolutionary Biology, University of Connecticut, [2]Biology, Haverford College, [3]Midwestern University, [4]Department of Molecular and Cell Biology, University of Connecticut, [5]Ecology and Evolution, University of Bern, [6]Biology Department, Carleton College

Evolutionary biology has long recognized the tendency for populations to be locally adapted to their ancestral habitat, resulting in higher resident fitness. However, immigrants can also introduce beneficial alleles. The resulting adaptive introgression is usually inferred retrospectively, rather than as a contemporary process. Here, we document exceptionally rapid ongoing adaptive introgression in a lake population of threespine stickleback (*Gasterosteus aculeatus*). In the first generations after a discrete immigration event, all chromosomes exhibited large increases in immigrant ancestry due to linkage disequilibrium. After a decade, the extent of introgression varied across the genome. The fastest-evolving genes included *Spi1b*, which enables an increased fibrosis defense against a previously common tapeworm, whose prevalence then declined dramatically. This case study highlights the capacity for immigration to supply beneficial alleles that drive rapid genome-wide evolution.

201W Bugs in the balance: local adaptation and balancing selection shape amino acid variation in *Anopheles*

Clara T Rehmann[1], Peter L Ralph[2], Andrew D Kern[1] [1]Institute of Ecology & Evolution, University of Oregon, [2]Department of Data Science, University of Oregon

Anopheles mosquitoes, the genus responsible for transmitting malaria, have large effective population sizes and maintain a high amount of ancestral polymorphism, limiting the effectiveness of standard population and phylogenetic approaches in describing their evolutionary histories and trajectories. Despite this, understanding the forces underlying patterns of genetic variation in these vectors is critical for informing genomic surveillance efforts. Here, we used a gene-based approach focused on putatively functional polymorphism within and between three *Anopheles* species, two from Africa and one from Southeast Asia, in order to understand and describe how variation is maintained in these vectors. Given their demographic histories and exposures to vector control methods, one might expect a priori to observe repeated instances of strong directional selection in the genome, similar to what is observed in Drosophila. However, in comparisons of polymorphism and divergence within and between species, this is not what we observe: instead, *Anopheles* seems to segregate an excess of nonsynonymous polymorphism. We find that genes involved in insecticide resistance, immune response, sensory perception, and environmental adaptation each violate neutral expectations in this way. In some cases, these genes also share common patterns of spatial nonsynonymous variation across species. Together, these observations suggest that *Anopheles*' maintenance of amino acid variation is due to balancing selection and local adaptation, rather than nearly neutral dynamics, presenting a striking contrast to the strong directional selection dynamics observed in similar systems and informing future genomic surveillance of these vectors.

202W Fixation Probabilities of Mutant Alleles in an Ecological Context

Kunaal Joshi[1], Susmita Halder[1], Adrian Gonzales Casanova[2], Michael Lynch[1] [1]Center for Mechanisms of Evolution, Arizona State University, [2]Arizona State University

For decades, population geneticists have relied upon a formula derived by Malecot and by Kimura to estimate the fixation probability of mutant alleles. Among other things, this formula leads to the conclusion that in sufficiently large populations the fixation probability asymptotically approaches 2s(Ne/N) (for small s), where s is the relative selective advantage of the mutant allele, and Ne and N respectively denote the effective and absolute (haploid) population sizes. In contrast, in sufficiently small population, fixation probability is 1/N, with the two domains of behavior being separated at the point where s is approximately 1/(2Ne). These results hold when s, N, and Ne remain constant during the fixation process, but require modification when populations are changing in size. Here, we show that if there are ecological effects associated with competing alleles, such that the genetic composition of the population influences the total population size, the fixation probability of a beneficial allele can substantially deviate from the levels suggested above, even in the domain of effective neutrality (i.e., $|s|<1/(2Ne)$). This suggests that, in certain ecological contexts, the rate of molecular evolution might be much greater than previously thought possible.

203W Mutation load at active tRNA genes in three nematode species

Avery Davis Bell[1], Corinne Simonti[1,2], Joseph Lachance[1], Annalise B Paaby[1] [1]School of Biological Sciences, Georgia Institute of Technology, [2]AAAS

Transfer RNAs (tRNAs) are ancient and essential components of protein synthesis. They are encoded by hundreds of genes in nuclear genomes and experience extremely high mutation rates, which has been hypothesized to impose mutation load on eukaryotic genomes. To investigate this hypothesis, we determined tRNA allelic variation in natural populations of three nematode species: the model organism *Caenorhabditis elegans* and its ~10-25 My diverged relatives *C. briggsae* and *C. tropicalis*. We find that tRNA genes carry strong signatures of both transcription-associated mutagenesis and purifying selection, and exhibit alleles predicted to disrupt function—including mismatches between tRNA anticodon and backbone. We next assessed how these putatively deleterious mutations were distributed among tRNA genes stratified by activity, a proxy for their functional demand. To do so, we ranked tRNAs by activity using ATAC-seq and PolIII ChIP-seq data from *C. elegans,* as chromatin accessibility and transcription strongly correlate with tRNA activity. Mutations occur in tRNA genes with expression levels across the full activity spectrum, suggesting that allelic variation does impact essential tRNAs. We also find that some tRNAs are uniquely expressed at specific developmental stages, raising questions about universal *vs* stage-specific tRNA requirements and about when transcription-associated mutagenesis induces heritable changes to the germline that shape the standing variation in tRNA pools in natural populations. Our results suggest that natural nematode populations bear mutation load at tRNAs; they also identify specific tRNA genes and developmental windows for experimental investigation into how allelic variation at cytosolic tRNAs affects fitness.

204W Harnessing citizen science to contextualize adaptation mechanism discovery

Xianran Li USDA-ARS

Species occupying broad geographic regions have evolved multiple mechanisms to regulate phenological characteristics, enabling adaptations to diverse native habitats. A main hurdle to explore the adaptation mechanisms is profiling phenology traits from native habitats. By developing computer vision AI to process citizen science observations across native habitats over North America, we uncovered a consistent latitudinal trend of earlier flowering at higher latitudes in warm-season perennial grasses. To explore the underlying adaptation mechanisms, we conducted common garden experiments with one species (switchgrass), and surprisingly, discovered the opposite latitudinal flowering-time trend. Integration of differential plasticity of haplotypes of flowering time regulatory genes, haplotype ranges, and local environmental profiles found that observations from native habitats capture only part of the genotype-environment-phenotype spectrum established in common garden experiments, therefore, reconciled the discrepancy. Two adaptive mechanisms emerged as key forces shaping the current haplotype ranges and influencing future shifts. Our study highlights the power of combining citizen science data with controlled experiments to uncover adaptation mechanisms across spatiotemporal scales.

205W Increased male frequency in *Caenorhabditis briggsae* intra-species hybrids occurs by genetic incompatibility that produces X nondisjunction

Karla Reyes Barajas, Kyle Asa Lee, Alyssa Stone, Srishti Kamboj, Joseph Ross Biology, California State University, Fresno

Postzygotic reproductive isolation barriers can arise from genetic divergence between two populations. Such Dobzhansky-Muller interactions involve alleles in different genes that do not function properly together in the context of a hybrid genetic background. Because the segregation of sister chromatids during gametogenesis is critical to viability and fertility, and because this process is regulated by many protein complexes, hybrid genetic incompatibilities involving meiosis-related genes might cause decreased fitness and speciation. The natural frequency of males in androdioecious species of *Caenorhabditis* (e.g. *C. elegans*, which has two sexes: males and self-fertile hermaphrodites) is low compared to the 1:1 sex ratio that typifies gonochoric species (having males and females). *Caenorhabditis* males are genetically XO, and the production of males is caused by X chromosome nondisjunction. Thus, a proxy for meiotic nondisjunction frequency is male frequency, which is often reported as ~1:1,000 individuals (0.1%). *C. briggsae* populations are useful for investigating genetic incompatibilities, particularly because of the structure and magnitude of their genetic diversity compared to *C. elegans*. We describe a genetic incompatibility where intra-species crosses between some populations produce F1 hybrids with increased meiotic nondisjunction rates, measured by an increase in male F2 offspring. Nonhybrid (control) crosses produce 0.1–0.6% males, and hybrids produce as high as 7% males. Cross direction effects suggest that a maternal effect might be involved. Weak correlation between F2 male frequency and F2 embryonic lethality implies that the incompatibility might specifically affect X and not autosomal nondisjunction. Mapping of epistatic loci in advanced-intercross recombinant inbred lines is underway, and future work will explore how epistatic interactions regulating the meiotic segregation of specific chromosomes can evolve. In *C. briggsae*, X nondisjunction is tolerated because of dosage compensation and thus this effect does not directly decrease hybrid fitness. In other species, where such interactions reduce hybrid fitness, evolution of meiotic machinery genes might result in speciation.

206W Can the Z:A ratio serve as a genomic index of sexual selection strength?

H. Luke Anderson[1,2,3], Kevin F. P. Bennett[2,4,5], Kira M Long[6,7], Aaron O'Dea[8], Jeffrey D Brawn[7], Thomas J Parsons[9], Philip L. F. Johnson[5], Michael J Braun[2,5] [1]Lab of Ornithology, Cornell University, [2]Vertebrate Zoology, Smithsonian National Museum of Natural History, [3]Ecology & Evolutionary Biology, Tulane University, [4]Biology, Penn State University, [5]Biology, University of Maryland—College Park, [6]Conservation Biology Institute, Smithsonian National Zoo, [7]Program in Ecology, Evolution, and Conservation Biology, University of Illinois Urbana-Champaign, [8]Smithsonian Tropical Research Institute, [9]Center for Innovations in Health Sciences, VinUniversity

Understanding sexual selection's role in shaping evolutionary trajectories is a key goal of biology, and thus developing a population genomic measure of sexual selection strength is desirable. In systems with ZW sex determination, one promising index is the Z:A ratio—the ratio of nucleotide diversity on the male-biased Z chromosome relative to the autosomal average—which is expected to decline under male-biased sexual selection due to reductions in male effective population size. However, other non-equilibrium processes can also influence Z-linked diversity, and studies in birds have yielded conflicting results about the degree to which Z:A ratios correlate with traditional proxies for sexual selection strength. As such, there is a need to empirically assess how the Z:A ratio varies under relatively well-understood demographic and selective circumstances to inform the situations in which it may serve as a valuable measure of sexual selection strength *per se*. Here, we leveraged the well-studied *Manacus* system in western Panama, which comprises a series of mainland populations (including parental *M. candei*, parental *M. vitellinus*, and their hybrid zone) as well as insular populations on the Bocas del Toro Archipelago. We found that the Z:A ratio was generally stable over time and space among mainland transect populations, with key exceptions: (1) two populations that experienced introgression of a sexually selected plumage trait between historical and contemporary sampling timepoints exhibited small but significant increases in Z:A ratios between timepoints; (2) a population located on the Tierra Oscura peninsula had an unusually low Z:A ratio relative to other mainland populations; and (3) there was high variation in individual-level Z:A ratios in the hybrid zone, likely due to fast-Z effects and recent admixture. Moreover, island populations exhibited striking reductions in Z:A ratios relative to mainland populations, which corroborates theoretical predictions that sex-linked nucleotide diversity should be disproportionately affected during population contraction events. The general stability of the Z:A ratio among mainland populations suggests its potential to serve as a yardstick of sexual selection when confounding processes are known or controlled. However, these results also demonstrate the strong effects of demographic history on Z-linked diversity, which should be considered in studies aiming to use the Z:A ratio as an index of sexual selection.

207W Modes of natural selection on maternal and zygotic gene expression in *Drosophila melanogaster* embryos

Thomas S. O'Leary[1], Brent L. Lockwood[2] [1]University of Vermont, [2]Biology, University of Vermont

Early embryonic development involves the coordination of gene expression from two distinct genomes, as mothers load eggs with gene products prior to the beginning of zygotic transcription. Thus, genetic variation for maternal transcription may segregate independently from that of zygotic transcription and be exposed to distinct selection pressures, despite existing in the same organism. To infer modes of selection on gene expression at maternal and zygotic developmental stages, we used single-embryo RNA-sequencing and conducted variance tests of selection on transcriptomes of parents and F_2 segregants of tropical and temperate *Drosophila melanogaster* embryos. We found that approximately 5–10% of genes showed signals of selection, with directional selection more common than stabilizing selection, and more genes with signals of selection in maternal transcripts than zygotic transcripts. Among core early developmental genes and transcription factors, maternal transcription showed patterns of both stabilizing and directional selection, whereas zygotic transcription was predominantly under stabilizing selection. Many heat shock genes showed patterns of directional selection between tropical and temperate embryos, consistent with local adaptation. Additionally, directional selection in piRNA-pathway genes suggests a role for germline defense during embryogenesis. Together, these results demonstrate that selective regimes shape developmental genetics over relatively short evolutionary timescales.

208W *New insights into the nature of genetic variation revealed by highly-resolved longitudinal studies*

Mark C. Bitter[1,2], Paul Schmidt[3], Dmitri Petrov[2] [1]Department of Biology, Stanford University, [2]Stanford University, [3]Department of Biology, University of Pennsylvania

Arguably the fundamental pursuit of population genetics research is to understand the nature of, and dominant evolutionary forces acting upon, genetic variation. Since the fields inception roughly 100 years ago, the primary inferential approach to this pursuit has been through the analysis of static snapshots of contemporary sequence data. The prevailing view that emerged from these inferences, and from which the standard null model of population genetics was constructed, posits that genetic variation is dominated by polymorphisms of neutral or nearly neutral fitness effect and genetic drift to be the primary evolutionary force. In my poster, I will discuss results gleaned from longitudinal assays of genetic variation that have challenged this paradigm. Specifically, I will focus on our findings from mesocosm experiments in large, genetically-diverse populations of *Drosophila melanogaster* in which we have directly quantified selection on individual loci throughout the evolutionary process itself and on ecological timescales. Across a range of experiments, we now have strong evidence that natural populations harbor hundreds of polymorphisms with strong effects on fitness that are actively maintained in populations, at least in part, by temporally fluctuating selection. I will highlight how these dynamics are likely driven by underlying fitness trade-offs and eco-evolutionary feedbacks. I will discuss the potential ubiquity of these patterns across taxa and propose ways forward in the development of more appropriate empirical studies and null models for evolutionary inference.

209W Evolutionary and Functional Conservation of Topologically Associating Domain Boundaries in Rice Genomes

Owen D Geyman[1,2], Michael Purugganan[1], Amina Kurbidaeva[1] [1]Biology Department, New York University, [2]Avenues: New York

Traditional evolutionary analyses have relied primarily on two-dimensional (2D) sequencing data; recent advances in three-dimensional (3D) genomics allow for deeper investigation of how genome architecture shapes evolutionary divergence. In this study, we analyze conserved and species-specific topologically associating domain (TAD) boundaries across multiple Oryza species to uncover both structural and functional evolutionary insights. Here, we characterize boundary conservation at numerous levels, from highly conserved to species-unique. We apply TCBF, a relatively novel and underutilized tool for TAD boundary conservation analysis, alongside a suite of complementary analyses, including gene ontology, RNA-seq, sequence conservation, and epigenetic profiles. Our approach not only validates canonical evolutionary relationships reflected in sequence data but also novel patterns unique to 3D genome organization. Most notably, our study uncovers strong associations between highly conserved TAD boundaries and RNA-processing–related genes, as well as other genetic and epigenetic profiles, and a contrasting absence in minimally conserved boundaries, highlighting strong functional relevance. Our findings also connote evolutionary and phylogenetic trajectories between common rice cultivars reframing empirically supported genomic relationships. Overall, our work highlights a bioinformatic framework that integrates 3D genome structure into evolutionary genomics, addressing a critical and underexplored dimension of comparative plant genomics while underscoring evolutionary novelty discernible only through three-dimensional genome architecture.

210W The evolutionary and functional impact of the human-specific acyl-CoA thioesterase (*ACOT1*) gene duplication

Carsyn Bonesteel, Omer Gokcumen, Charikleia Karageorgiou University at Buffalo

ACOT1 plays an important role in lipid metabolism and has been associated with physiological traits including liver function, fat mass, diabetes, and breast milk composition. Although previous studies have reported signatures of selection in *ACOT1* in Melanesians, the evolutionary origin and genomic complexity of *ACOT1* have not been fully characterized. Here, we investigate the genomic, evolutionary, and functional impact of the *ACOT1* gene. Using long-read sequencing from 326 individuals together with short-read sequencing data from unrelated individuals in the 1000 Genomes Project, we confirm that cytosolic *ACOT1* is a human-specific gene that arose from duplication of the mitochondrially localized paralog *ACOT2*. We observe extensive copy number variation and reconstruct eight distinct structural haplotypes at the *ACOT* locus, with pronounced population-specific distributions, including a significant reduction in *ACOT1* frequency in Southeast Asian populations. Analysis of all available high-coverage archaic genomes reveals that the *ACOT1* duplication is absent in Neanderthals and Denisovans, indicating that it arose in anatomically modern humans. Leveraging over 500 ancient genomes, we further reconstruct the spatiotemporal history of the locus and show that the *ACOT1* duplication predates the major out-of-Africa dispersal, placing its origin earlier than 50,000 years ago. We also identify complex haplotypes involving subsequent duplications and deletions of *ACOT1* and *ACOT2*, including tandem duplications spanning both genes. Building on these evolutionary insights, we will integrate GTEx tissue-specific genotype and expression data to link haplotype structure and *ACOT1* copy number to gene expression variation across relevant tissues. We further aim to apply haplotype-aware GWAS approaches to test associations between *ACOT1* copy number and metabolic traits while reducing spurious signals at this structurally complex locus. Future work will examine how *ACOT1* copy number influences lipid composition in human cell lines exposed to defined environmental stressors. Together, these findings provide a framework for understanding gene-environment interactions in recent human evolution and may inform more personalized approaches to studying metabolic disease.

211W The genetic basis of intrinsic postzygotic reproductive isolation in necrotic Mimulus

Pia Franziska Schwarz, Jenn M Coughlan Ecology & Evolutionary Biology, Yale University

Hybridization is a dynamic process that can lead to diverse genomic and phenotypic outcomes in hybrid offspring. Epistatic interactions between novel allele combinations in hybrids result in lowered hybrid fitness and gene-flow barriers in the form of Bateson-Dobzhansky-Muller incompatibility. Here we explore the genetic basis of necrosis and dwarfism in F2 hybrids of Mimulus guttatus and Mimulus decorus. We hypothesize that antagonistic evolution between plant immune system and pathogens leads to the evolution of polymorphic alleles of plant immune genes. We analyze the quantitative trait loci (QTL) to detect genetic markers segregating with the necrosis and dwarfism trait values, and evaluate the role of the plant immune system in hybrid necrosis. We aim to identify potential candidate genes associated with these traits, and how selective pressures on immune adaptations in Mimulus species and populations contribute to post-zygotic reproductive barriers. Hybrid necrosis and dwarfism can serve as a model to gain insight into the mechanism guiding the evolution of incompatibility alleles in the context of rapid diversification of plant disease resistance genes in a genetically diverse species complex.

212W Population Genetics of Source–Sink Dynamics

Ziyang Xia[1], Andrew H Vaughn[2,3], Philipp W Messer[1] [1]Department of Computational Biology, Cornell University, [2]Bakar Computational Health Sciences Institute, University of California San Francisco, [3]Center for Computational Biology, University of California Berkeley

Organisms typically inhabit environments that are at least partly heterogeneous, for example due to local differences in habitat quality or environmental conditions. As a result, some parts of a population may systematically produce more descendants than others, generating a net flux of individuals from these regions into the rest of the population. However, the effects of such so-called source-sink dynamics are rarely incorporated into population genetic models. Here, we use mathematical models and simulations to study the impact of source-sink dynamics on coalescence times and patterns of genetic diversity. We find that even a small source with modest net outflux into a consistently large population can dramatically reduce coalescence times and levels of genetic diversity, while variance and inbreeding effective population sizes can remain close to the population's census size. When demographic inference is applied to such a population, the inferred history resembles a strong recent expansion. Nevertheless, these dynamics leave characteristic signatures of tree imbalance in coalescent genealogies that allow source-sink dynamics to be distinguished from true population expansions. We further extend our analyses to scenarios with multiple sources and to non-stationary settings in which sources exist only temporarily. Finally, we show how unequal resource distribution in a spatial population can naturally give rise to source-sink dynamics. Interestingly, lineages sampled close to the source in our spatial model typically exhibit denser branching histories and shallower local divergence, contrary to the common assumption that ancestral regions should be enriched for basal, early-diverging lineages. Our study provides an alternative and potentially widespread mechanism for reducing a species' effective population size and confounding demographic inference without requiring selection or changes in census size over time, while also generating spatial genealogical patterns that may depart from classical biogeographic assumptions.

213W Analytical expectations for ancestry junction accumulation in admixed genomes

Shirin Nataneli[1], Aydin L Karatas[1], Roshni A Patel[2], Tessa Ferrari[1], Jazlyn A Mooney[1] [1]Quantitative and Computational Biology, University of Southern California, [2]Data Science, University of Oregon

Complex demographic events have shaped human history, leaving signatures of genetic variation across the genome. Here, we investigate the recent evolutionary history of admixed populations formed from multiple ancestral sources. We present a discrete, generalizable model of admixture that leverages ancestry switches, which are recombination breakpoints that mark changes in ancestral origin along a chromosome. We derive analytical expectations for the number of ancestry switches within a genomic segment as functions of recombination rate, ancestry heterozygosity, and effective population size, and extend these expectations to incorporate population-specific recombination maps. Forward-time simulations tracing ancestry junctions for ten generations after admixture show close agreement with theoretical predictions under constant and variable recombination models. We observe minimal variability in switch counts across ten replicates, underscoring the robustness of the theoretical expectation. Furthermore, model-based switch parameterized using literature-informed demographic values counts agree with empirical observations from African American individuals in the 1000 Genomes Project. For example, when modeling human chromosome 1, we found a mean of approximately six switches per haplotype, which aligns with the theoretical expectation under an initial African ancestry proportion of 0.85, and agrees with published estimates from other African-American cohorts. Overall, the model provides a new route for using ancestry switches to study how recombination and demography jointly shape ancestry patterns in admixed populations without requiring separation into parental sources.

214W Simple and complex: mapping the genetic basis of insecticide resistance across genotypes and environments to predict resistance evolution

Amir Gabidulin, Rene Shahmohamadloo, Seth Rudman School of Biological Sciences, Washington State University

High-resolution genetic mapping has greatly improved our ability to identify variants underlying trait variation. Can these data be translated into accurate forecasts of the genomic basis of adaptation? First, we tested the repeatability of variant identification by mapping insecticide resistance across multiple genetic backgrounds and ecological contexts in Drosophila. We then combined these results with a field resistance evolution experiment and data from resistant populations sampled from agricultural fields to test the predictability of mapping studies for resistance evolution. Overall, we find patterns of concordance in the genetic basis of resistance across contexts in mapping, but limited translatability to forecasting genomic evolution suggesting pleiotropic constraints may be a key hurdle to using mapping studies for evolutionary prediction.

215W Disentangling Reticulated Evolution: Genomic Signatures of Introgression in Darwin's Finches

Gerardo Cendejas Mendoza[1], Matteo Sebastianelli[2], Leif Andersson[2], Carlos A. Valle[3], Erik Enbody[1] [1]Computational Biology, Cornell University, [2]Uppsala University, [3]Universidad San Francisco de Quito

Studying the extent and persistence of hybridization is fundamental to understanding how adaptive radiations generate and maintain biodiversity. Darwin's finches are a classical model of ecological speciation, but our understanding of their evolutionary history is increasingly shifting from a bifurcating tree to a complex network of reticulate lineages. Long-term studies on the small island of Daphne Major have documented a local system where three sympatric species hybridize; however, because most insights into the radiation come from this single location, it remains unclear how widespread these patterns are across the wider archipelago and thus the broader impact of hybridization on the radiation's history.

In this study, we leverage whole-genome SNP data to reconstruct an updated phylogeny of the *Geospiza* genus and investigate hybridization dynamics using D-statistics (ABBA-BABA) and f_branch statistics. While we survey the entire archipelago, we focus specifically on Genovesa Island, where we uncover a complex history of admixture that challenges simple speciation models. We detect a network of gene flow among three sympatric species on the island (*Geospiza propinqua, G. magnirostris*, and *G. acutirostris*). The study also reveals evidence of long-distance introgression from two external sources.We observe signals of introgression (f_branch) between the Pinta Island Cactus Finch (*Geospiza scandens*) and *G. propinqua* on Genovesa, which may contribute to the distinct large, pointed beak of the Genovesa population. We also detect significant introgression from Española Island (*Geospiza conirostris*) (located roughly 180 km away) into *G. propinqua* and *G. magnirostris* on Genovesa. This finding parallels the founding of the hybrid "Big Bird" lineage on Daphne Major by an immigrant from Española, reinforcing the capacity of *G. conirostris* for long-distance dispersal and introgression. Together, these findings support the view that hybridization in this system is more widespread than previously thought, underscoring that dissecting these reticulate histories is essential to understand how gene flow contributes to the rapid speciation and ecological diversification characteristic of this adaptive radiation.

216W Study pollinator syndrome switches through evolution experiment with *Mimulus*

Foen Peng Biology, Haverford College

Flowering plants attracting different pollinators, e.g. birds vs. bees, often have different suites of floral traits, so called pollinator syndrome. Switches between pollinator syndromes are presumably difficult as they require concerted polygenic adaptation, but these switches are frequently documented in nature. To understand how this process could occur, my lab experimentally evolved a hybrid *Mimulus* population with abundant floral variation towards a new pollinator which they have never been exposed to. In the early phase of this experiment, we have witnessed rapid phenotypic and genomic changes.

217W Genetic incompatibilities and compensatory adaptation drive hybrid genome evolution in yeast

Artemiza A Martinez, Gregory I Lang Biological Sciences, Lehigh University

Interspecific hybridization can break down species barriers to reveal interspecific genetic incompatibilities. We generated a panel of recombinant hybrid genomes from a cross between *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*, sister species that diverged ~5 million years ago and which remain only ~85% identical at the amino acid level. Recombinant hybrids are on average less fit than either parent and show broad phenotypic variation. For 628 conserved protein complexes we classified each as hybrid (containing at least one subunit from each species) or uniparental. Inheritance patterns of protein complexes in the hybrid genomes support a model of pervasive negative genetic interactions affecting fitness of hybrid protein complexes. To select for compensatory mutations, we evolved the recombinant hybrids strains for 10,000 generations. We identify a small subset of hybrid protein complexes that accumulate more mutations when hybrid compared to when uniparental. This includes the Ku70/80 complex, which is known to be rapidly evolving in *Saccharomyces* yeasts. Together, our results suggest that weak genetic incompatibilities are widespread and shape fitness and adaptation in hybrid genomes.

218W Microsyntenic Divergence with Conserved Protein Domains of *chico* Across *Drosophila* Subgenera During Species Radiation

Elizabeth Wasson, Logan Cohen, Laura K. Reed The University of Alabama

The insulin/TOR signaling pathway is a well conserved pathway across metazoans and plays a central role in regulating growth and metabolism. Disruption of this pathway can have many wide-ranging physiological consequences that impair metabolic regulation. In *Drosophila*, the gene *chico* is homologous to the vertebrate *insulin receptor substrates* (*IRS*s), which codes for an adaptor protein linking insulin receptors to downstream signaling cascades early in the pathway. Loss of function mutations in *chico* result in reduced body size and growth. Because *chico* is the first gene in this pathway and acts as a highly connected adaptor protein that mediates interactions between the insulin receptor and downstream signaling components, it is expected to be under strong evolutionary constraint at the protein level. As a result, *chico* serves as a useful model for understanding how genes in complex biological networks evolve across species occupying different ecological niches. This allows us to study how quickly protein coding regions are evolving and could the environment have a role in this. Using gene models generated by the Genomics Education Partnership course-based undergraduate research experience (thegep.org), we are working towards a multi-ortholog analysis of chico across 32 species of *Drosophila*. We examined *chico*'s genomic microsynteny (local neighborhood) and domain structure and our results show that though *chico*'s microsynteny shifts between clades, protein domain architecture remains conserved across all 32 species. This pattern suggests that while *chico* maintains core functional regions, its genomic neighborhood may be more flexible during evolutionary divergence. We hypothesize that environmental pressures associated with species radiation may contribute to these microsyntenic changes without disrupting protein function. Future work will include quantifying and visualizing microsyntenic rearrangements using RStudio and using Multiple Sequence Alignments in Jalview to visualize evolutionary patterns.

219W A Molecular Evolution Analysis of *Phosphoinositide-dependent-kinase-1 (Pdk1)* Orthologs Across *Drosophila*

Emie K Vandiver, Logan Cohen, Laura K. Reed The University of Alabama

The insulin/TOR signaling pathway is a conserved pathway across many species that is critical for regulating growth, metabolism, and aging. Disruption of this pathway can cause serious biophysiological consequences that influence metabolic syndromes such as diabetes, obesity, and heart disease. The gene *Phosphoinositide-dependent-kinase-1 (Pdk1)* encodes for a protein that serves as a master regulatory kinase in the insulin signaling pathway, apoptotic processes, and cell growth regulation. The *Drosophila* gene *Pdk1* is homologous to the *3-phosphoinositide dependent protein kinase 1 (PDPK1)* gene in humans. Characterizing the evolution of *Pdk1* and its role in the insulin signaling network in *Drosophila* can give insight into how essential genes evolve within the context of complex biological networks. In this study we used gene models generated by the Genomics Education Partnership course-based undergraduate research experience (thegep.org) to perform a multi-ortholog analysis of *Pdk1* across 32 species of *Drosophila*. A microsynteny (local neighborhood) analysis shows that there are gene insertions, deletions, and possible duplication events in the genomic neighborhood of *Pdk1* across the phylogeny. These events could disrupt regulatory elements and expression levels of genes within the genomic neighborhood. While *Pdk1's* sequence and isoform structure remains well conserved across species, some species groups show evidence of novel isoforms relative to the informant species *Drosophila melanogaster.* Sequence-based domain analysis results display that *Pdk1* protein function is maintained despite syntenic rearrangements and novel isoforms. Future work will include Multiple Sequence Alignments in Jalview to visualize evolutionary patterns within the gene sequence, and utilization of SyntenyPlotteR in RStudio to quantify microsynteny changes.

220W Sex differences in recombination in house fly, *Musca domestica*

Basanta Bista, Yesbol Manat, Carolynn Porter, Richard P Meisel Biology and Biochemistry, University of Houston

Sex-specific recombination rates are a common feature of meiosis across eukaryotes. Sex-limited recombination is an extreme form of sex-specific recombination rates where meiotic recombination only happens in one sex. Sex-limited recombination is especially important for sex chromosome evolution. Recombination is often suppressed in the heterogametic sex (XY males or ZW females), either genome-wide or on the sex chromosomes specifically. Suppressed recombination is thought to stabilize linkage between sex-determining regions and sexually antagonistic alleles, thereby promoting X-Y (or Z-W) divergence and differentiation, along with Y or Z chromosome degeneration. While these general patterns have been observed across many eukaryotic taxa, the evolutionary origins of recombination suppression on nascent sex chromosomes is not well understood. To address this gap, we directly measured sex-specific recombination in the house fly, *Musca domestica*. House fly has multiple nascent “proto-XY” chromosome pairs, allowing us to test how sex-specific recombination suppression evolves during the earliest stages of sex chromosome evolution. We measured recombination in *M. domestica* using low-coverage whole-genome sequencing of offspring from controlled test crosses involving four genetically distinct strains, with males carrying two different proto-XY chromosomes. We developed and implemented a novel and broadly applicable strategy that enables genome-wide detection of recombination from low-coverage whole-genome sequencing data by combining low-confidence progeny genotypes with high-confidence parental genotype calls and aggregating information across many low-confidence variant sites to reconstruct parental haplotype inheritance along the length of chromosomes. Across strains, females exhibited higher average numbers of crossovers than males (females: 0.558 per chromosome, 2.79 per genome; males: 0.098 per chromosome, 0.490 per genome), consistent with strongly reduced recombination in males that is indistinguishable from no recombination. Our results provide evidence that female-limited recombination facilitates the evolutionary formation and subsequent divergence of proto-XY chromosomes.

221W Theory of Epimutation-Selection Balance

Gregory Chernomas[1], Cortland Griswold[2] [1]Integrative Biology, University of Guelph, [2]University of Guelph

Epigenetic variation can be a source of phenotypic and fitness variation in natural populations. Here I present theory that incorporates epigenetics within the classic framework of mutation-selection balance. For example, a gap in theory was how inbreeding affects deleterious epigenetic variation. Our results align with classic theory; however, departures from classic theory arise due to the unique biology of epialleles, such as the relatively high reversion rate of epialleles and unique epigenetic phenomena such as paramutation. A multi-locus model of deleterious epimutation-selection balance highlights a potentially novel process when a transposon is adjacent to a genetic site because a transposon can influence local epimutation rates. In particular, the cis epigenetic effects of transposons may lead to new mechanisms that can influence the evolution of recombination rates. We find conditions where recombination can be indirectly beneficial or deleterious, based on the effective epimutation rate, which is expected to vary considerably across species due to the variability in epigenetic resetting rates. Therefore, species that vary in epigenetic properties may experience distinct local and possibly global qualitative effects on recombination rate evolution. Overall, we provide fundamental predictions on heritable epigenetic variation viewing it as a deleterious process on average.

222W Integrating competition into quantitative genetics

Yasmine McDonough University of British Columbia

The genetic variance-covariance matrix (G-matrix) plays a central role in predicting the rate and direction of adaptation. The theory underlying these predictions, however, tend to assume that traits are subject solely to stabilising selection to an optimum. In natural populations, populations may also be shaped by interactions between coexisting individuals, such as competition over resources. Here we investigate how incorporating alternative selection pressures such as intraspecific competition may affect the evolutionary dynamics of the G-matrix. Using individual-based simulations, we characterise the G-matrices of populations experiencing stabilising selection and/or intraspecific competition (i.e, negative frequency-dependent selection) on two quantitative traits. We consider adaptation to a stable, gradual, or suddenly moving environment. We show that competition can (1) substantially improve G-matrix stability, including during environmental change and in small populations, (2) increase the rate of adaptation to new optima by reshaping the fitness landscape at the leading edge, and (3) create misalignment between the leading eigenvector and the genetic line of least resistance by obscuring directions most amenable to novel mutations. In total, our results demonstrate that intraspecific competition can alter rates and trajectories of adaptation in ways that cannot be predicted by the G-matrix.

223W Deer mice (*Peromyscus maniculatus*) of California: Investigating evolutionary history, conservation genomics of island populations, and patterns of co-diversification with Sin Nombre hantavirus

Tommy M Herrera[1], Sharon Messenger[2], Lydia Smith[1], Michelle DaVila[3], Chris Conroy[3], Rauri Bowie[3], Erik Enbody[4], Michael Nachman[3] [1]Integrative Biology, University of California Berkeley, [2]Viral Immunoserology and Molecular Diagnostics Section, California Department of Public Health, [3]Museum of Vertebrate Zoology, University of California Berkeley, [4]Computational Biology, Cornell

Deer mice (*Peromyscus maniculatus*) are the most abundant and widely distributed small mammal species in California, inhabiting all major biomes across the mainland as well as the Channel Islands off the southern coast. Their high prevalence makes them significant reservoirs and vectors of disease. Across the western United States, deer mice serve as the primary reservoir for Sin Nombre hantavirus (SNV), the causative agent of hantavirus pulmonary syndrome in humans, which has a case fatality rate of approximately 35%. Using genomic data from the California Conservation Genomics Project (CCGP), we evaluate the genetic structure and evolutionary history of deer mice populations across both the mainland and Channel Islands. This analysis largely corroborates existing regional subspecies designations and elucidates the processes underlying diversification across California's diverse ecogeographic regions. Additionally, we examine the conservation genomics of federally listed Channel Island populations to assess the genomic impacts of their isolation from mainland populations and to compare levels of genetic diversity between island and mainland groups. Finally, by integrating data from the CCGP and the California Department of Public Health, we explore patterns of co-diversification between deer mice and SNV to determine whether strong host-virus interactions have produced a shared evolutionary history across California. These findings offer insights into the evolutionary processes shaping deer mouse populations and their conservation status. Additionally, they advance our understanding of the epidemiological patterns of deer mouse-SNV interactions across the region.

224W Premature hair-greying caused by melanocyte-specific gene regulation in American black bears

Emily E Puckett[1], Christopher B Kaelin[2], Greg S Barsh[2] [1]Biological Sciences, University of Memphis, [2]Genetics, Stanford University

Hair greying is an apparent and early signal of aging in humans. Superficially, this is due to the loss of melanin from hair follicles, although there may be multiple causative mechanisms. The grey color morph of the American black bear is colloquially known as the Glacier bear and presents the phenotype by 6 months of age. Microscopic investigation of hairs plucked from glacier animals reveals the phenotype is due to the complete depigmentation of a subset of individual hairs across the body. We used a genome-wide association on 10 glacier and 82 non-glacier bears from the Yakutat population and identified a single major locus ($P = 3.92 \times 10^{-14}$) that was 829kb. The association interval contains three genes including *BCL2*, a strong candidate given the grey phenotype observed in *bcl-2* knockout mice. We investigated the cis regulatory landscape, and the transcriptional effect of 243 SNPs and 51 indels identified at high frequency within the association interval of glacier bear using a massively parallel reporter assay (MPRA). We identified an upstream variant in a TFEB transcription factor binding site that is a strong candidate for a causal genetic variant. An ancestral recombination graph around the association peak estimated the TMRCA of the glacier haplotype to be 7kya, which contributes to the low frequency (<1%) of the phenotype across the landscape.

225W Massively parallel evolution reveals a scaling law between cell size and cell density

Alexander B Sastokas[1], Wei-Chin Ho[2], Kara Schmidlin[1], Parker Crossland[1], Peter Brown[3], Douglas Shepherd[3], Michael Lynch[1], Kerry Geiler-Samerotte[1] [1]Center for Mechanisms of Evolution, Arizona State University, [2]Biology, University of Texas at Tyler, [3]Center for Biological Physics and Department of Physics, Arizona State University

Scaling relationships are useful in predicting evolutionary outcomes. Cell size often scales with other phenotypes and varies a lot across the tree of life, thus understanding how it affects other phenotypes is of great interest. Here, we investigate phenotypes that co-vary with cell size by evolving approximately 300,000 barcoded Saccharomyces cerevisiae lineages under selection for increased cell size. Our evolution experiment revealed a consistent scaling relationship in which intracellular density decreased as cell size increased. Using whole genome sequencing, we uncovered numerous unique mutations specific to our selection, suggesting that increasing cell size is not restricted to a small number of genetic pathways. Yet no matter the genetic basis for the increase in cell size, we see the characteristic decrease in intracellular density. To further test the generality of this scaling relationship, we analyzed a publicly available dataset spanning 68 species, which recapitulated the negative relationship between cell size and density. This suggested that the scaling relationship we observed from our experimental evolution could persist across the tree of life. Together, our findings suggest fundamental constraints that may shape evolutionary trajectories across diverse taxa.

226W Molecular and In Silico Characterization of Xylanase Enzyme in *Sordaria fimicola*

Muhammad Saleem[1], Maimoona Ilyas[2] [1]University of the Punjab, [2]Botany, Women University Mardan, Pakistan

Xylanase is a key hydrolytic enzyme encoded by the Endo-1,4-beta xylanase gene, which plays a role in breaking down β1,4-glycosidic linkages in xylan. This study aimed to identify and analyze the Endo-1,4-beta xylanase gene in six strains of Sordaria fimicola. Molecular, genetic, and in silico analysis of the gene from S. fimicola collected from different north and south slopes of Evolution Canyon valley, a site known for divergent evolution, revealed polymorphisms at the nucleotide level that led to changes in amino acids, affecting various post-translational modifications (PTMs) and protein functions. A total of 87 polymorphic sites were identified in the nucleotide alignment, resulting in changes to 87 amino acids across the 31 kDa domain. This suggests that the gene is part of the jumping gene family. However, all SFS strains showed more variation, especially S3 with 69 polymorphic sites, compared to the NFS strains, which were more conserved. PTMs such as phosphorylation and glycosylation were predicted for serine, threonine, tyrosine, and asparagine residues. The study reports 28S, 19T, and 11Y phosphorylation sites in SFS strains and 22S, 16T, and 9Y sites in NFS strains for Endo-1,4-beta-xylanase. Glycosylation sites were found on 16 and 15 serine/threonine residues in SFS and NFS strains, respectively. Notably, S3 exhibited a higher number of potential phosphorylated and glycosylated sites compared to other S. fimicola strains. The presence of the Endo-1,4-beta xylanase gene confirms that S. fimicola is capable of producing xylanase. Stressful environmental conditions on the South Side induced genetic variations, leading to new combinations at this locus and across the Sordaria genome, which may enhance xylanase production and aid species survival. PTMs generate protein diversity that is challenging to quantify.

Key Words: Endo-1,4-beta xylanase gene, Post-translational Modifications (PTMs), Xylanase genes, Evolution Canyon, and xylanases.

227W Adaptive Dynamics of Repeated Local Adaptation in *Amaranthus*

Elizabeth S Polston, Julia Kreiner University of Chicago

Whether the availability of adaptive variation limits the rate of adaptation across natural landscapes is not well understood. Here, we look to resolve this question by quantifying the role of standing variation, gene flow, and *de novo* origins of genome-wide variation underlying adaptation to agriculture. *Amaranthus tuberculatus*, a now problematic agricultural weed, is an ideal system for investigating these dynamics, as recent shifts in the use and management of its native habitat are likely to have reshaped genome-wide diversity. Past investigations in the species have found repeated phenotypic and genetic differentiation between nearby agricultural and natural habitats consistent with local adaptation. With expanded sequencing efforts, totalling to 440 individuals from 17 agricultural and natural paired sites (34 populations), we have resolved 865 independent loci across all 16 chromosomes that are consistently diverged allele frequencies (via Cochran-Mantel-Haenszel), while differences in extended haplotype homozygosity identifies many more. Across these putatively agriculturally-adapted loci, we quantify the frequency of hard, soft, and partial selective sweeps with haplotype and gene tree inference at local and regional scales. Further, we determine how allele age and repeated origins contribute to these patterns with explicit models of convergent evolution. If we can understand how the origin and spread of adaptive variation underlie repeatability and the ability to rapidly respond to environmental change, we will enhance evolutionary-informed management of invasive and at-risk species, benefiting both our economy and biodiversity under climate change.

228W Spontaneous un-selected mutation patterns in *Candida albicans* reveal the relative stability of alternative ploidy states

Kevin Bao[1,2], Dannie Yin[3], Andrew Wu[4], Jiaxin Xie[3], Jacob Fredette-Roman[3], Nathaniel Sharp[3] [1]Princeton International School of Mathematics and Science, [2]University of Wisconsin-Madison, [3]Genetics, University of Wisconsin-Madison, [4]Massachusetts Institute of Technology

Spontaneous mutation generates not only allelic variation throughout the genome, but also variation in the architecture of the genome itself. The opportunistic human pathogen *Candida albicans* is a budding yeast with the potential to grow and evolve in several ploidy states, providing diverse avenues for adaptation. We characterized spontaneous rates of un-selected genetic change in this organism, both in terms of transitions between ploidy states and the generation of aneuploid karyotypes, as well as genome-wide rates of point mutation in each cell type. We conducted mutation accumulation in 160 lines derived from the SC5314 genetic background, with over 700 generations per line. We detected over 1500 point mutations, numerous karyotypic changes, and loss-of-heterozygosity events. Diploids showed point mutation rates similar to other yeast species, and few karyotypic changes. In tetraploids, which can be formed as part of a parasexual process of mating between diploids, we found relatively high point mutation rates, with a shift towards deletions, and frequent karyotypic changes. It is hypothesized that tetraploids undergo chromosome loss to return to diploidy without meiosis, but we find that a third of karyotypic changes in this cell type involved increases in chromosome copy number, and that the return to diploidy is a random mutational process. In haploids, which are believed to be a particularly unstable cell type, we observed frequent diploidization events, with the resulting diploids exhibiting aneuploidy and higher point mutation rates than their haploid progenitors. Mutations were generally deleterious in diploids, but were more likely to be beneficial in the other cell types. We find no evidence that spontaneous karyotypic variation in tetraploids provides resistance to the antifungal drug fluconazole. Our study reveals how the mutation process both responds and contributes to genomic variation in a highly evolvable human pathogen.

229W Comparative single-cell analysis of transcriptional bursting reveals the role of genome organization in de novo transcript origination

UnJin Lee, Cong Li, Chris Langer, Nico Svetec, Li Zhao Rockefeller University

Unlike duplicate gene evolution, the evolutionary and biochemical mechanisms underlying de novo gene birth remain poorly understood. Analysis of the earliest stages of de novo gene origination, non-coding RNAs, reveals that their promoters appear significantly closer to existing protein-coding genes than expected by chance in Drosophila. Recent population genetic analyses have shown that these observations are inconsistent with models of enhancer-facilitated de novo gene birth since 1) low frequency de novo appear to arise relatively randomly while 2) older, higher frequency de novo genes tend to be retained if they arose near existing promoters. To further investigate the earliest steps of de novo gene birth, we performed a comparative scRNA-seq study of testis tissue in Drosophila. In order to identify homologous cell types within our multi-species testis scRNA-Seq data, we developed a novel selection criteria that allowed for the transfer of cell-type labels when transcriptomic divergence occurs simultaneously in both a species- and cell type-specific manner. Further, we demonstrate how expression patterns of key cell type markers, e.g. RBP4, can diverge rapidly between closely related species like D. melanogaster and D. ananassae (div. ~25My), challenging the commonly held, but rarely confirmed, assumption that transcriptomic function in D. melanogaster testis tissue is conserved across species. Using the resulting cross-species scRNA-Seq data set, we then show that the activation of de novo transcripts significantly increases the transcriptional burst size, but not frequency, of neighboring genes during the meiotic transition. As burst size is a kinetic property related to promoter strength, while burst frequency is related to enhancer activity, this result suggests that the act of transcription itself, not sequence-based, enhancer-mediated cis- or trans- regulatory activity, drives the retention of these new promoters. Additionally, we find that this increase of burst size is restricted only to immediate neighboring genes, suggesting that retention of these genes is driven by the interaction between active transcription of both the de novo transcript and its pre-existing neighboring gene. Such interaction is consistent with under-appreciated biophysical mechanisms of short-range transcriptional regulation, e.g. supercoiling mediated coupling, highlighting how genome architecture plays a key role in driving the origination of de novo genes

230W Rare intronic variants drive high variance in human ancestral fitness under a high deleterious mutation rate

Ulises Hernandez[1], Walid Mawass[2], Joseph Matheson[3], Gavin Huttley[4], Joanna Masel[1] [1]Ecology and Evolutionary Biology, University of Arizona, [2]Human Genetics program The University of Chicago, [3]University of California San Diego, [4]Australian National University

There has been fierce debate regarding the importance of rare genetic variants to disease. We construct a model of deleterious load, in ancestral human environments, from first principles. At demographic equilibrium, deleterious variants segregate according to mutation-selection-drift balance and impose a genetic “load” on individuals. Variation in load within a population depends on the whole genome deleterious mutation rate U_d, the distribution of fitness effects (DFE), and the effective population size. Human U_d is estimated as $U_{obs}=2\mu G f_d$, based on the degree f_d to which functional regions diverge more slowly than repetitive intergenic regions, the point mutation rate μ, and the functional genome size G. We fit non-equilibrium nucleotide substitution models to infer f_d across different classes of sites. Preliminary results suggest that human $U_{obs}>4$, higher than previous estimates. Surprisingly, 84% of human U_{obs} corresponds to introns. Using a whole genome DFE to correct for the degree to which weakly selected mutations fail to impact divergence rates, we find that $U_d \sim 1.7 U_{obs}$ under the conservative assumption that all deleterious mutations have a similar DFE to that of ultra-conserved non coding mutations, and $U_d \sim 1.38 U_{obs}$ even when all deleterious mutations share the DFE of non-synonymous mutations and hence of somewhat higher mean effect size. Our model, under a range of $U_{obs}=2.5\text{-}9.7$, implies an ancestral fitness difference between a 75th quartile and a 25th quartile human of 18-39%, or 12-26%, respectively in these two scenarios. A large Q75-Q25 gap is robust to reasonable levels of global epistasis. We propose that variation in load that is this high has broad health implications. Variation in load is higher in populations with low effective population size, making our results also relevant to historically threatened species. From the total variation in load, 83% comes from a previously unappreciated source: weakly deleterious variants at extremely high allele frequencies. These deleterious mutations fixed, and then experienced back mutations at a low rate . Because of their low frequencies and weak selection coefficients, the dynamics of these back beneficial mutations are dominated by drift. Our results suggest that GWAS are under-powered to detect most genetic variation in diseases associated with load.

231W Evolutionary genomics of *Paramecium* mitochondrial genomes

Ravina Shyam Telkar, Farhan Ali Arizona State University

Mitochondrial genomes are central to cellular metabolism and are widely used to infer evolutionary history, yet their population genetic properties remain poorly characterized outside of plants and animals. Unicellular eukaryotes represent a critical but understudied component of eukaryotic diversity, offering unique opportunities to examine how evolutionary forces operate during major evolutionary transitions. Here, we investigate the evolutionary genomics of *Paramecium* mitochondrial genomes using 315 globally sampled isolates spanning multiple species.

Using high-quality mitochondrial genome assemblies, we reconstructed protein-based phylogenies, characterized gene content and gene order variation, and quantified patterns of genetic diversity and selection. Phylogenomic analyses revealed the presence of previously unrecognized cryptic lineages, expanding the number of described *Paramecium* species to 22. Silent-site nucleotide diversity (π) at four-fold degenerate sites was used to infer effective population sizes within species, while synonymous divergence (dS) and gene-wise dN/dS ratios were estimated to assess evolutionary constraint. Core electron transport genes exhibit strong purifying selection, whereas ciliate-specific mitochondrial genes show elevated evolutionary rates consistent with relaxed constraint or functional divergence.

Despite extensive geographic sampling, we observe little correspondence between geographic distance and genetic divergence, providing no evidence for isolation by distance. Closely related mitochondrial haplotypes are often separated by large geographic distances, supporting the hypothesis that high dispersal decouples spatial and genetic structure in this system. Additionally, we find no convincing evidence for recombination in *Paramecium* mitochondrial genomes, indicating that mutation, selection, and genetic drift are the primary forces shaping mitochondrial evolution.

Together, these results support the “everything is everywhere” paradigm for unicellular eukaryotes and establish *Paramecium* as a powerful model for mitochondrial population genomics and evolutionary dynamics.

232W Sex, DNA Repair, and the germline mutational landscape in *Drosophila melanogaster*

Rob Melde[1], Nathaniel Sharp[2] [1]Integrative Biology, University of Wisconsin - Madison, [2]University of Wisconsin - Madison

Mutation is the ultimate source of genetic variation upon which selection acts, yet the processes underlying mutation are not fully understood. Mutation rates and spectra vary across biological scales, from between species to within single genomes, and nearly a century ago, Haldane proposed that sex itself may shape mutational processes. While a male mutation bias is now well-documented across amniotes, the underlying mechanisms remain elusive. Here we aim to test the hypothesis that sex-specific differences in DNA repair contribute to sex-biased mutation patterns. To do this, we employed a modified mutation accumulation (MA) design in *D. melanogaster*, serially passaging chromosomes through either the paternal or maternal germline for 30 generations prior to whole genome sequencing. We exposed a subset of lines to methyl methanesulfonate (MMS) to elevate DNA damage levels in both sexes, enabling direct comparison of sex-specific mutation rates and spectra under baseline and stressed conditions. In parallel, we conducted a complementary MA study examining how endogenous meiotic double-strand breaks influence spontaneous mutation. Together, these experiments reveal how DNA damage, repair, and sex interact to shape the mutational landscape of the germline in *D. melanogaster*.

233W Investigating the effect of mild heat stress on pollen viability in multiple species across a variety of scales

Derek Denney[1,2], Emily B Josephs[3], David B. Lowry[2,3] [1]Michigan State University, [2]Plant Resilience Institute, Michigan State University, [3]Plant Biology, Michigan State University

Natural plant populations are threatened by the rising global temperatures associated with climate change. These elevated temperatures disrupt the critical components of reproduction at a lower threshold than photosynthesis, growth, and survival. While crop scientists have begun focusing on the vulnerability of heat stress on reproduction, evolutionary and ecological studies often focus on other metrics to estimate fitness and may overestimate a population's ability to survive under warming temperatures. We advocate for integrating pollen and ovule developmental metrics into ecological and evolutionary studies to improve predictions for plant population dynamics under future climate scenarios. Additionally, we present two studies investigating the dynamics of warming temperature on pollen viability using monkey flower sourced from along an elevational gradient in the Sierra Nevada range and shepherd's purse collected from across the span of the United States. These studies illustrate that natural populations may respond differentially to elevated temperatures. Studies such as these will increase our understanding of how natural populations will respond to increasing temperature stress and are likely to reveal novel mechanistic insights that can be utilized to improve crop resilience in a warming world.

234W Genomic and transcriptomic signatures of proto-neo-Y chromosome evolution in *Drosophila americana*

Zethus Avery, Anne Scuderi, Yasir Ahmed-Braimah Syracuse University

Degenerate sex chromosomes, like the mammalian and *Drosophila* Y, are theorized to evolve from ancestral autosomes through recombination arrest followed by gene loss and degeneration. Multiple theories have been proposed about the progression of this degenerative process, and we test these hypotheses using *Drosophila americana* as a model. *D. americana* exhibits a latitudinal cline for a centromeric fusion of the 4th and X chromosome, ranging from a 100% frequency of the fusion in northern latitudes to a 0% frequency in southern ones with a gradual transition in the intermediate latitudes. The unfused 4th chromosome in northern males, inherited with the Y chromosome, is a candidate for a newly emerging sex chromosome that has not yet reached fixation, i.e. a proto-neo Y. We investigated the genomic and transcriptomic changes to the male 4th chromosome in northern populations of *D. americana* relative to the southern populations by constructing a species pangenome with three complete genomes from across the latitudinal gradient. We called variants for 132 strains of *D. americana* from across 32 populations against this pangenome and calculated the mutation rate, finding that there is a higher mutation rate on the 4th and X chromosome then the rest of the genome. We also predicted the effect of these variants and found that many of the variants on the 4th chromosome are missense or nonsense mutations. Then we aligned the transcripts from bulk RNA-seq from each of the strains to the pangenome to test the hypothesis of gene expression divergence in sex chromosome evolution. Through this analysis, we characterize the degeneration and regulatory divergence that arise during the early evolution of sex chromosomes.

235W Lineage specific selection in the mound-building mouse Mus spicilegus

Mara Baylis[1], Rachel Brem[2], Samantha Jackson[3], Linda Wilbrecht[3] [1]Center for Computational Biology, University of California Berkeley, [2]Center for Computational Biology, UC Berkeley, [3]Neuroscience, UC Berkeley

The steppe mouse Mus spicilegus exhibits unique behaviors within its genus, including mound-building, pubertal delay, and dispersal behaviors correlated with day length, making it an ideal system for linking lineage-specific molecular evolution to the genetics underlying these traits. Toward this end, we first analyzed coding alignments spanning Rattus and diverse Mus lineages, fitting codon models in PAML to test for lineage specific acceleration of nonsynonymous substitution on the M. spicilegus branch. Likelihood ratio tests identified 51 significant genes as candidates of positive selection along the M. spicilegus lineage, an excess relative to other Mus, including several characterized in model mice with roles in pubertal development, circadian regulation, and synaptic processes. Separately, we used phyloP to identify cis-regulatory loci whose sequences were significantly accelerated on the M. spicilegus branch, and identified elements under accelerated evolution. These elements are identified as regulatory regions with target genes involved in neuronal development and metabolism. To further contextualize gene candidates, we intersected gene sets with cell-type-resolved expression from mouse brain resources. Together, our data establish that M. spicilegus has adapted under evolutionary pressures distinct from other house mice, and they open a first window onto the genetic architecture of photoperiod gated dispersal, risk behavior, and mound-building phenotypes.

236W A chromosomal inversion underlies life history divergence on a microhabitat scale in yellow monkeyflower

Madison Plunkert[1,2], Annie Zell[3], Avin Niknafs[3], Dena Grossenbacher[3], David B Lowry[1,2] [1]Plant Biology, Michigan State University, [2]Plant Resilience Institute, Michigan State University, [3]Biological Sciences, California Polytechnic State University - San Luis Obispo

Perennial plants must survive through seasonal drought, cold, or other seasonal stressors, while annuals can complete their life cycles before these stressors occur. At the Laguna Lake and South Hills sites in San Luis Obispo, California, annual individuals of the yellow monkeyflower (*Mimulus guttatus* syn. *Erythranthe guttata*) live on a rock outcrop while perennial individuals live in a seep within tens of meters. This small spatial scale allows us to examine life history variation in the absence of complicating factors like climatic differences and isolation by distance. Through common garden experiments, we showed that genetic variation contributes to the morphological differences between annuals and perennials at Laguna Lake but not at South Hills, suggesting that phenotypic plasticity and genetic variation play different roles at the two sites. Strikingly, only the Laguna Lake perennials produced belowground stems known as rhizomes, a perennating organ that the Laguna Lake annuals and South Hills population did not produce in the greenhouse. We used pooled whole-genome sequencing to uncover genomic regions underlying life history divergence on a microhabitat scale. A 6-Mb region of chromosome 8 showed elevated genetic differentiation between annuals and perennials at each site. This genomic region corresponds to a chromosomal inversion previously shown to underlie life history divergence between populations. Here, both orientations of the inversion are segregating *within* populations. The San Luis Obispo populations are the first known instance where this inversion is driving life history divergence on the microhabitat scale, offering insight on the role of chromosomal inversions and repressed recombination in phenotypic divergence with gene flow.

237W Turtles reveal novel insights into PRDM9's role in recombination rate landscape evolution

Logan Havard[1], Chipper Johnson[1], Matthew Wolak[1], Laurie S Stevison[2] [1]Biological Sciences, Auburn University, [2]Auburn University

Functional PRDM9 drives rapid recombination evolution across vertebrates, while its absence, as shown in yeast, birds and dogs, slows evolution. This reduced turnover is due to 'default' hotspots near gene promoters, which are more stable over time. Recent work in snakes challenged this dichotomous framework that intact PRDM9 recruits hotspots away from promoters. This duality is also prevalent across mammals, but functional PRDM9 still directs the majority of hotspots. Here, we constructed a novel LD-based fine-scale recombination map in the diamond-backed terrapin. Similar to snakes, we uncovered more balanced hotspot usage between PRDM9 and default mechanisms in turtles, with default hotspots more common on microchromosomes and PRDM9 hotspots more common on macrochromosomes. Our results suggest this balanced hotspot usage is shared across reptiles, having evolved at least ~250-300 mya. We complement our recombination analysis with a PRDM9 survey targeting reptiles, specifically squamates and testudines, and identify at least 5 novel losses. Our findings increase the number of independent losses across vertebrates by 50%, and suggest these clades are particularly prone to loss, including several quite recent. When combined with documented PRDM9 loss in Archosauria (birds and crocodiles) this suggests a possible connection between PRDM9 loss and its role in suppressing default hotspots. Finally, expanding genome coverage in underrepresented, species-rich clades, such as reptiles, enabled this work and promises further evolutionary insights. Future work on these and other clades could yield novel insights on the function of PRDM9, as well as its role in fine-tuning the evolutionary rate of the recombination landscape.

238W Insect Compound Eye Cell Type Evolution Across a Speciation Continuum

Wei Lu[1], Nicholas VanKuren[2], Marcus Kronforst[1] [1]Ecology and Evolution, University of Chicago, [2]University of Chicago

Distinct eye types have evolved independently across the animal kingdom, and these ancestral architectures have been further modified to suit specific visual ecologies. A central question remains: how do these complex structures, often constrained by pleiotropy, undergo modification during rapid adaptive radiation? Traditionally, comparative studies of insect visual systems have focused on opsins and photoreceptors. Consequently, the contributions of non-opsin genes and support cell types to visual system evolution remain largely unknown. Notably, comparative transcriptomics in mammalian camera-type eyes suggests that downstream retinal ganglion cells, not photoreceptors, are the most rapidly evolving cell type.

Here, we integrated population genomics, chromatin conformation capture, and single-nucleus RNA-sequencing (snRNA-seq) to characterize cell type evolution in *Heliconius* butterflies. We sampled adult compound eyes across a speciation continuum, encompassing a single polymorphic population, fully interfertile geographical races, and sister species with partial hybrid infertility. Our analysis reveals distinct cellular mechanisms driving compound eye evolution. First, we found that genes associated with visual mate preference variation are enriched not only in photoreceptors but also in support cells, including pigment cells and lamina glia. Second, transcriptome-wide evolutionary rates analysis shows that photoreceptors evolve significantly faster than downstream lamina neurons. This pattern contradicts the mammalian model where downstream neurons drive diversity. We propose this discrepancy arises from fundamental differences in eye architecture (compound eye vs. camera-type eye) and distinct evolutionary timescales. To our knowledge, this represents the first comparative analysis of insect compound eye evolution at single-cell resolution.

239W A comparative genomic view of Tandem Repeat evolution in primates

Carolina de Lima Adam[1], Joana L Rocha[2,3], Peter H Sudmant[2], Rori Rohlfs[1] [1]Institute of Ecology and Evolution, University of Oregon, [2]University of California, Berkeley, [3]New York University

Tandem repeats (TRs) are highly mutable DNA elements that comprise nearly 8% of the human genome and can influence gene regulation, protein structure, and disease. Yet, comparative analyses have historically been limited by challenges in sequencing and annotation. Using telomere-to-telomere (T2T) reference genomes, we identified more than 3 million TRs in each of eight primate species and nearly 2 million homologous loci between 46 humans and 23 chimpanzees. We found that diversity is strongly governed by genomic context, with coding and 5' UTR sequences exhibiting reduced polymorphism and elevated length conservation, suggesting stronger stabilizing selection in these regions. Heterozygosity and mutation rates vary systematically with motif length and show concordance between indirect and pedigree-based estimates. Leveraging our population-aware, cross-species dataset, we introduced an HKA-like approach to identify TRs with signatures of selection while accounting for mutation rate variation across loci, which can vary from <10-8 to ~10-2 per generation. We found that genes containing TRs with extreme divergence-diversity ratios are enriched in nervous system development, synaptic function, and cell signaling pathways. Some of the highest-ranked diverse genic loci were found in immune-related genes, consistent with traditional targets of balancing selection. We also recovered a weak but significant correlation between TR length divergence and gene expression divergence. Altogether, our comprehensive cross-species catalog provides a valuable resource for studying TR evolution in primates. We show that TR length conservation extends beyond coding regions, underscoring the functional impact of 5' UTRs. We also show that TR diversity is a product of the interplay between mutation rate variation across motif lengths and selection in functionally relevant loci. Finally, we identified thousands of loci with signatures of natural selection and pinpointed those occurring within genes linked to essential biological processes.

240W The Genetic Basis of the Repeated Postzygotic Isolation in *Mimulus*

Hagar Soliman, Jenn Coughlan EEB, Yale University

F1 hybrid seed inviability (HSI) is a potent and widespread postzygotic barrier in flowering plants, contributing to speciation at both early and late stages of divergence. Understanding the evolutionary drivers of this barrier requires first dissecting its genetic basis. Here, we focus on three taxa within the *Mimulus guttatus* species complex that differ in their phylogenetic relatedness, yet all exhibit strong asymmetric HSI patterns when crossed. Previous work first documented such asymmetry between Northern *M. guttatus* and Northern *M. decorus*: crosses sired by Northern *M. guttatus* produced large inviable seeds, whereas the reciprocal direction yielded small but viable seeds. We recently identified a parallel pattern between Northern *M. guttatus* and a closely related subset population of *M. guttatus* from the Sierra foothills (herein "Sierran *M. guttatus*"). In this instance, the direction of asymmetry is reversed—crosses sired by the Sierran lineage, rather than Northern *M. guttatus*, produced large inviable seeds. To identify loci associated with HSI, we grew and sequenced ~6,000 offspring of reciprocal backcrosses to both parental lines to identify loci with parent-of-origin segregation distortions. Mapping the loci underlying HSI in this system will allow us to address key and long-standing evolutionary questions, such as: Do loci associated with HSI have parent-of-origin effects, as suggested by asymmetric patterns of HSI? What is the genetic architecture of asymmetric HSI? And lastly, are the HSI loci shared or distinct between the Northern *M. guttatus*–*M. decorus* and Northern *M. guttatus*–Sierran contrasts?

241W Contemporary sex-differential selection across organisms and life stages

Matthew J Ming[1], Alexa Huzar[2], Coral Loockerman[3], Daniel I Bolnick[4], Arbel Harpak[1,5], Mark Kirkpatrick[1] [1]Integrative Biology, University of Texas at Austin, [2]University of Texas at Austin, [3]Center for Biomedical Research Support, University of Texas at Austin, [4]Ecology and Evolutionary Biology, University of Connecticut, [5]Population Health, University of Texas at Austin

Contemporary sex-differential selection (SDS) generates between-sex genetic divergence. It is unclear, however, how and when in life SDS acts. Further, it is difficult to distinguish true signals of selection from artifacts which may result from, for example, biases in sampling. To better understand SDS, we carried out two studies. First, we performed meta-analysis across three human genetic studies, testing for signals of SDS that consistently replicate despite heterogeneous study participation patterns. We found twelve genes with concordant cross-study signals and identified four types of SDS, such as sex-differential viability effects as previously hypothesized, and also by selection on individual sperm cells. Second, we investigated signals of SDS at three life stages in natural populations of stickleback fishes (*Gasterosteus aculeatus*) we sampled from five lakes in Canada. We found putative targets of SDS including genes involved in development (such as *suclg2*) and in response to the environment (such as *e2f7*). Our results imply that sex-differential viability selection acts throughout the human and stickleback genomes, suggesting that SDS may be common in other species with separate sexes.

242W Inferring models of ancient introgression with single unphased genomes and a two-locus statistic

Nicholas W Collier, Aaron P Ragsdale Integrative Biology, University of Wisconsin-Madison

The emerging picture of deep human population history is one of complexity, population structure, and gene flow. Model-based demographic inference techniques that use two-locus statistics, which summarize the information contained in genealogical correlations between linked loci, are one way to resolve this picture. So far, such techniques have relied on large samples from present-day populations. Although the inclusion of ancient DNA (aDNA) in demographic inference may provide increased power to distinguish between competing historical models and estimate parameters, existing inference frameworks based on two-locus statistics are challenging to apply to aDNA samples, which are few in number, unphased, and time-stratified. Here, we introduce D+, a two-locus statistic that can be estimated from a single high-coverage diploid genome, making it well-suited for application to aDNA. D+, which is defined as the probability of observing heterozygosity at two loci in a two-haplotype sample, captures one-locus diversity and genealogical covariance through a simple relation. We show how the statistic can be extended to multi-population settings and how its properties make ancestral population structure or reticulation theoretically identifiable. We also connect the statistic to a tractable system of two-locus summaries, whose expectations can be computed numerically under arbitrarily complex demographic models. With this, we develop a model that describes the ancestry of eight hominid individuals: one contemporary and three ancient anatomically modern humans (AMH), three Neanderthals, and one Denisovan. In agreement with earlier work, we infer two episodes of gene flow from the ancestors of AMH to Neanderthals and an introgression from an outgroup hominin lineage to the ancestors of Denisovans. Our model also relates early AMH lineages in Europe. We show that the lower proportion of Neanderthal ancestry observed in western Eurasians relative to other contemporary non-African populations may be explained by ancestry contributions from a population which diverged basally from other non-African lineages and did not receive Neanderthal introgression, also in agreement with earlier work. We use a formal likelihood-ratio test to quantify the statistical support for these model features, and show that the accurate estimation of the associated parameters requires their joint inference due to their correlated effects on diversity.

244W k-mer counts in pooled sequencing data rapidly estimate frequencies of known alleles in a pangenome graph

Miles Roberts Integrative Biology, University of California Berkeley

Pangenomic references, as opposed to single-genome references, are the new standard for population genetic analyses. However, genotyping new sequencing samples against a reference pangenome can become computationally costly as either the reference pangenome scales in size or sample scales in coverage. Genotyping known variants can be massively sped up by looking for short exact matches between sequencing reads and known variants (i.e., k-mers) using tools like BayesTyper and Pangenie. However, these and similar tools are restricted to estimating only diploid genotypes whereas many population genetics datasets rely on pooling DNA from many genotypes into one sequencing library (i.e., pool-seq) to cut labor costs. Instead of diploid genotypes, pool-seq data can only estimate allele frequencies. We present a method for counting allele-specific k-mers in pool-seq data that can estimate frequencies of known pangenomic alleles in much less time than vg giraffe. We first test the method on simulated datasets, where the underlying genomes have k-mer distributions following power-laws based on all currently available plant and animal reference genomes. We found that vg giraffe and k-mer counts give similarly accurate allele frequency estimates in about 1/10th the time. Finally, we investigated empirical sequencing datasets and observed that allele-specific k-mer counts in sequencing reads recapitulate allele frequencies in previously published variant sets. Overall, counts of allele-specific k-mers can quickly estimate frequencies of known alleles in a pangenome from pool-seq data, either providing preliminary results before committing to pangenome alignment or allow scaling to large pool-seq datasets.

245W Chromosome-level assembly and annotation of the genomes of two Antarctic pinniped sister species: the leopard seal (*Hydrurga leptonyx*) and the Weddell seal (*Leptonychotes weddellii*).

Juan Pablo Aguilar Cabezas[1], Sarah Kienle[2], Douglas Krause[3], Renato Borras-Chavez[2], Emily Sperou[2], Krista van der Linde[4], Ingrid N. Visser[5], Eric Lewallen[6], Carolina Bonin Lewallen[6] [1]Biological Sciences, Hampton University, [2]Natural Resources Science Department, University of Rhode Island, [3]United States Antarctic Marine Living Resources Program, Southwest Fisheries Science Center, [4]Whale Watch Kaikoura, [5]The Orca Research Trust, [6]Marine and Environmental Sciences Department, Hampton University

Weddell and leopard seals are Antarctic seal sister species that diverged ~2.89 Mya. While the Weddell seal (WS) is an exceptional diver (can dive up to 90 min and 1000 m) and a distribution associated with fast ice, the leopard seal (LS) dives for shorter periods (max <23 min), is a shallower diver (max 355 m) and prefers fluid pack ice. These differences highlight possible adaptations that evolved during sympatric speciation and should be reflected in the genome of each species. High quality, annotated genomes for both species will facilitate comparative genomics to identify the genetic architecture and its functional significance related to speciation. PacBio HiFi reads were generated from a WS skin sample, and a preliminary assembly was built. Next, we leveraged a publicly available Hi-C scaffolded genome (DNA ZOO) to scaffold our WS assembly into pseudo-chromosomes. For the LS, we generated Omni-C libraries and scaffolded the leopard seal genome assembly (NCBI) to pseudo-chromosomes. Finally, we annotated the WS and LS genome assemblies with protein and RNA-seq data. The theoretical and empirical estimates of genome size were 2.2/2.4 Gigabases (Gb) for the WS and 2.1/2.5 Gb for the LS. Both species had 17 main scaffolds corresponding to the conserved number of chromosomes in phocids. The statistical kmer analysis indicated that the heterozygosity of the Weddell seal was half that of the leopard seal (0.2%/0.5%). Both assemblies had very high complete BUSCO score (complete carnivora orthologs), 99.4% (LS) and 99.1% (WS), and the proteomes also had a complete BUSCO score of 95.8%. We identified 22,929 genes for WS and 22,517 genes for LS, surprisingly more than other species of the subfamily. These chromosome-scale annotated genome assemblies will enable us to conduct comparative genomic analyses elucidating species divergence and adaptations contributing to our understanding of sympatric speciation in polar marine mammals.

246W Resolved globally, conflicted locally: Multi-scale genetic architecture of intralocus sexual conflict in human metabolism

Anasuya Chakrabarty[1], Kaulick Mitra[2] [1]Ashoka University, [2]BRIC-National Institute of Biomedical Genomics

Despite having almost shared genomes, males and females are strikingly different across species. This shared genetic architecture between sexes creates constraint in their evolutionary divergence under sex-specific selection when both the sexes are displaced from their phenotypic optimum, a phenomenon known as Intralocus sexual conflict (IASC). Traditionally, IASC has been studied at the level of single traits, despite traits being genetically correlated and single locus affecting many traits simultaneously. Furthermore, genome-wide genetic architecture can obscure conflict localized to specific genomic regions. In humans, large-scale genomic data now allow these questions to be addressed at unprecedented resolution.

We investigated IASC in the multivariate and local genetic architecture of 17 human metabolic and physiological traits, and lifetime reproductive success (a fitness proxy). Using large-scale GWAS summary statistics from the UK Biobank, we estimated the sex-specific additive genetic (co)variance matrices ($\mathbf{G}_{mf}$) and local cross-sex-cross-trait genetic correlations. At the genome-wide level, we found that multivariate genetic architecture was highly similar between the sexes, with only 1.15% asymmetry in the between-sex covariance matrix (**B**) which acts as a strong constraint to the evolution of sex differences under random sexually antagonistic selection. The pronounced sexual dimorphism in these traits despite such a shared architecture is consistent with a largely resolved IASC at the global (genome-wide) scale through sex-specific regulatory mechanisms rather than sex-specific multivariate genetic architecture.

Contrastingly, local analyses revealed substantial heterogeneity with 12% of regions showing significant cross-sex-cross-trait genetic correlation being consistently sexually antagonistic, having opposite fitness effects in the sexes, while 20% showed mixed antagonistic and concordant effects across trait pairs, though majority of the regions were concordant. These regions showed distinct enrichment in biological pathways with cannonical WNT signalling being enriched for antagonistic region implying conflict driven by developmental constraint. Our results demonstrate that IASC operates at multiple scales, and that integrating multivariate and local genetic architecture is essential in understanding the biology of sexual conflict.

247W Pangenome-based investigation of genomic diversity in *Kaolinonychus*: A harvestmen lineage with deep admixture history and troglophilic adaptation (Arachnida, Opiliones, Paranonychidae)

Dongyoung Kim[1], Yong-Gun Choi[2], Young-Hyun You[3], Jung-Sun Yoo[4], Choongwon Jeong[1] [1]Seoul National University, [2]The Korean Institute of Biospeleology, [3]National Institute of Biological Resources, [4]National Institute of Wildlife Disease Control and Prevention

The Korean endemic harvestmen in the monotypic genus *Kaolinonychus* represents a compelling model for evolutionary dynamics, characterized by deep phylogenetic divergence and ancient admixture. Furthermore, spanning both cave-dwelling and surface-dwelling populations, it offers a unique system to investigate the genomic mechanisms of environmental adaptation. However, deciphering these dynamics between divergent lineages is challenging with standard linear reference-based genomics approaches, which suffer from reference bias and often fail to capture lineage-specific sequences and structural variants (SVs), which are increasingly recognized as pivotal drivers of evolutionary adaptation and speciation. To address these limitations, we employed a comprehensive pangenome strategy. As a foundational step to ensure representative coverage, we identified 5 major genetic lineages by analyzing 200 nationwide resequencing samples against a single reference genome. Guided by this framework, we sequenced representative individuals from each of the 5 lineages using PacBio HiFi reads, yielding 10 highly contiguous haplotype assemblies, which were subsequently integrated using the Minigraph-Cactus pipeline. The resulting pangenome graph significantly expands the genomic representation from 822 Mb (linear reference) to 1.27 Gb, capturing 445 Mb of novel non-reference sequences. This comprehensive topology explicitly encodes complex variation patterns, recovering over 17.8 million variants, including approximately 344,000 structural variants (≥50 bp) that were previously inaccessible. Furthermore, leveraging this graph substantially improved short-read mapping and variant calling sensitivity, particularly in divergent regions. Ultimately, this resource serves as a cornerstone for unraveling the species' reticulate evolutionary history and deciphering the precise genomic signatures of cave adaptation.

248W Detecting Positive Selection Using Spatial Autocorrelation and Haplotype Structure in Populations with Limited Dispersal

Amatur Rahman, Zachary A. Szpiech Department of Biology, Pennsylvania State University

Most population genetic analyses assume discrete, well-mixed populations, yet natural populations are often spatially structured, with limited dispersal and geographic constraints on mating. These deviations from panmixia can bias demographic inference, generate spurious associations with environmental variables, and complicate the detection of selective sweeps. A selective sweep occurs when a beneficial mutation rapidly increases in frequency, carrying along linked neutral variants. Genomic regions under positive selection are therefore expected to show characteristic signatures such as long haplotypes at elevated frequency. Haplotype-based approaches generally improve power over allele-frequency-only methods by leveraging this expected pattern, but limited dispersal dynamics modify these signatures. Under limited dispersal, the slow spatial spread of adaptive alleles weakens extended haplotype homozygosity, increases intermediate-frequency variants, and promotes sweeps on multiple haplotype backgrounds, thus reducing the power of standard statistics. Accounting for spatial structure is therefore crucial for detection of selective sweeps.

Recent advances have demonstrated that incorporating geographic information can distinguish positively selected variants from neutrally evolving ones. Here, we extend this insight to haplotype-based approaches by integrating haplotype structure with spatial autocorrelation to construct a joint statistic for detecting positive selection in spatially structured populations with limited dispersal. Using individual-based simulations under a limited-dispersal model, we show that naive application of classical haplotype-based methods misses signals that are recovered by our statistic, while controlling the false positive rate at 1%. Our method also identifies geographic regions where the adaptive allele is likely segregating. We incorporate this new method into the software selscan for easy application.

249W Gene Expansion and Regulatory Rewiring Shape Sex-Biased Evolution of the Mouse Submandibular Gland

Luane Jandira Bueno Landau[1], Shikha Jain[2], Nathan Griffin[3], Achisha Saikia[4], Sarah Knox[3], Stefan Ruhl[2], Omer Gokcumen[1] [1]Biological Sciences, University at Buffalo, [2]Oral Biology, University at Buffalo, [3]Department of Cell & Tissue Biology, University of California San Francisco, [4]Biological Sciences, University at Bufalo

Mammalian saliva plays essential roles in digestion, immunity, and host-microbiome interactions, yet its protein composition varies markedly across species and sexes. The evolutionary mechanisms underlying this molecular diversity remain poorly understood. We compared mouse and human saliva secretomes at genomic, transcriptomic, and proteomic levels to investigate how saliva composition evolves. We performed RNA-seq of the three major mouse salivary glands (parotid, submandibular, and sublingual) from both sexes, with liver and pancreas as reference tissues, and integrated them with proteome data of whole saliva and reanalysis of published data. We found that evolution of genes in mouse saliva is driven by rapid gene turnover and sexual dimorphism. In the submandibular and sublingual glands respectively, 68% and 73% of expression from genes encoding secreted proteins derives from lineage-specific genes that lack one-to-one human orthologs. Mouse submandibular gland shows striking sexual dimorphism, with 1615 sex-biased genes, five times higher than in liver, which is considered a model of sex-biased expression. These genes are not randomly distributed but instead cluster in regions shaped by recent gene duplication. One such region is the kallikrein (*Klk*) gene family, a mouse-specific expansion that accounts for ~20% of male-biased submandibular expression. Our analyses suggest that this bias arises through regulatory changes that are expanded by gene duplication, including the spread of a testosterone-associated regulatory motif and the expansion of a shared chromatin domain that promotes coordinated gene regulation. Together, our results reveal how lineage-specific gene duplication and regulatory rewiring drive rapid, sex-specific evolution of the mammalian salivary secretome.

250W MHC class II DQB heterozygosity confers transgenerational fitness advantage in wild dolphins

Oliver Manlik[1], Janet Mann[2], Michael Krützen[3,4], Vivienne Foroughirad[5], Molly McEntee[2], Anna M. Kopps[6], Simon J. Allen[3,4,7], Georgeos Hardo[1], Celine H. Frère[8], Richard C. Connor[9], Lars Bejder[10], William B. Sherwin[11] [1]Biology Department, United Arab Emirates University, [2]Georgetown University, [3]Department of Evolutionary Anthrophology, University of Zurich, [4]University of Western Australia, [5]Texas A & M University, [6]Independent Consultant, [7]University of Bristol, [8]The University of Queensland, [9]Florida International University, [10]University of Hawaii at Manoa, [11]UNSW-Sydney

Natural selection drives evolution by favoring advantageous genetic variants, resulting in an accumulation of beneficial fitness traits in populations over generations. However, direct evidence linking parental genetic variation to offspring fitness in long-lived species is lacking. Here we show transgenerational fitness effects mediated by the major histocompatibility complex class II (MHC-II DQB) in a natural population of bottlenose dolphins (*Tursiops aduncus*). Our multi-decadal study (1984-2022) revealed that MHC-heterozygosity for the peptide-binding region of MHC DQB—crucial for the adaptive immune response—was a strong predictor of offspring viability. MHC-heterozygous mothers did not produce more offspring, but their offspring were about twice as likely to survive. In contrast, neutral genetic diversity was unrelated to fitness. Interestingly, MHC-heterozygous females were more successful but less sociable mothers, possibly prioritizing investment in offspring over social bonds. This is in contrast to other social mammals that rely on sociability to mediate fitness. This study offers the first empirical evidence that MHC-diversity confers a fitness advantage across generations in long-lived mammals, a finding consistent with theorized but rarely tested expectations for MHC-dependent mate choice.

251W Selective sweeps of recessive beneficial alleles in populations with continuous spatial structure

Freya Kailing, Philipp W. Messer Department of Computational Biology, Cornell University

Spatial population structure can profoundly alter the spread of adaptive mutations relative to panmictic populations with random mating. In particular, previous studies have shown that in populations inhabiting large geographic ranges, limited dispersal reduces the fixation probability of beneficial mutations and prolongs their fixation times; however, these studies have focused exclusively on co-dominant alleles. Recessive beneficial mutations could behave differently because they experience positive selection only in homozygotes, which occur more frequently in spatially structured populations due to elevated inbreeding. This effect could increase their initial rate of spread and fixation probability. Here, we use simulations of continuous-space populations to investigate the population dynamics of recessive beneficial mutations. We find that reduced dispersal increases their fixation probability relative to panmixia, particularly at higher selection coefficients, although recessive alleles remain less likely to fix than co-dominant alleles across all dispersal regimes tested. Moreover, we show that reduced dispersal accelerates the spread of recessive mutations at low frequencies but slows it at higher frequencies, resulting in the fastest fixation times at intermediate dispersal rates. Our results demonstrate that the interaction between spatial structure and dominance fundamentally shapes both fixation probabilities and fixation times, and that limited dispersal can have qualitatively opposite effects on recessive versus co-dominant beneficial alleles. These findings highlight the need to reassess population-genetic inference methods for selection that assume panmixia when applied to spatially structured populations.

252W Characterizing genomic structural variation within and between populations of *Aedes aegypti*

Gabriela Valente-Almeida[1], Remi Ketchum[1], Shaquira Little[1], William W Booker[1], Daniel R Matute[2], Daniel R Schrider[1]
[1]Genetics, University of North Carolina Chapel Hill, [2]Biology, University of North Carolina Chapel Hill

Aedes aegypti is the primary vector of several tropical diseases, including dengue, Zika, chikungunya, and yellow fever. Because vaccines and treatments for these diseases are limited or difficult to implement at scale, population control of mosquitoes has been the main strategy to lower transmission. Vector control has relied heavily on the use of insecticides, and consequently, multiple *A. aegypti* populations have acquired some degree of insecticide resistance (IR). Understanding the genetic basis of IR is critical not only for developing more effective vector control strategies but also for studying adaptive evolution in response to recent dramatic changes in the selective environment. While single nucleotide polymorphisms (SNPs) have been widely investigated in the context of rapid adaptation, structural variants (SVs) remain understudied. In this project, we investigate the content of SVs in 122 *A. aegypti* individuals sampled from North America, South America, and Africa. The shape of the site frequency spectrum suggests that many of these SVs are deleterious. However, we also found several SVs that were located near genes known to be involved in IR and that showed evidence of recent positive selection according to SweepFinder. For example, we found deletions present in putative selective sweeps encompassing cytochrome P450 genes in Senegal and Gabon, and in glutathione-S-transferase genes in Colombia and Brazil. Our results demonstrate that SVs account for a substantial fraction of genetic variation in *Ae. aegypti*, and that they may play an important role in the rapid adaptation to exceptionally strong selective pressures like insecticides.

253W Evolution of mutational fitness effects in island populations

Emma K Howell[1], Felix Baier[2,3], Hopi E Hoekstra[3], Lauren E Nolfo-Clements[4], Bret A Payseur[1] [1]University of Wisconsin-Madison, [2]Max Planck Institute for Brain Research, [3]Harvard University, [4]Framingham State University

The distribution of fitness effects (DFE) quantifies the selective consequences of newly arising mutations. Theoretical and empirical investigations of the DFE suggest that it is highly context-specific, shaped by both intrinsic properties of an organism (e.g., biological complexity) and extrinsic properties of a population (e.g., environment). Despite accumulating insights about how the DFE compares across populations and species, less is known about how this distribution changes over shorter evolutionary timescales. Island colonizations provide a powerful framework for understanding the impact of recent shifts in selection on the DFE, as founding populations often experience abrupt environmental changes. Using whole-genome, population-level sequence data, we investigate how such extreme transitions shape the DFE in two island radiations of *Peromyscus* mice: white-footed mice (*P. leucopus*) in Massachusetts's Boston Harbor and deer mice (*P. maniculatus*) in the Gulf Islands of British Columbia. Collectively, these data span distinct populations, island systems, and species, providing a comparative framework for examining the generality of DFE evolution across multiple scales. To measure the extent to which the selective effects of mutations have diverged between island and mainland populations, we leverage recent advances that extend DFE inference to multiple populations, enabling direct estimation of the correlation in fitness effects (i.e., the joint DFE correlation) between populations. We partition our analyses by functional annotation, fitting distinct joint DFEs to mutations in 3' UTRs, 5' UTRs, introns, and exons. Despite the recency of island colonization in these continental island systems, we find evidence for island-mainland divergence in fitness effects across functional annotation classes. The extent of this divergence is modulated by the level of constraint; annotation classes under stronger constraint (i.e., larger average selection coefficients, $|s|$) exhibit the highest correlations in fitness effects between island and mainland populations, while those under weaker constraint (i.e., smaller average $|s|$) exhibit lower correlations. Our findings reveal that island colonization promotes broad-scale shifts in selection that are detectable in population genomic data and that the impact of these selective shifts varies across mutation types.

254W Comparative Population Genetics of Octopus Sister Species

Silas B Tittes[1], Gabrielle C Coffing[2], Scott Small[1], Andrew Kern[1] [1]Institute of Ecology and Evolutio, University of Oregon, [2]Biology, Indiana University Bloomington

The coleoid cephalopods have convergently evolved many traits similar to vertebrates including camera-type eyes, large brain to body size ratios, and complex behaviors. Most studies have viewed cephalopod evolution over large times scales, making comparisons among individual genomes of taxa that have diverged tens to hundreds of millions of years ago. Studies at these time scales have elucidated many important features of cephalopod evolution, including evolution of proteins involved in brain and eye development, chromosomal sex determination, and more. However, very few cephalopod studies have focused on shorter time scales from a population genetics perspective, and we thus know far less about the role of recent evolutionary forces on cephalopods. In this study, we used a comparative population genetic approach using whole-genome resequencing to investigate the sister species *Octopus bimaculatus* and *Octopus bimaculoides*, which despite similar morphologies, differ substantially in their life histories, ecologies, and geographic distributions. Using demographic inference, we found that the two species have diverged relatively recently, approximately one million years ago, and that *O. bimaculatus* has consistently maintained a much larger effective population size after divergence. Consistent with their demographic histories, we found that positive selection has had a larger impact on *O. bimaculatus*, where we found stronger evidence for an association between recombination rate and nucleotide diversity, more selective sweeps, and a higher proportion of mutations fixed by adaptation; all of which are consistent with higher efficiency of natural selection in larger populations. Protein coding genes overlapping with selective sweeps were enriched for functions related to brain and eye development, suggesting that the distinct traits shared among all coleoid cephalopods continue to be shaped by positive natural selection more recently in these two species.

255W Regulation of gene expression shapes genotype x environment interactions in switchgrass

Colette Berg[1], Xiaoyu Weng[2], Jason Bonnette[2], Yuko Yoshinaga[3], Chris Daum[3], Anna Lipzen[3], Thomas Juenger[2], David Lowry[1] [1]Department of Plant Biology, Michigan State University, [2]Department of Integrative Biology, University of Texas at Austin, [3]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory

Understanding the genetic basis of local adaptation is a fundamental goal in evolutionary biology. Local adaptation is contingent on genotype-by-environment interactions, which are often mediated by variation in gene expression. The causes and consequences of variation in gene expression are complex, and they are rarely evaluated under the multi-dimensional environmental regimes that occur in the field. In this study, we investigated gene regulatory variation in the bioenergy crop switchgrass (*Panicum virgatum*) at two highly contrasting field sites in Texas and Michigan. The study leveraged two major field experiments at both sites, from which we collected leaf tissue for RNA sequencing. These experiments included a diversity panel of over 500 tetraploid genotypes, ideal for expression-based genome-wide association studies (eGWAS), and controlled crosses used for gene network and allele-specific expression analyses. eGWAS of the switchgrass diversity panel revealed over 100,000 cis- and trans- SNPs regulating gene expression. To account for deep population structure, the diversity panel was split into Midwest, Gulf, and Atlantic populations, and the eGWAS was mapped separately for each population in the Michigan and Texas common gardens. In all three populations, 30 – 50% of the mapped cis- and trans- regulatory SNPs were present in both common gardens, while the remaining cis- and trans- SNPs were unique to either the Texas or Michigan garden. In all gardens and populations, trans- regulatory SNPs had a lower mean minor allele frequency than cis- regulatory SNPs ($p < 2.2e-16$ for all comparisons), suggesting that variants regulating genes in trans are under stronger negative selection than those regulating genes in cis. Network analyses of gene co-expression in upland and lowland switchgrass in Michigan and Texas common gardens identified gene modules shaped by genotype and environment, and all of the top 9 hub genes in a genotype-associated module were heat shock protein genes. Ongoing analyses will explore allele-specific expression in AP13 x DAC6 F_1's in both gardens and identify likely candidate genes mediating phenotypic differences between upland and lowland switchgrass in both habitats. Taken together, these analyses thoroughly characterize how environmental conditions and genetic background shape gene expression regulation, and ultimately, phenotypic variation in nature.

256W Reused Genes and Divergent Paths: Genomic Predictability in Stickleback Ecotypes

Diana J Rennison[1], Matthew Farnitano[2], Zachary Tobias[2] [1]Ecology, Behavior & Evolution, University of California San Diego, [2]University of California San Diego

A central question in evolutionary biology is how predictable evolution is at the genetic level. When populations adapt to similar environments, do they diverge using the same genes and genomic regions, or does adaptation proceed through alternative genetic routes? The threespine stickleback offers a powerful natural experiment, having repeatedly evolved distinct ecotypes in parallel across independent freshwater populations. We use whole-genome resequencing data from twenty independent ecotype pairs (645 individuals) to evaluate the repeatability and limits of genomic divergence. By comparing patterns of population differentiation across ecological contrasts, we identify both shared and ecotype-specific signatures of divergence. Of particular interest, were a set of variants shared by sympatric benthic ecotypes, but differentially utilized by solitary freshwater populations; use of these variants helped explain the magnitude of divergence from the ancestral marines. In general, we observed substantial heterogeneity in the genomic architecture of divergence among ecotype pairs. Differences in the magnitude, location, and clustering of divergent regions highlight the influence of demographic history, recombination landscape, and local genetic context on evolutionary outcomes. Together, our results from this comprehensive dataset suggest that evolutionary predictability in stickleback is partial rather than absolute: parallel ecological divergence is underpinned by a core set of repeatedly used genomic regions, but shaped by context-dependent processes that generate variation among populations. Our findings clarify when and why genomic evolution is predictable and provide broader insight into the repeatability of adaptive diversification.

257W **From Monogamy to Polygamy: Mating Systems as Drivers of Sex-Biased Gene Evolution in Syngnathids**

Morjina Solaiman[1], Adam Jones[1], Sarah Flanagan[2] [1]Biological Science, University of Idaho, Department of Biological Science, [2]School of Biological Sciences, University of Canterbury

In many species, sexual selection drives the evolution of sex-biased gene expression, with a significant number of genes often showing male-biased expression, particularly in reproductive tissues such as the testes. However, little is known about gene expression patterns in species where sexual selection acts strongly on females, especially in sex-role reversed systems. Within the fish family Syngnathidae, both reproductive role reversal and mating system variation are present: in monogamous systems, sexual selection acts weakly on females, whereas in polygamous systems, sexual selection acts strongly on females. This study aims to address this gap by comparing tissue-specific transcriptomic differences across polygamous and monogamous syngnathid species. We performed a comparative transcriptomic analysis of four syngnathid species with contrasting mating systems: two polygamous species (*Syngnathus scovelli* and *Syngnathus typhle*) and two monogamous species (*Hippocampus erectus* and *Hippocampus zosterae*). We quantified the abundance and range of expression levels of sex-biased genes and estimated the nonsynonymous-to-synonymous substitution ratio (dN/dS) to detect signatures of positive selection. Genes with log□ fold changes > 2 and p-values < 0.05 were identified as significantly differentially expressed sex-biased genes. Overall, a higher number of sex-biased genes were observed in the two polygamous species compared to the monogamous species, and reproductive tissues exhibited the greatest expression differences relative to somatic tissues in both mating systems. Our findings suggest that species experiencing stronger sexual selection pressures harbor more sex-biased genes and exhibit higher evolutionary rates. These results contribute to ongoing debates about whether sexual selection drive the evolution of sex-biased gene expression across species and mating systems.

258W Germline mutation rate evolution in species with contrasting mating systems

Ivanna Ostapchuk[1,2], Elizaveta Grigoreva[3], Guy Sella[1], Magnus Nordborg[3] [1]Columbia University, [2]Gregor Mendel Institute of Molecular Plant Biologu, [3]Gregor Mendel Institute of Molecular Plant Biology

Mutation rate is a fundamental parameter in evolutionary processes, contributing to the generation of genetic variation and the accumulation of deleterious mutations. Discoveries of mutator alleles in mice (Sasani et al. 2022, Sasani et al. 2024) provide evidence that germline mutation rate can vary among individuals and therefore can be a trait subject to selection pressure. Mating systems represent one of the factors that can influence the extent of linkage disequilibrium between mutator alleles and deleterious mutations they generate (Nordborg 1999). This work investigates how mating systems - self-fertilization, clonality, and phylogenetically established inbreeding - influence the evolution of mutation rates. We focus on two opposing forces: an excess of deleterious mutations associated with mutator alleles, which strengthens indirect selection against them and reductions in effective population size , which weaken purifying selection through genetic drift.

We extend a modifier-locus model (Milligan et al. 2022) to jointly account for selfing-induced reductions in effective recombination rate and effective population size. Analytical results indicate that moderate to high selfing rates strengthen associations between mutator alleles and deleterious mutations, leading to a reduced modifier-driven increase in mutation rates relative to outcrossing populations.

We also compare theoretical predictions to available pedigree-based, mutation-accumulation, and synonymous substitution data from sister species pairs across plants, fungi, nematodes, and arthropods. Empirical patterns are broadly consistent with theoretical expectations, with outcrossing species often exhibiting higher estimates of μ or dS than selfing or inbred species. Exceptions include Arabidopsis and Caenorhabditis clades, in which dS cannot be used as a direct proxy for μ due to unknown differences in generation times in natural populations. Overall, our combined theoretical and empirical approach highlights an interplay between the strength of associations of mutator alleles with deleterious mutations and genetic drift in mutation-rate evolution, offering a framework for interpreting variation in μ across taxa.

259W Evolution of gene expression in the trout retina

Emily EK Kopania, Nathan L Clark Biological Sciences, University of Pittsburgh

Many species experience changes in their visual environment as they develop from juvenile to adult stages, and it is unclear how ecological shifts across life history shape eye evolution. In parallel, many genes essential to eye development are highly conserved, so species must respond to differences in their visual environments within the constraints of eye development, often through changes to gene expression in the retina. Salmonid fishes provide an ideal system for investigating retinal transcriptome divergence across ontogeny. For example, rainbow trout, which live in shallow streams their whole lives, and steelhead trout, which migrate from shallow streams to deeper oceans or lakes, are the same species, *Oncorhynchus mykiss*. Although they both develop in shallow streams, their adult forms exist in different visual environments, leading to the predication that gene expression in their retinas should diverge across development. Here we ask: What are the transcriptional differences between steelhead and rainbow trout retinas, and do these represent evolved differences among populations? We sampled retinas from three types of trout from Pennsylvania fish hatcheries: steelhead trout, rainbow trout, and an outgroup, brown trout (*Salmo trutta*). The steelhead in this population migrate to Lake Erie, which has much deeper and open water compared to nearby streams. We performed RNAseq to investigate expression differences among trout types and identified 33 shared differentially expressed (DE) genes between steelhead and either rainbow or brown trout, many of which have putative eye functions. These fish were sampled before the steelhead migrated to Lake Erie, and therefore these DE genes likely represent evolved differences in expression rather than responses after exposure to a new visual environment. In our ongoing work, we are using single-nucleus RNAseq to investigate the evolution of cell-type composition and cell-type-specific expression in trout, because the retina is a complex tissue made up of many cell types with different functions in vision. Our work suggests that there are evolved differences in the trout retina transcriptome associated with different visual environments. This establishes trout as an exciting model system to investigate retina transcriptome evolution in the context of interactions between developmental constraints and ecological selective pressures.

260W Sequence context and methylation interact to shape germline mutation rate variation at CpG sites

Sheel Chandra[1], Ziyue Gao[2] [1]Biology, University of Pennsylvania, [2]Genetics, University of Pennsylvania

The DNA sequence surrounding a nucleotide strongly influences its probability of mutation. A classic example is the elevated transition rate at CpG sites, which is largely attributed to cytosine methylation. However, CpGs with different flanking sequences also exhibit mutation rate variation, which is only partially correlated with context-specific methylation level. To investigate processes underlying this variation, we introduce a framework for modeling CpG mutation rates across contexts, treating methylation level as a continuous predictor and explicitly modeling the probability of recurrent mutations. Applying our model to human polymorphisms from gnomAD and the 1000 Genomes project, we estimated CpG mutation rates of unmethylated and methylated CpGs, respectively, in each unique 4-mer and 6-mer context. We found that unmethylated and methylated CpGs exhibit different context-dependent mutation patterns, suggesting that they act as distinct mutational substrates. In addition, we found that upstream and downstream sequences exert largely independent effects on CpG mutability. Extending this analysis to other primates reveals both conserved and species-specific patterns. Notably, an upstream adenine consistently elevates CpG mutation rates across species, independent of methylation status or downstream sequence. In contrast, methylated CpGs show greater divergence in context-dependent mutation rates among primates than unmethylated sites, pointing to recent evolutionary changes in context effects potentially linked to DNA demethylation and repair.

261W The genomic landscape of structural variation during parallel divergence in three-spined stickleback

Matthew C Farnitano, Zachary JC Tobias, Diana J Rennison Ecology, Behavior, and Evolution, University of California San Diego

While most studies of genomic variation focus on single-nucleotide polymorphisms (SNPs), there is increasing appreciation for the important role of structural variants (SVs) as drivers of phenotypic variation, adaptation, and divergence. Yet we still know little about the genomic distributions, effect sizes, evolutionary histories, and adaptive roles of structural variants within diverse and dynamic natural populations. We use the recent colonization of replicate freshwater habitats by ancestrally-marine three-spined stickleback (Gasterosteus aculeatus) as a backdrop to investigate the genomic distribution of copy-number variants, indel polymorphisms, and other structural variants within and between diverging populations. We leverage a whole-genome resequencing dataset of 645 individuals across 3 benthic-limnetic ecotype pairs, 7 lake-stream pairs, 10 solitary lakes, and 3 marine populations, giving us unprecedented resolution of structural variation dynamics in natural populations. We quantify the extent of shared and unique structural variation across replicate populations and freshwater ecotypes. By directly comparing the distributions and patterns of structural variation to those of single-nucleotide polymorphisms, we assess whether structural variants have played a unique role in rapid, parallel adaptation. Furthermore, we utilize demographic inference of surrounding genomic regions to address whether structural variation is more often enriched from ancient standing variation or is more recently derived. Our investigations provide a powerful look into how structural variation impacts the evolutionary dynamics of rapid divergence in diverse natural populations.

262W Adaptation vs sexual selection: Exploring the genes and selective pressures underlying a pheromone difference between natural Drosophila populations

Myron B Child[1], Henry Chung[2], Collin Gehl[1], Emma Cabage[1], John Pool[1] [1]Genetics, UW Madison, [2]Michigan State University

For new species to form, some degree of reproductive isolation needs to occur to prevent gene flow between interbreeding groups. Adaptation and sexual selection can cause populations to diverge based on what phenotypes are advantageous in their unique environments, leading to phenotypic and genetic differences that are incompatible between populations. In Drosophila melanogaster, cuticular hydrocarbons (CHCs) act as both mating pheromones and as a protective outer layer against dry environments, meaning they are affected by adaptive and sexual selection. Male flies from West Africa and the Caribbean have a CHC profile uniquely dominated by 7-pentacosene (7-P), whereas most other populations have males with a predominantly 7-tricosene (7-T) profile. It has been hypothesized that this difference may be a result of environmental adaptation to West Africa's hotter environment, however, the amount of 7-T versus 7-P on a male has been shown to affect female mating preference. Our experiments on a highly 7-P biased line from West Africa show that it has improved desiccation resistance and stronger intra-population female preference compared to a highly 7-T biased line. Population genetic analysis has identified two candidate genes for the observed 7-T/7-P ratios, which lie in previously identified QTL regions and are predicted to be fatty acyl-CoA reductase and elongase genes. We are now generating strain-specific knockouts of each gene to do reciprocal hemizygosity assays to test whether these genes may have evolved in response to adaptation or sexual selection. Additionally, we are performing desiccation resistance and female mate choice assays on wild-derived populations from multiple geographic locations to determine how these traits vary between and within geographic and environmental groups, and whether that difference correlates with levels of 7-T and 7-P.

263W The genetics of gene expression divergence in wild baboons

Shuyu He[1], Jordan Anderson[1], Iker Rivas-González[2], Arielle S Fogel[1,3], Ruth K Nyakundi[4], Mercy Y Akini[4], Lucy Ochola[4], Elizabeth A Archie[5], Susan C Alberts[1,6], Jenny Tung[2,6,7] [1]Department of Evolutionary Anthropology, Duke University, [2]Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, [3]Department of Molecular Biology & Genetics, Cornell University, [4]Kenya Institute of Primate Research, [5]Department of Biological Sciences, University of Notre Dame, [6]Department of Biology, Duke University, [7]Duke University

Understanding primate evolution requires studying the contribution made by gene regulation. Regulatory variation plays an important role in shaping phenotypic diversity and can contribute to the emergence of species-specific traits. While much progress has been made in understanding the genetic basis of gene regulatory traits in humans, much less is known about how gene regulation in primates evolves on deeper timescales. To address this gap, we integrated genome-wide genotyping and RNA-seq data from 175 wild baboons in the Amboseli ecosystem of Kenya. This population has experienced both ancient and recent hybridization between yellow baboons (*Papio cynocephalus*) and anubis baboons (*P. anubis*; ~1.4 million years diverged). We collected paired control and lipopolysaccharide (LPS)-stimulated blood samples to simulate the response to bacterial infection. We identify *cis*-acting expression quantitative trait loci (eQTL) for over 55% of blood-expressed genes (FDR<0.05). For ~5.5% of expressed genes, we also identified response expression quantitative trait loci (reQTL), where genotype predicts the magnitude of the within-individual *response* to LPS stimulation. Genes with at least one significant (r)eQTL were enriched for key biological processes, including T cell–mediated immunity, consistent with the view that adaptive immune regulation is a frequent target of natural selection and undergoes rapid evolutions. We then used high-coverage resequencing data from two panels of unadmixed yellow baboons and anubis baboons to measure F_{ST} and d_{xy} between species, as well as pi within each species, across the genome. Compared to all tested variants, (r)eQTL SNPs were enriched for variants that show high between-species divergence but low within-species diversity, suggesting that our ability to link them to gene expression derives in part from working with an admixed population. Together, our results illustrate how natural hybrid zones facilitate the identification of functionally diverged genes and variants important during primate speciation.

264W Numerical challenges when inferring selection and population size history from the site frequency spectrum

Jonathan Wiese, Matthias Steinruecken Ecology and Evolution, University of Chicago

Inferring both non-equilibrium demography and selection from observed genetic data is a central goal in population genetics. The site frequency spectrum (SFS) is a widely used summary of observed genetic variation, and in order to utilize it for inference, we need to model how the shape of the expected SFS depends on the parameters of interest. Here we implement and analyze several numerical approaches, including the exact Wright-Fisher (WF) model, diffusion-based approximations, and moment-based approaches, to compute the expected SFS for non-equilibrium demographic models with selection. We test each approach on simulated data using a maximum likelihood framework for parameter inference. For positive selection, we find that all methods produce comparable results in the weak selection, small population size regime. However, diffusion-based methods diverge from the exact WF model when selection is moderate to strong, even when standard approximations assuming weak selection are relaxed. Furthermore, moments-based methods are only numerically stable for weak selection, while exact WF-based methods are computationally intractable for large population sizes. We conclude that the optimal method to compute an expected SFS depends on both the strength of selection and the population size history. For likelihood-based inference of demographic and selection parameters, we recommend a blended approach, in which different methods are used in their appropriate regimes in the parameter space.

265W Genomic analysis of *Solanum polygamum* suggests convergent evolution of sex determination via a response regulator gene

Aldo Carmona Baez[1], Sarah Carey[2], Kira McQuaig[3], Amy M LaFountain[4], Gregory Anderson[4], Alex Harkess[2], Rafael Guerrero[3] [1]North Carolina State University, [2]HudsonAlpha Institute for Biotechnology, [3]Biological Sciences, North Carolina State University, [4]Ecology and Evolutionary Biology, University of Connecticut

The study of sex determination systems can inform key evolutionary questions, such as how genomic regions evolve under suppressed recombination and dosage compensation. Recent studies of dioecious plant species have uncovered independently evolved cases in which the genetic basis of sex determination involves response regulator genes, suggesting that this may represent convergent evolution across the angiosperm clade. This study explores the genetic basis of the previously uncharacterized dioecious species *Solanum polygamum* through de novo assembly and comparative genomics. We show that *S. polygamum* has a ZW sex-determination system that is potentially controlled in a 'switch fashion' by the presence or absence of a response regulator gene. This candidate sex-determinant gene is located within an inversion that likely suppresses recombination near the sex-determinant region, consistent with theoretical expectations. In addition, we show that *S. polygamum* has a ~2Gb genome and is phylogenetically closer to other species with similar-sized genomes that have undergone a dramatic TE expansion. Our results support the potential for convergent evolution in the genetic basis of sex determination across plant species and provide another empirical example of species that underwent a large expansion of TEs and also transitioned to dioecy. This study demonstrates the potential of plant sex determination systems to serve as models for complex evolutionary questions.

266W Spatial population structure in the hyperpolymorphic Pacific acorn barnacle

Angel G Rivera-Colon, Scott T Small, Andrew D Kern Institute of Ecology and Evolution, University of Oregon

The Pacific acorn barnacle (*Balanus glandula*) is a widespread intertidal species along the Pacific coast of North America (Alaska to Baja California) that harbors exceptionally high genome-wide genetic diversity, consistent with very large effective population sizes. While prior work has explored evolutionary consequences of this diversity, the geographic distribution of genetic diversity across the species' range remains poorly characterized at a genome-wide scale. We sequenced whole genomes from 54 individuals sampled at 12 coastal sites spanning the species' distribution. Genome-wide polymorphism is extreme, with segregating-site density roughly an order of magnitude higher than in well-studied model systems such as *Drosophila*. Population genomic analyses reveal strong spatial structure across the range, including divergence between northern and southern populations associated with a previously described cline near coastal California. We also find suggestive evidence of an additional transition in the northern portion of the range, separating mid-latitude populations in Oregon and Washington from higher-latitude populations in British Columbia and Alaska. Together, these results provide the first genome-wide assessment of diversity and population structure in *B. glandula* and establish a foundation for future work on how hyperpolymorphism is maintained across heterogeneous environments.

267W Using an extinction-recolonization framework to model *P. falciparum* populations

Cobi M Henry[1], Parul Johri[2] [1]Biology, University of North Carolina at Chapel Hill, [2]Biology/Genetics, University of North Carolina at Chapel Hill

P. falciparum, the deadliest species of malaria-causing parasites, exhibits complex population dynamics shaped by severe transmission bottlenecks, selfing, and alternating sexual and asexual reproduction. Moreover, malaria-endemic areas are characterized by higher rates of transmission between hosts than in regions with a low malaria prevalence. To understand how these complex dynamics affect genomic variation in *P. falciparum* populations, we employ forward-in-time simulations to model both within-host and between-host dynamics using an extinction-recolonization framework in a metapopulation where demes represent hosts. While populations with high transmission are characterized by low extinction rates and high migration rates (resulting in more mixed infections), populations where malaria is less prevalent tend to have fewer mixed infections, and greater rates of extinction. We examine the effects of transmission bottlenecks, extinction rate, and migration rate on the site frequency spectrum, linkage disequilibrium, and genetic differentiation between hosts under neutrality as well as selection. Finally, we estimate extinction and migration probabilities in various populations of malaria and compare simulated patterns of variation to observed data. We find that malaria populations may be characterized by high rates of extinction (i.e., a minority of infected hosts pass on their infection to the next generation) and/or low rates of migration between hosts. These scenarios are most likely to result in high ratios of nonsynonymous to synonymous diversity and similar site frequency spectra between neutral and selected sites, as commonly observed in *P. falciparum* populations. Such a framework may provide a way towards developing an evolutionary null model to perform population genomic inference in *P. falciparum* populations.

268W Population history and migration in sub-Saharan African *Drosophila melanogaster*

Austin Daigle[1], Parul Johri[1,2] [1]Genetics, University of North Caronlina at Chapel Hill, [2]Biology, University of North Caronlina at Chapel Hill

Demographic inference in species with functionally dense genomes is challenging because pervasive selection distorts variation in ways that can mimic or obscure demographic signals. Here, we infer the demographic history of three sub-Saharan African *Drosophila melanogaster* populations, including flies sampled from remote regions proposed to lie within the ancestral range of the species. We combine high-quality genome annotations with conservation information to parameterize theoretical expectations for background selection that explicitly incorporate the distribution of fitness effects of new mutations in coding and conserved non-coding regions, and use these expectations to identify and retain sites least affected by selection for downstream inference. Using these putatively neutral sites, we estimate the best-supported three-population topology and obtain parameter estimates that include asymmetric, bidirectional migration rates among populations, consistent with previously reported unique mating behaviors in flies from West Africa that confer reproductive isolation with flies from other *D. melanogaster* populations. These findings refine our understanding of *D. melanogaste*r's ancestral dynamics while highlighting the challenges of inferring complex demographic history in functionally dense genomes highly influenced by selection.

269W The harmonic mean recombination rate: properties, measurement, and implications

Tyler V Kent[1], Carl Veller[2] [1]University of Chicago, [2]Ecology & Evolution, University of Chicago

Metrics of genome-wide recombination should measure what recombination does. In evolutionary genetics, metrics of genome-wide recombination should therefore measure the degree to which recombination shuffles allelic combinations genome-wide, with reference to the evolutionary processes that this shuffling influences. Here, we explore properties of the harmonic mean recombination rate, r_{har}, which is known to mediate several important evolutionary-genetic processes, such as the genome-wide reduction in genetic diversity due to linked selection and the reduction in genetic variance of traits under stabilizing selection. We show that, relative to the arithmetic mean recombination rate r_{bar}, r_{har} is extremely sensitive to both fine- and broad-scale variation in the recombination landscape, and to the density along chromosomes of the loci between which it is measured. Consequently, the effect of recombination on the evolution of a complex trait is sensitive to details of the trait's genetic architecture. We further demonstrate that the number of chromosomes and the distribution of selected sites along chromosomes, via their influence on r_{har}, likely play a major role in shaping differences between species in the loss of genetic diversity due to linked selection.

270W Identifying key adaptations facilitating functional differentiation of duplicate genes

Liaoyi Xu[1], Arbel Harpak[2] [1]University of Texas at Austin, [2]Integrative Biology, University of Texas at Austin

Gene duplication is a major source of adaptation since it allows gene duplicates to assume distinct functions. However, the high sequence similarity between segmental duplicates can lead to a specific DNA repair error called interlocus gene conversion (IGC). IGC involves the accidental use of a paralogous sequence as a template for repair, resulting in a "copy-paste" event between different loci. Frequent IGC homogenizes differences between duplicates and can lead to "concerted evolution" whereby the duplicates maintain the same alleles and function. However, if natural selection favors distinct alleles at a given site, IGC would be selected against. Here, we developed a method to detect "IGC islands"–small (several bp long) regions of segmental duplicates where IGC events appear to be purged from the population, within SDs that otherwise experience frequent IGC–hypothesizing these islands would confer key adaptations involving the differentiation of function of the duplicates. We applied our method to data from the Human Pangenome Reference Consortium (HPRC) and identified IGC islands across the ~5% of the human genome that is segmentally duplicated. For example, in the opsin genes OPN1LW and OPN1MW, we detected several IGC islands. Among these, three contain nonsynonymous mutations that are known to shift spectral sensitivity between longer (red) and shorter (green) wavelengths, with effects that are nearly additive. Overall, our method offers a promising way to identify specific sites under strong, often interpretable selection pressures in regions of the genome that have traditionally been among the hardest to study.

271W **Decoupling Migration from Coalescence under the Structured Serial Coalescent**

Hao of Shen, John Novembre Human Genetics, University of Chicago

As population genomic data increasingly span both geographic space and evolutionary time, there is a growing need for demographic inference methods that better utilize time-series sampling and spatial resolution. We contribute to this by developing a theoretical foundation for pairwise genealogical processes under the structured serial coalescent (SSC), deriving evolution equations for pairwise branch length distributions. By classifying these relationships according to their parameter dependencies, we identify a computationally tractable class of branch lengths—scaling at $O(d^3)$ for d demes—that depends primarily on lineage migration rates, independent of local population sizes.

This decoupling provides a mathematical basis for addressing the long-standing identifiability problem between gene flow and effective population size, particularly when sampling intervals are short relative to the coalescence timescale. We demonstrate that these relationships are highly flexible, extending to diverse summary statistics including allele frequencies, the length of IBD tracts, and the distribution of pairwise coalescence times inferred from Ancestral Recombination Graphs (ARGs). Finally, we propose an inference framework for estimating absolute, time-varying migration rates and demonstrate its utility through a toy example. This work provides a scalable path for reconstructing high-resolution migration dynamics in evolution, epidemiology, and conservation

272W **Spatial Transcriptomics to Characterize a Novel Pigmentation Trait in Swordtail Fish**

Rhea Sood, Molly Schumer Stanford University

Coloration and pigmentation are remarkably variable within and between species and can serve as an important window into the mechanisms of adaptation. We find that the trait is driven by the recent retrovirus insertion near the gene alkal2a. This insertion is associated with changes in the chromatin landscape, upregulation of alkal2a, and accumulation of iridescent cells that adhere to the scales. Alkal2a is expressed differentially across age groups, and we can see this difference across individuals that we tested for gene expression and constructed genomes for. We also characterize the spatial expression of alkal2a in different regions of the body of the fish to understand the development of the trait further.

273T Integrating Pan-Genome and AI to Identify Adaptive Causal Variants for Maize Environmental adaptation

Wei-Yun Lai, Jingjing Zhai, Terry Casstevens, Lynn Johnson, Zachary Miller, Ana Berthel, Jose A. V. Franco, M. Cinta Romay, Michelle C Stitzer, Sheng-Kai Hsu, Germano Costa-Neto, Edward S Buckler Cornell University

Maize (Zea mays ssp. mays) was domesticated in tropical lowlands but later expanded into highland and temperate regions that impose temperature stresses during development. Although highland landraces show enhanced seedling cold tolerance and accelerated development, the evolutionary origins and molecular basis of these adaptations remain unclear, in part due to marker sparsity and linkage disequilibrium. To overcome these limitations, we generated 17 high-quality long-read maize genomes capturing extensive tropical diversity and combined them with existing assemblies to construct a practical haplotype graph (PHG) of 80 diverse genomes. Using Zea mays ssp. huehuetenangensis, an ourgroup subspecies isolated from subsequent hybridization, as the reference reduced reference bias, yielding ~250 million SNPs. We leveraged this PHG to impute genome-wide genotypes for 3,300 locally adapted maize landraces from skim sequencing data. With this, we test two hypotheses: whether temperature-adaptive alleles reflect the evolutionary legacy of ancient whole-genome duplication and introgression from highland teosinte, and whether temperature adaptation primarily targets protein-coding variation rather than regulatory change. We observed an elevated proportion, up to 20%, of highland teosinte haplotypes in highland landraces. Environmental GWAS identified 454 loci associated with elevation adaptation. Contrary to the prevailing views emphasizing regulatory importance, we observed 3-fold stronger enrichment in coding than regulatory regions among these loci. Coding signals were further enriched 13-fold among syntenic duplicate genes, highlighting the role of ancient WGD in preserving protein isoforms optimized for diverse thermal regimes. Finally, integrating zero-shot scores from a plant DNA foundation model (PlantCAD) prioritized putative causal variants in 63% of candidate genes, including HPC1, COP1L, SS7, and SUS6. Together, our results clarify the roles of WGD and protein optimization in maize adaptation and provide a scalable framework for identifying climate-resilient alleles.

274T The influence of abiotic factors on a selfish centromere

Findley Finseth[1], Carolyn R Fishman[2], Katya Gilmore[2], Rayne McNerney[2], Clair Wu[2], Evan Stark-Dykema[3], Tim Wheeler[3], Lila Fishman[3] [1]Department of Natural Sciences, Scripps, Pitzer, and Claremont McKenna Colleges, [2]Department of Natural Sciences, Scripps College, [3]Division of Biological and Biomedical Sciences, University of Montana

Gene drivers violate Mendelian inheritance by biasing their own transmission, often at a cost to their carriers. The asymmetry of female meiosis creates an arena for centromeres to drive, as “strong” centromeres that are preferentially segregated to the egg or megagametophyte gain a transmission advantage. Although selfish centromere drive has a genetic basis, drivers evolve within biological and physical environments that can influence their function and costs. In plants, centromere drivers may be particularly sensitive to fluctuating environments, as heat stress alters centromere loading and chromosomal segregation and can cause pollen infertility. The yellow monkeyflower, *Mimulus guttatus,* possesses a centromere-associated, female meiotic drive locus (MDL11) that imposes fitness costs, offering an opportunity to study how abiotic factors influence drive dynamics. First, we characterized the distribution of MDL11 across a landscape. While MDL11 appeared widespread in a previous study, re-analysis of whole-genome data revealed that MDL11 is restricted to the Cascades suggesting environmental factors may limit its spread. Second, we explored how thermal stress impacted the fitness costs of drive. We exposed F2s segregating for drive and non-drive centromeres to a range of temperatures and assayed pollen fertility. Finally, we used RNA-Seq to explore how changes in gene expression underlie drive and heat-associated pollen infertility costs. Preliminary results suggest that increasing thermal stress dramatically reduces pollen fertility, particularly at early stages of development. Investigation of MDL11’s interactions with abiotic stressors will shed light on both the empirical dynamics of this model system and the basic biology of centromeric drive.

275T Mapping the Genomic Landscape of Convergent Dietary Adaptation Across Mammalian Liver Transcriptomes

Gwendolyn Johnson[1], Harshal Wasnik[2], Amelia Harralson[2], Derrick Um[3], Isabelle Tedesco[3], Nicholas DiVittorio[3,4], Sara Goldin[2,3], Allie Graham[5], Irene Kaplow[6], Wynn Meyer[2] [1]Biological sciences, Lehigh University, [2]Biological Sciences, Lehigh University, [3]Lehigh University, [4]Moravian, [5]University of Kansas, [6]Carnegie Mellon University

The mammalian phylogeny encompasses a staggering diversity of nutritional strategies, with repeated, independent acquisition of similar dietary phenotypes. Thus, dietary differences in mammals, ranging from generalist herbivory to highly specialized insectivory, provide a powerful model for investigating convergent evolution.

Traditional dietary classifications rely on broad categorical labels, such as 'omnivory’, which obscure significant complexity in the relative contribution of distinct nutrient sources. By quantifying diet as a compositional continuum, we achieve greater rigor in identifying liver transcriptomic variation across multiple dietary groups.

In this study, we analyzed publicly available liver RNA-seq datasets across hundreds of mammalian species, quantifying protein-coding transcript abundance and identifying orthologous genes. By applying phylogenetic generalized least squares (PGLS), we tested for relationships between gene expression levels and dietary phenotype, accounting for phylogeny to isolate changes in expression due to convergent evolution from those due to shared evolutionary history.

Our preliminary results identify significant hepatic expression differences by dietary specialization. Specifically, herbivorous taxa exhibit increased expression levels in pathways for carbohydrate processing and lignocellulose degradation, potentially reflecting the regulatory changes required for shifts to plant-based diets. Conversely, we observe higher expression in genes governing lipid catabolism and bile acid metabolism in carnivorous lineages, potentially reflecting metabolic adaptations to high-protein, high-lipid intake. We find that immune-related pathways like xenobiotic metabolism show differences in expression across diet. Additionally, our data suggest that the broad ‘carnivore’ classification masks metabolic differences. While consumers of vertebrates and invertebrates are conventionally grouped together under one label, divergent transcriptomic signatures in vertivores and invertivores, such as within intestinal lipid absorption pathways, challenge this assumption.

In identifying candidate genes and pathways with expression changes associated with convergent dietary specialization, we shed light on the expression changes linked to metabolic evolution, with implications for the physiological diversity and disease susceptibility shaped by diet across the mammalian phylogeny.

276T Detecting Convergent Evolution Beyond Individual Genes: Insights from Cluster-Level Selection

Ethan Talley[1], Xuan Zhuang[2] [1]Biological Sciences, University of Arkansas, [2]University of Arkansas

Comparative evolutionary genomics provides valuable insights into how natural selection shapes species' traits. Genomic analysis of lineages that have independently evolved similar traits (convergent evolution) can identify which genetic changes are linked to the shared trait and which are specific to each lineage, and we can also discover previously unrecognized molecular, functional, and phenotypic similarities among those lineages. In recent years, many methods have been developed to test for convergent evolution on the level of individual genes, but these often fail to produce results with explanatory power. This may indicate that convergence below the level of phenotypes is uncommon, but it is equally possible that convergence in groups of structurally similar or functionally related genes is at play. However, tools for detecting convergent evolution in sets of functionally related genes remain comparatively underdeveloped. Using cold-adapted fish as a study system, we employed two tests to identify candidate gene clusters that are likely involved in cold adaptation, using species without the phenotype of interest to differentiate between phenotype-related selection and phenotype-unrelated selection. We identified many promising groups of genes, including gene sets with prior support in the literature as well as novel, functionally coherent clusters that would likely have been missed by singlegene analyses. These results illustrate the power of geneset–based tests of convergence for uncovering the molecular basis of complex adaptive phenotypes.

277T Population Structure and Genetic Divergence in Marine Threespine Sticklebacks Along the Western Coast of North America

David Ferranti, Caitlin Wise, Diana Rennison Ecology Behavior and Evolution, University of California San Diego

Adaptive evolution facilitates the colonization of new habitats and persistence of populations in a world increasingly marred by anthropogenic activity. Characterizing the processes that affect genetic variation in natural populations is key to understanding the dynamics of rapid adaptation to shifting fitness landscapes. To investigate the forces that shape genetic variation in natural populations, we use threespine stickleback (*Gasterosteus aculeatus*) collected along the western coast of North America. Although threespine stickleback have been used as a model system to examine rapid adaptation from a marine anadromous form to a freshwater ecotype, little attention has been paid to the marine ecosystem, which is often assumed to represent a static, panmictic population. Much of the ecology of marine stickleback remains a mystery, including the extent of their natal homing behavior. It is known, however, that marine sticklebacks face different selection pressures, such as fluctuations in sea surface temperature, across their range, which may lead to local adaptation. Here we present an analysis of threespine sticklebacks sampled from Alaska to southern California to determine patterns of population structure, genetic diversity, and migration. We analyzed the patterns of genetic diversity at both large and small geographic scales and found evidence of differentiation between California estuarine fish and their northern counterparts as well as finer population structure in distinct ecoregions such as the Salish Sea and Puget Sound. Selection scans within population clusters identified shared and unique genomic targets of selection, allowing comparison to known windows of divergence important for adaptation in postglacial freshwater populations. We also quantified patterns of isolation by distance along the coast to determine the extent and directionality of stickleback migration. Together, our findings provide insight into the historical and ongoing evolution of a species across a geographic range of thousands of kilometers.

278T A statistical method to estimate mutational variance in humans

Arun Durvasula[1], Nicholas Mancuso[1], Evan Koch[2] [1]Population and Public Health Sciences, USC, [2]Genetics, Yale University

We have learned much about the influence of segregating genetic variation on complex diseases and traits but still know little about the impact of new mutations on complex trait variation in humans. Traditional approaches to estimate the mutational variance rely on measuring phenotypic variation over successive generations in isogenic lines, which is not possible in humans. Despite this, recent large-scale whole-genome and exome sequencing efforts in biobanks provide an opportunity to characterize the impact of rare and ultra-rare variation on complex diseases and traits.

Here, we describe a statistical approach that models the distribution of observed effects from exome-wide association studies using a frequency-dependent mixture model. We estimate the distribution of effect sizes for new mutations by taking the limit as the frequency approaches zero. Our approach is highly scalable, and capable of running on both CPU and GPU machine architectures. Through realistic simulations, we show our method produces accurate estimates of the distribution of mutational effects compared with the simulated values.

We apply our method to exome-wide association summary statistics from two complex traits, BMI and LDL and estimate the mutational variance. Taken together, our approach enables inference of the mutational variance in humans across a wide range of complex traits and diseases.

279T The rights and wrongs of rescaling in population genetics simulations

Parul Johri[1], Fanny Pouyet[2], Brian Charlesworth[3] [1]Biology, University of North Carolina at Chapel Hill, [2]Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, [3]Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh

Computer simulations of complex population genetic models are an essential tool for making sense of the large-scale datasets of multiple genome sequences from a single species that are becoming increasingly available. A widely used approach for reducing computing time is to simulate populations that are much smaller than the natural populations that they are intended to represent, by using parameters such as selection coefficients and mutation rates, whose products with the population size correspond to those of the natural populations. This approach has come to be known as rescaling, and is justified by the theory of the genetics of finite populations. Recently, however, there have been criticisms of this practice, which have brought to light situations in which it can lead to erroneous conclusions. This paper reviews the theoretical basis for rescaling, and relates it to current practice in population genetics simulations. It shows that some population genetic statistics are scaleable while others are not. Additionally, it shows that there are likely to be problems with rescaling when simulating large chromosomal regions, due to the non-linear relation between the physical distance between a pair of separate nucleotide sites and the frequency of recombination between them. Other difficulties with rescaling can arise in connection with simulations of selection on complex traits, and with populations that reproduce partly by self-fertilization or asexual reproduction. A number of recommendations are made for good practice in relation to rescaling.

280T A catalog of structural variation in sample of high-quality *Culex quinquefasciatus* genome assemblies

Remi Ketchum, Anna Trotter, Daniel Matute, Daniel Schrider University of North Carolina at Chapel Hill

The southern house mosquito *Culex quinquefasciatus* is a major vector of zoonotic pathogens, including West Nile fever. Insecticides are widely used to control *C. quinquefasciatus* populations, but resistant mosquitoes have emerged in numerous locations across the globe including the United States. Much progress has been made in recent years in identifying insecticide resistance (IR) mutations in various pest species. In most species these studies have focused primarily on SNPs as they are readily detectable. However, several genomic structural variants (SVs) have been experimentally shown to impact IR in *Drosophila melanogaster*, and we have previously observed structural variation in *Drosophila* at many genes that may be relevant to IR (e.g. cytochrome P450s). Thus, it is possible that SVs make a larger contribution to the evolution of resistance than is currently visible without the use of long-read sequencing data. To this end, we sequenced 7 *C. quinquefasciatus* specimens collected from Puerto Rico (*n*=3) and Zambia (*n*=4) with PacBio HiFi and constructed high-quality assemblies for each. We found a large number of SVs in this data set, with each genome exhibiting tens of thousands of large insertions and deletions relative to the reference genome. Roughly half of these were shared with at least one other genome in our sample. IR genes were located near or within many SVs, including several cytochrome P450 genes. Many of these variants would have gone undetected using short-read approaches, underscoring the importance of long-read sequencing for comprehensive population genetic analyses of adaptation to pesticides and other strong selective pressures, not only in insects but across the tree of life.

281W Differentiating shared ancient origin and recent admixture for genetic similarities between populations

Eaaswarkhanth Muthukrishnan, Ji Tang, Charleston W.K. Chiang University of Southern California, Keck School of Medicine

To date, Native Hawaiians (NH) are one of the most admixed ethnic minorities in the United States, with over 70% of individuals reporting multiple ancestries. Among the major non-Polynesian ancestry components in NH are Asian ancestries, reflecting the long-standing presence of East Asian (Japanese and Chinese) and Island Southeast Asian (Filipino) populations in Hawai'i. Here, using a series of analyses based on inferred local ancestries, we wish to distinguish whether genetic similarities between NH and Filipinos largely reflect shared ancient Austronesian (ANS)-like ancestry or post-colonial admixture. Using genotype data from 183 Filipino participants living in Hawai'i from the Multiethnic Cohort (MEC), we first characterized Filipino population structure through unsupervised admixture analysis. We identified a dominant ancestry component (averaging 95%) shared across Filipino individuals that is likely Austronesian in origin, alongside variable East Asian and European contributions consistent with documented historical contacts. To model genetic ancestry in NH, we compared a four-way ancestry model (Polynesian [PNS]-, East Asian-, European-, and African-like ancestries) to a five-way model that additionally included Filipinos as a reference for ANS-like ancestry. We found that approximately 42% of markers assigned as PNS-like ancestry in the four-way model were reclassified as ANS-like ancestry in the five-way model. Moreover, all NH individuals were estimated to harbor a minimum of 17% ANS-like ancestry, consistent with ANS-like ancestry representing a basal component shared across the NH population. Furthermore, the distribution of ANS-like ancestry segment lengths was shifted toward shorter segments (mean = 1.8 cM) compared to European-like ancestry segments (mean = 2.5 cM), arguing against post-colonial era gene flow as the primary source of this ancestry. Notably, ANS-like ancestry segments most frequently misclassified as PNS-like segments were predominantly short (<1 cM), supporting a deeper coalescent origin rather than recent admixture. Together, these results suggest that while some ANS-like ancestry in NH likely reflects recent admixture with Filipino populations, the majority of the genetic similarity between Filipinos and NH may derive from shared ancestry associated with the Austronesian expansion through Island Southeast Asia in the ancestors of Polynesians.

282T Pangenome reveals diverse structural variants underlying repeated evolution of herbicide resistance in *Amaranthus tuberculatus*

Jacob Montgomery, Julia Kreiner Ecology and Evolution, University of Chicago

Herbicide resistance exemplifies convergent evolution, arising repeatedly and independently through diverse molecular mechanisms across species and populations. Yet the role of structural variation in this convergence, particularly whether complex genomic rearrangements evolve repeatedly and yield similar or distinct architectures, remains poorly understood. While previous work has suggested structural variation to be responsible for herbicide resistance in the agricultural weed *Amaranthus tuberculatus*, its structure and origins remain unresolved due to constraints of short read resequencing. Here, we utilize long-read and HiC sequencing to produce *de novo*, chromosome-level, phased genome assemblies for 15 individual diploid plants (30 total haplomes) derived from landscape-wide collections. We use these assemblies to construct a pangenome graph that captures extensive structural variation across the genome. Here, we focus on the duplication of the *5-enolpyruvylshikimate-3-phosphate synthase* (*EPSPS*) gene, whose protein is targeted by the herbicide glyphosate. We identify *EPSPS* duplicated in diverse structural contexts, including within tandem arrays on primary chromosomes and on extrachromosomal circular DNA cassettes of varying size and structure. Integrating this pangenome with short-read resequencing from 450 range-wide samples will reveal whether this diverse structural variation derives from a single origin with stepwise divergence or represents unique mutational events. Finally, by mapping the frequency of these variants across the range, we will work to identify ecological and management factors driving their relative spread. By characterizing structural variation across populations, this study reveals how complex structural variants evolve and spread during rapid adaptation.

283T A Simple Genetic Basis for Necrosis in *Mimulus guttatus*

Erin Howard, Matthew Farnitano, Andrea Sweigart Genetics, University of Georgia

Genetic divergence between geographically distant populations can result in reproductive incompatibilities, which may represent the early building blocks of speciation. Identifying the genetic mechanisms of these barriers can tell us about the evolutionary forces that shape populations, and whether divergence is driven primarily by local adaptation or random drift. Here, we aim to determine the genetic basis of severe hybrid necrosis between geographically distant populations of *Mimulus guttatus*; crosses between these populations result in F1 progeny with brown leaves, highly stunted growth, and early plant death. We self-fertilized the few F1s that flowered to generate F2s, and approximately 25% of them exhibited necrosis. Using these F2s, we have genetically mapped the phenotype to two interacting loci. Additionally, we are performing genetic crosses between *M. guttatus* collected from throughout the species range to understand the geographic distribution of this strong F1 barrier. Ultimately, this work will help illuminate the evolutionary pressures driving divergence and speciation among *M. guttatus* populations and across all taxa.

284T The genetic basis of two melanic pigmentation traits in *Xiphophorus nezahualcoyotl*

Lyle Given[1], Tristram Dodge[1], Molly Schumer[1,2] [1]Stanford University, [2]Howard Hughes Medical Institute

Understanding the genetic architecture of sex-linked traits has been a longstanding challenge. Recent advances in long-read sequencing enable complex regions of the genome to be assembled and investigated, but how genes on sex-chromosomes affect sex-linked traits is still unknown for many non-model taxa. For many species in the genus *Xiphophorus*, diverse melanic pigmentation patterns are sex-linked and thought to be underpinned by an unknown macromelanophore determining locus called Mdl. Past molecular studies have proposed the gene *Xmrk*, which causes melanoma in some species, also drives variation in melanic patterns. However, other studies did not detect *Xmrk* in all species with sex-linked macromelanophore patterns, casting doubt on the role of *Xmrk* in pattern formation. Revisiting this question with genomic data, we find for the first time that *Xmrk* is present and tightly associated with macromelanophore patterns in *X. nezahualcoyotl*, a species where *Xmrk* was previously thought to be absent. We characterize two macromelanophore patterns, “spotted side” and the previously undescribed “marmoratus,” and quantify phenotypic variation between and within these patterns. Using GWAS in a wild population, we map these patterns to a small genomic region containing *Xmrk*. In analyzing long-read assemblies of spotted individuals, we find the sex chromosomes vary in gene content and higher-order repeat structure and uncover variation in the coding and noncoding region near *Xmrk*. Moreover, we find the *Xmrk* haplotype associated with spotted side is linked to the X and Y chromosomes while marmoratus is potentially limited to the Y chromosome. Our results clarify longstanding controversy surrounding the genetic architecture of sex-linked traits in *Xiphophorus*. We showcase how phenotypic variation can be linked to dynamic genomic regions, which may contribute to the evolution of diverse sex-linked traits.

285T Re-interpreting Deleterious Variant Burden in Populations with Elevated Runs of Homozygosity

Leqi Tian[1], Dorcus Malomane[1], Ji Tang[1], Malak Abedalthagafi[2], Charleston Chiang[1] [1]University of Southern California, [2]Tufts University

Runs of homozygosity (ROH) are extended genomic regions of consecutive homozygous genotypes arising from identity-by-descent due to recent parental relatedness, historical population bottlenecks, or geographic isolation. Previous studies have reported enrichment of deleterious variation within ROH in human populations. Using variants in the top and bottom 5th percentile of deleteriousness inferred by Genomic Pre-trained Network (GPN) scores, we observed similar patterns in Saudi Arabians. However, this enrichment was only detectable when counting homozygous deleterious genotypes, driven by systematic differences in frequencies between deleterious and presumed neutral alleles. When counting by the number of variants, deleterious variants are not enriched, if not somewhat depleted, in ROH compared to neutral variants. To further investigate if differences in demographic histories impact the burden of deleterious variation in ROH, we performed forward simulations under two demographic scenarios while sampling additive selective coefficients from published distributions of fitness effects: (1) a consanguinity model reflecting Saudi Arabian mating practices, and (2) an Out-of-Africa bottleneck followed by population expansion without sustained consanguineous mating. We matched the simulated populations under the two models by nucleotide diversity (π) to enable direct comparison between models.

As expected, the consanguinity model produced more and longer ROH than the bottleneck model. Moreover, when evaluated by the number of deleterious variants in ROH, we found that individual-level increases in ROH coverage were associated with greater depletion of deleterious variants under both demographic scenarios. This depletion was driven by long (and more recent) ROH, and is notably more pronounced under the consanguinity model compared to the bottleneck model. This pattern suggests that purging of deleterious recessive alleles is more efficient under sustained consanguinity, where repeated homozygosity exposes greater deleterious burden to selection. The effect diminishes in shorter ROH, consistent with the expectation that ancient, weakly deleterious variants—which accumulate in short ROH segments—are less efficiently purged. These findings highlight how the demographic origin of ROH influences the landscape of deleterious variation and may explain discrepancies between homozygous genotype-based and variant-based enrichment analyses.

286T Transgenerational inheritance of thermal tolerance reveals generationspecific geneexpression responses to environmental change

Alison Hall, Melissa Pespeni Biology, University of Vermont

Rapid environmental change elicits immediate plastic responses, inherited transgenerational effects, and longerterm evolutionary divergence. Recently, we showed that epigenetic and genetic mechanisms contribute in complementary but largely distinct ways over 25 generations of adaptation. However, we lack direct tests of how thermal tolerance phenotypes and geneexpression responses are transmitted across early generations following a single episode of environmental stress. Here, we used a multigeneration splitbrood experiment and measured thermal tolerance and geneexpression phenotypes over four generations. We found that a single generation of development at elevated temperature (22°C; no mortality) increased thermal tolerance, which was inherited for two subsequent generations despite a return to ambient conditions (18°C). The most dramatic transcriptional differences occurred in the second generation, when offspring of heatexposed parents were returned to ambient conditions, revealing substantial regulatory and physiological adjustments driven by environmental transition. Notably, firstgeneration expression changes were enriched for chromatin organization and RNA catabolism, whereas secondgeneration changes involved mRNA localization and developmental processes, suggesting that inherited states and developmental context jointly shape stress responses. Integration of these transgenerational transcriptional phenotypes with genetic and epigenetic divergence identified in longterm experimental evolution will enable tests of whether early inherited regulatory states persist, transform, or decouple from the molecular targets of adaptation. Together, this work provides rare, generationbygeneration resolution of inherited stress responses and establishes a framework for linking shortterm transgenerational effects with longerterm evolutionary change.

287T Investigating the role of interspecific introgression in flowering time variation in *Capsella bursa-pastoris*

Maya Wilson Brown, Emily Josephs Plant Biology, Michigan State University

A central goal of evolutionary genetics is to understand the demographic and selective processes that contribute to trait variation within a species. In tetraploid *Capsella bursa-pastoris*, population structure and variation in the onset of flowering has been investigated significantly in the species' ancestral range of Eurasia and North Africa. In the ancestral range, *C. bursa-pastoris* in the Northern Eurasian group has a documented history of introgression from diploid congener, *Capsella rubella*. Additionally, previous association studies have identified candidate genes associated with flowering time variation in the Northern Eurasian group and East Asian group. However, the genetic basis of early flowering in the Mediterranean group is still unknown. This study will investigate the genetic basis of flowering time variation in *C. bursa-pastoris* within the introduced North American range and will assess whether introgression from *C. rubella* contributes to this trait variation in both groups. We find that Mediterranean and Northern Eurasian genotypes are represented in the United States and both groups have introgressed genetic material from *C. rubella*. Using linear mixed models, we aim to identify loci associated with variation in flowering time and characterize interspecific introgression at these loci. The results of this study will provide a greater insight to the impact of interspecific introgression on the evolution of this globally distributed species.

288T A diffusion framework for the steady-state allele frequency distribution under fluctuating selection and population size

Hisham AA Ali, Kunaal Joshi, Michael Lynch Biodesign Institute, Arizona State University

Previous theory has characterized the steady-state distribution of allele frequencies under the assumption that microevolutionary processes remain fixed over time. Extensions of this framework have incorporated fluctuating selection, but rely on a caveat of quasi-neutrality: although an allele may experience selection at any given moment, it is effectively neutral when averaged over long timescales. Similarly, fluctuations in census population size can be accommodated through effective population size estimates, but only within models that treat selection as constant.

Here, we develop a theoretical framework to calculate the steady-state allele frequency distribution under *simultaneously fluctuating* population size and selection, without invoking quasi-neutrality. We derive novel drift and infinitesimal mean coefficients for the diffusion approximation that jointly account for these dynamics. We validate our theory using forward simulations of populations at steady state experiencing rapid, generation-to-generation fluctuations in both selection and population size. Under the assumption of rapid fluctuations, these simulations demonstrate that our theory accurately predicts the steady-state allele frequency distribution.

Our theory enables inference of past demographic history from the site frequency spectrum without assuming constant selection, thereby broadening the scope of demographic inference under realistic evolutionary dynamics.

289T The trans-Pacific voyage that shaped the pattern of genetic variation and enrichment of functional alleles in Native Hawaiians

Ji Tang[1], Eaaswarkhanth Muthukrishnan[1], Bryan L. Dinh[2], Xin Sheng[1], Fei Chen[1], Kekoa Taparra[3,4], Lynne Wilkens[5], Loïc Le Marchand[5], Christopher A. Haiman[1], Charleston W.K. Chiang[1,2] [1]Center for Genetic Epidemiology, Department of Population and Public Health Sciences, University of Southern California, [2]Department of Quantitative and Computational Biology, University of Southern California, [3]Department of Health Policy and Management, Fielding School of Public Health, University of California, Los Angeles, [4]Department of Radiation Oncology, David Geffen School of Medicine, University of California, Los Angeles, [5]Epidemiology Program, Cancer Center, University of Hawai'i, Honolulu

The evolutionary history of a population can shape the allelic architecture of complex traits and disease risk. An example is the Native Hawaiians (NH) population, which has experienced serial bottlenecks and recent admixture. Using whole-genome sequencing (WGS) data from 1,065 NH individuals from the Multiethnic Cohort (MEC) study, we modeled their ancestry as a recent four-way admixture and defined the Polynesian (PNS)-like ancestry as the dominant pre-colonial component on the Hawaiian Islands. We then conducted three lines of investigation linking this demographic history to the distribution of complex phenotypes and disease risk in NH. First, using identity-by-descent segments while accounting for admixture, we inferred a stronger bottleneck in the PNS-like ancestry of NH than other ancestries, with an effective population size as low as ~600 individuals, possibly reflecting the combined effect of colonization and previous serial bottlenecks during trans-Pacific voyages. Second, applying our recently developed genealogy-based method, we detected fine-scale substructure within the PNS-like ancestry, consistent with long-term isolation. This structure shows geographic variation across islands and correlates with Austronesian ancestries predating the founding of the NH population. Third, we observed enrichment of a class of predicted functional alleles due to this demographic history. Compared with WGS data from an additional 9,656 MEC participants, we identified ~5 million alleles that are rare (<0.5%) in individuals from other ancestries but common (>5%) among NH. Driven by the PNS-like ancestry, NH individuals carry 15–27% fewer rare and 4–19% more common functionally important alleles than other MEC individuals ($p = 5\times10^{-3}$). Among functional alleles with the greatest enrichment (>20% higher frequency) between PNS-like and other continental ancestries, 76 (and 56) alleles associated with increasing (and decreasing) risk of multiple diseases. Taken together, these results suggest how demographic history may have shaped genetic variation and contributed to molecular and phenotypic diversity among Native Hawaiians.

290T Disentangling the biogeography of multiple fermentative yeast species

Yu-Chen Yeh[1,2], Cheng-Ju Yang[1], Wei-Ting Chien[1], Yu-Ching Liu[1], Huei-Mien Ke[3], Isheng Jason Tsai[1,2] [1]Biodiversity Research Center, Academia Sinica, [2]Genome and Systems Biology Degree Program, National Taiwan University, [3]Department of Microbiology, Soochow University

Broadleaf Fagaceae forests in East Asia are one of the geographical origins of the baker's yeast *Saccharomyces cerevisiae*. However, the environmental drivers of adaptation in other yeast species remain unresolved. This project leverages comparative population genomics to infer the evolutionary histories of multiple yeast species across the Saccharomycetaceae.

We enriched and sequenced the genomes of 311 new fermentative yeast strains from the natural forests in Taiwan, Vietnam, the Philippines and Japan. Accounting for an estimated 113 million years of divergence between the species examined (*Lachancea fermentati*, *Lachancea kluyveri*, *Lachancea thermotolerans*, *S. cerevisiae, Saccharomyces kudriavzevii* & *Torulaspora delbrueckii*). we analyzed a total of 955 isolates in order to quantify inter- and intra-specific diversities. Population structures inferred from single nucleotide polymorphisms revealed novel lineages for all six species, including earliest diverged lineages for *Lachancea thermotolerans*, *Lachancea fermentati*, and *Torulaspora delbrueckii*. Notably, we identified recurrent phylogeographic splits between Okinawan and Taiwanese yeasts. The discovery of wild populations sister to anthropogenic clades in *L. thermotolerans* and *L. fermentati* suggests independent domestication events within these species. Regional genomic diversity (π) confirmed that East Asia harbored the highest value for *S. cerevisiae*, *Saccharomyces kudriavzevii*, *L. thermotolerans*, and *L. fermentati* while *Lahcancea kluyveri* exhibited greater diversity in the Americas.

As yeast populations support the model of diversification with respect to the geographical patterns, we are interrogating shared allele frequency changes on orthologous genes to detect parallel evolution. This multispecies framework facilitates the association of genomic variation with ecological context across diverged fermentative yeast species in nature.

291T Heavy-tailed dispersal and its impact on local adaptation: Revisiting the British peppered moth

Sherif Negm, John Novembre Human Genetics, University of Chicago

Local adaptation arises when selection varies across space, and gene flow is limited. The melanism cline of the British peppered moth (Biston betularia) is a canonical example, yet it sparked controversy: classical migration–selection models could not fit the Liverpool–Wales transect using selection coefficients from mark–release–recapture (MRR) studies (Bishop 1974), and fits were often rescued by invoking inflated dispersal and non-visual selection advantage (Cook 1980; Mani 1980); this fueled doubts about the classical interpretation (Coyne 1998; Hooper 2002). We argue the mismatch reflects a modeling limitation—assuming Gaussian dispersal—rather than a failure of the selection narrative. We derive a general approximation for migration–selection clines under broad classes of symmetric dispersal distributions, including leptokurtic/heavy-tailed distributions, and under arbitrary spatial selection gradients. We develop a maximum-likelihood framework that tests for heavy-tailed dispersal and jointly estimates the tail-decay exponent and a local dispersal scale parameter from spatial allele-frequency clines. We connect the approximation to previous theoretical work done on the topic of clines, including existence conditions in Slatkin (1973) and Nagylaki (1975), and recover the Gaussian diffusive limit as a special case. Applied to the historical transect, our method fits the moth data using the MRR-measured selection coefficients, inferring realistic local dispersal and a tail-decay parameter indicative of heavy-tailed dispersal, consistent with wind-borne larval ballooning in Biston betularia (Liebert and Brakefield 1987). Our work provides evidence for a genetic signature of heavy-tailed dispersal that may matter for many systems experiencing local adaptation.

292T Convergent genome- and gene-level constraints shape repeated environmental adaptation in grasses

Sheng-Kai Hsu[1], Aimee J Schulz[2], Charles O Hale[2], Germano Costa-Neto[1], Zachary R Miller[1], Michelle C Stitzer[1], Travis Wrightsman[2], Jingjing Zhai[1], Henry D Dawson[2], Lara J Brindisi[1], Elad Oren[1], Yun Luo[1], Beatrice A Konadu[2], Jonathan O Ojeda-Rivera[1], Taylor M AuBuchon-Elder[3], Elizabeth A Kellogg[3], M Cinta Romay[1], Edward S Buckler[1,2,4] [1]Institute for Genomic Diversity, Cornell University, [2]Section of Plant Breeding and Genetics, Cornell University, [3]Donald Danforth Plant Science Center, [4]USDA-ARS, Robert W. Holley Center for Agriculture and Health

Grasses (Poaceae) dominate terrestrial ecosystems and sustain global food security, yet the genomic principles enabling their repeated adaptation to extreme environments remain unresolved. Here, we combine dense phylogenomic sampling, global environmental data, and state-of-the-art nucleotide and protein foundation models to characterize the mutational targets underlying environmental adaptation in grasses. Analyzing 707 genomes from 569 species spanning 17 climate zones, we identify dozens of phylogenetically independent transitions into extreme temperature, water, and soil environments. These repeated adaptations are accompanied by convergent shifts in genome-scale molecular properties, including Nitrogen-to-Carbon balance and biosynthetic energetic cost of the proteome, revealing predictable biochemical constraints imposed by environmental selection. At the gene level, through an AI-informed phylogenetic mixed modeling framework, we identified 330 genes that repeatedly underlie different axes of environmental adaptation, 16 of which were supported by three separate lines of evidence. Together, our results show that grass adaptation is channeled by layered constraints acting at genome-wide and gene-specific scales, producing predictable evolutionary trajectories.

293T Population-level divergence in 3D genome organization revealed by comparative chromatin contact matrix analysis in threespine stickleback

Fahad S Gilani Molecular & Cell Biology, University of Connecticut

Three-dimensional (3D) genome organization is increasingly recognized as a key layer of gene regulation, yet most studies characterize chromatin architecture using single individuals or pooled samples, limiting evolutionary inference. In particular, methods for comparing 3D genome structure across populations remain underdeveloped.

Here, we present a population-level analysis of 3D genome organization in threespine stickleback using Omni-C data from 24 individuals sampled across six independently evolved freshwater lake populations from coastal Alaska. Taking a genetics-driven perspective, we focus on population differences in chromatin interaction patterns and their potential regulatory consequences. Rather than focusing on individual chromatin features alone, we develop and apply a matrix-based comparative framework that treats chromatin contact maps as quantitative objects. Using both traditional methods such as eigenvector decomposition and insulation profiles and innovative tensor-based similarity metrics, we directly compare interaction matrices across individuals to quantify population-level structure, variance, and divergence in 3D genome organization. Notably, population-level differences in contact structure may coincide with genes implicated in transcription regulation, suggesting a link between chromatin architecture and gene expression divergence.

This approach reveals consistent differences in chromatin interaction patterns among populations, despite substantial within-population variation, demonstrating that aspects of 3D genome architecture are structured systematically between populations. Importantly, our framework enables statistical comparison of 3D genome organization across evolutionary and population replicates, moving beyond descriptive analyses of loops or domains in single genomes.

By integrating chromatin contact matrix analysis with quantitative genetics thinking, this work establishes a generalizable method for studying the evolution and comparison of 3D genome architecture in natural populations. Our approach treats chromatin interaction matrices as quantitative phenotypes and provides a foundation for future studies linking chromatin structure to genetic variation and adaptive evolution.

294T Comprehensive Genomic Landscape and Selection History of Adaptive Introgression in Peruvians

Susanna J. K. Gutierrez[1], Autumn Allemon[1], Xinjun Zhang[1,2] [1]Human Genetics, University of Michigan, [2]Ecology and Evolutionary Biology

Archaic introgression introduced novel genetic variation into modern human populations and is repeatedly implicated to facilitate local adaptations. In the Americas, including modern Peruvians, numerous loci have been reported as candidates for adaptive introgression despite the absence of direct contact between archaic humans and Native Americans. However, most existing approaches rely on demographic assumptions that incompletely capture the complex history of the Americas. As a result, these candidates include substantial false positives and false negatives, obscuring the true genomic landscape of adaptive introgression and limiting biological interpretation.

Here, we systematically re-evaluate all previously reported candidates for adaptive introgression in Peruvians using a comprehensive, multi-layered validation framework. Integrating evidence from selection scans and haplotype-based visual inspection, we narrow hundreds of putative candidates to 20 high-confidence archaic haplotypes with convergent evidence for positive selection. These include both established candidates (e.g. *WARS2*, *SCN1A*, and *SYT14)* and novel loci (e.g. *DDX6P2*, *MGAT5*, and *DAB2)*. Leveraging whole-genome sequences from modern, ancient, and archaic humans, we traced the ancestry and allele frequency trajectory of each haplotype and identified Neanderthal or Denisovan origins and the modern human source population responsible for introduction into the Peruvian gene pool. We then applied ancestral recombination graph methods to infer the strength and timing of selection acting on archaic haplotypes. Across loci, selection onset clusters into two major periods: shortly after the initial settling of Peru (< ~10 kya) and during the transition to agriculture and increasing population size (~5 kya). We report that all adaptive introgression candidates from either Neanderthal or Denisovans were introduced via the ancestral Beringian populations upon the peopling of the Americas.

Finally, we show that the reconstructed selection history of archaic adaptive variants provides independent constraints on key demographic timelines, offering a complementary route to refining models of American population history. Building on these insights, we introduce ongoing work on *MaLAdapt 2.0*, an extension of our machine learning framework that explicitly incorporates complex and uncertain American demographic histories to reduce false positives in adaptive introgression inference. Overall, this work establishes a high-confidence genomic landscape of adaptive introgression loci in Peruvians and underscores how archaic introgression introduced standing genetic variation serves as a basis for adaption in populations without direct contact with archaic humans.

295T Gene expression noise evolves more slowly and by different molecular mechanisms than gene expression in a model eukaryotic genus

Brian Metzger Purdue University

Gene expression is a stochastic process and varies among genetically identical cells in the same environment. This 'noise' in expression can be selectively beneficial in unpredictable environments but is often thought to be a nuisance that is selected against. At the same time, mutations that alter expression level also often alter expression noise. The extent to which this coupling of mutational effects on expression level and expression noise impacts the evolution of gene regulation is largely unknown. In particular, we know little about the rate at which expression noise evolves and the molecular mechanisms that it uses to do so. To address these questions, we used single-cell RNA-seq on two distantly related *Saccharomyces* yeast species. We found that expression noise was more similar between these species than gene expression. This suggests that gene expression noise is either more robust to mutations than expression level or that the strength of selection acting on expression noise is greater. To identify the molecular mechanisms underlying differences in expression noise, we also performed single-cell RNA-seq on the hybrid of the two species. We then used estimates of allele-specific expression in individual cells to identify *cis*- and *trans*-regulatory differences in expression noise. We found that expression levels and expression noise evolved via distinct molecular mechanisms, with *trans*-regulatory differences more common than *cis*-regulatory differences in expression noise. Furthermore, expression of the two alleles in a hybrid cell was often poorly correlated, suggesting independent control of the two alleles and that differences in noise are often attributable to intrinsic noise. This could arise if there were substantial evolutionary differences in the molecular basis of noise between homologous genes in the two species or if the molecular factors typically responsible for noise act independently across alleles. If the former is true, this would suggest considerable compensatory changes in expression noise given the slow absolute rate at which it evolves. By contrast, the latter explanation would be consistent with evolutionary changes in diffusible factors, such as transcription factors, being the primary source of evolutionary change in expression noise. These findings indicate that the evolution of gene expression and expression noise are often decoupled, and that estimates of differences in expression noise between species can provide mechanistic insights into the evolution of gene regulation.

296T The mechanisms maintaining a narrow and environmentally correlated cline in an avian plumage polymorphism

Vicens Vila-Coury[1], Augie Kramer[2], Sahid M.R. Bello[2], Paul R. Sweet[2], Brian T. Smith[2] [1]American Museum of Natural History & Columbia University, [2]American Museum of Natural History

The geography of local adaptation is determined by a balance between selection and gene flow. When selection is sufficiently strong, local adaptation can occur over narrow distances, despite even elevated gene flow. A population of birds on the small Eastern Caribbean island of Grenada is potentially a rare example of this possibility. In Grenada, Bananaquits exhibit two distinct color morphs, one yellow and the other black. The distribution of morphs is environmentally correlated, whereby yellow morphs are most abundant in the island's xeric scrub and black morphs dominate the island's humid uplands. Morph phenotype is determined by a derived autosomal dominant variant at MC1R. Although gene flow between morphs is common, this cline spans less than 18 km. To identify the mechanisms underlying this cline and how its patterns vary genome-wide, we prepared a de novo Bananaquit genome assembly and sequenced over 100 individual birds from Grenada. In this presentation we discuss how population structure, gene flow, and signatures of selection indicate the potential mechanisms maintaining this exceptionally narrow cline within an insular population of a highly vagile species.

297T Parallel selection across agricultural weed genomes

Acer VanWallendael Horticulture, North Carolina State University

Agricultural weeds include some of the best contemporary examples of rapid eukaryotic evolution under both natural and artificial selection, showing repeated evolution of traits such as herbicide resistance and phenology. However, it remains a challenge to diagnose the genomic, environmental, and chance factors that contribute to any particular population of a species becoming resistant, partially because most studies focus on either a single species or a single generation. To better understand how parallel selection drives genetic variation across space, we sampled tissue from five nuisance weed species with different mating systems and genome structure at twenty locations in the southeast as well as one outgroup in Michigan. These species include Amaranthus palmeri (dioecious diploid), Erigeron canadensis (native selfing diploid), Cyperus esculentus (invasive asexual holocentric), Ipomoea purpurea (invasive outcrossing diploid), and Sorghum halepense (invasive perennial tetraploid). This project is designed to be replicated yearly to follow population-level allele frequency changes over time using pooled whole-genome sequencing and annotation of rapidly evolving genes. In the first year, we found clear differences in population genetic structure across outcrossing and selfing/asexual species. We tracked genetic variation in genes related to herbicide resistance and metabolism across space to detect parallelism. These results contrast the roles of mutation and gene flow driving genetic diversity across species facing similar selection, but different life history traits. In future years, measuring directional changes in allele frequencies in these same populations will allow testing of hypotheses about the drivers of evolution rate in plants under high artificial selection pressure.

298T Shifts in target and off-target stress resistance following selection for copper resistance in *Drosophila melanogaster*

Elizabeth R Everman[1], Kazzrie A Arnold[2] [1]School of Biological Sciences, University of Oklahoma, [2]University of Oklahoma

Heavy metals are an important source of environmental and anthropogenic stress with clear effects on ecosystem health and the evolution of populations. Often, heavy metals are introduced into the environment through industry and agriculture, which raises additional concerns about risks to human health. Adaptive resistance to heavy metal stress involves shifts at multiple loci with links to genes with diverse effects, including genes that are specific to heavy metal detoxification to genes that are involved in oxidative stress response, xenobiotic resistance, and mitochondrial function. The underlying genetic complexity of heavy metal resistance presents the potential for the evolution of cross tolerance to different forms of stress. To characterize the potential for phenotypic and genetic cross tolerance following heavy metal selection, we sampled *Drosophila melanogaster* from a retired copper mine and an active commercial fruit farm and subjected replicate wild-derived laboratory populations to selection for copper resistance over many generations. We tracked phenotypic shifts in copper resistance of all replicated cages and found substantial (up to 16-hour) increases in response to selection. We explored the effects of copper selection on resistance to other non-target heavy metals (lead and cadmium) as well as on non-metal stress resistance (starvation resistance) and on life history traits (longevity and fecundity). Overall, we found clear evidence that copper selection can shape multiple stress and life history traits. Ongoing genomic analyses will allow characterization of the genomic architecture as it is shaped by copper selection over time. We will ultimately explore the potential for pleiotropy between target and off-target responses to selection to begin characterizing the complex genomic responses to selection that influence multiple traits.

299T Tandem Repeat Variation near Signatures of Selection

Myco Gabrielli I. Torres, Carolina de Lima Adam, Rori Rohlfs University of Oregon

Tandem repeats (TRs) are segments in DNA composed of repeating motifs and exhibit high mutation rates. Changes in TR characteristics, such as the number of repeats, can affect transcription and downstream traits. Long-read sequencing allows for further systematic study of the role of TRs in an evolutionary context. While SNP-based methods are most commonly used to identify regions of the genome with signatures of selection, it remains unclear how TRs might underlie these signals. Here, we used TR and SNP data from 46 samples of the Human Pangenome Reference Consortium (HPRC) to identify and characterize TRs near signatures of natural selection. For each TR, we calculated the local SNP-based Tajima's D. Through this analysis, we hope to identify how patterns of TR variation might differ in regions under selection by considering how TR diversity varies with local Tajima's D. Identifying these characteristics will further elucidate the role of TRs in human natural selection and how TR variation is shaped by selection.

300T Testing Network-Based Predictions of Genetic Constraint in Melanogenesis

Yemko Pryor[1], Kirk Lohmueller[2], Arslan Ziadi[3], Tina Lasisi[1] [1]Anthropology, University of Michigan, [2]University of California, Los Angeles, [3]University of Minnasota

A gene's position within biological networks should predict its tolerance for genetic variation. Core, highly connected genes with broad pleiotropy face stronger constraint than peripheral genes with specialized functions. However, this network-constraint relationship remains largely untested at the pathway level using population genetic data. We use melanogenesis, a well-characterized pigmentation pathway with 400-500 genes, to test whether network position predicts patterns of genetic constraint and recent selection across human populations.

We integrate a curated melanogenesis network (spanning biosynthesis, trafficking, and regulatory functions) with constraint metrics including loss-of-function intolerance (LOEUF), evolutionary conservation (PhyloP), nucleotide diversity (π), and selection statistics (PBS, iHS) computed across global populations from the 1000 Genomes Project and Human Genome Diversity Project. . Our preliminary analysis of LOEUF scores supports the network-constraint hypothesis: pigment-specific enzymes (e.g., TYR, TYRP1) are uniquely tolerant (median LOEUF = 1.89), while signaling and developmental regulators are highly constrained (median LOEUF = 0.30–0.36), with complete separation between genes causing syndromic disorders (max LOEUF = 0.66) and those causing isolated pigmentation phenotypes (min LOEUF = 0.86). To distinguish constraint imposed by network position from constraint shaped by phenotypic selection, we compare African and Melanesian populations, which share deeply pigmented phenotypes achieved through independent evolutionary paths We test three hypotheses: (1) genes with higher network centrality show greater constraint regardless of population; (2) genes with broader tissue expression show distinct constraint patterns reflecting pleiotropy; (3) population-specific selection signatures concentrate at peripheral network positions while core pathway genes remain universally constrained.

This framework synthesizes polygenic selection theory and mutation-selection-drift balance to understand how network architecture shapes functional variation. By testing network-aware predictions of genetic constraint, this work moves beyond locus-by-locus analysis toward pathway-level evolutionary models applicable to any complex trait.

301T tsNN: Learnable Embeddings for Ancestral Recombination Graphs

Amjad Dabi[1], Murillo Rodrigues[2], Nathaniel Pope[1], Andrew Kern[3] [1]Biology, University of Oregon, [2]Oregon National Primate Research Center, [3]Biology, Universtiy of Oregon

Ancestral Recombination Graphs (ARGs) have been a topic of recent interest in the field of population genetics due to their ability to represent genetic variance in a compact manner as relationships between ancestral haplotypes. Currently, no methods exist to predict features of the edges or nodes comprising the structure of an ARG. While machine learning has enabled state-of-the-art predictions in population genetics for problems such as demographic inference or the detection of selection, one of the hurdles of developing such methods for ARGs is finding flexible and useful lower-dimensional vector representations (embeddings) that allow for downstream inference of parameters of interest. We developed an architecture based on a message-passing algorithm that operates directly on ARGs with a custom CUDA kernel for efficiency. Using this architecture, we developed a two-step training strategy that first produces embeddings by predicting fine-scale features and sample population origin as a 'pretext task' and subsequently uses these embeddings for downstream inference of demographic parameters. We show that this architecture learns well on the pretext tasks and produces useful embeddings for the downstream task. Our results demonstrate the utility of graph-based machine learning methods for ARG-based inference.

302T Spatiotemporal genomics of adaptation to human-mediated selective pressures in Amaranthus tuberculatus

Rozenn Pineau, Julia Kreiner University of Chicago

Understanding the predictors shaping past and current patterns of genetic diversity is critical for accurate forecasting of evolutionary dynamics. In agroecosystems, plant populations face various human-mediated selective pressures, such as herbicide application and land-use change, that vary across space and time and interact with rapidly changing climatic conditions. Predicting allele frequency change in this context is a challenging problem considering heterogeneity across scales, multivariate selective pressures, and the non-linearity of gene-environment relationships.

To address these challenges, we are studying the genetic basis and temporal dynamics of adaptation to agricultural stressors under climate change in Amaranthus tuberculatus, a native plant turned problematic agricultural weed. In recent work, we conducted a drought selection experiment and identified 43 loci associated with adaptation to drought in contemporary populations. By integrating genomic data from herbarium specimens, we resolved the role of fluctuating selection in the maintenance of climate adapted alleles across the past century. Building on this work, we are extending our analyses to finer spatiotemporal resolution and a broader set of changing climatic and land management practices. To this end, we have curated a high resolution herbarium spatiotemporal dataset spanning much of the US in the last century. By developing gradient forest modelling on spatiotemporal datasets, and leveraging a prior known sets of herbicide, drought, and agriculturally adaptive loci, we plan to identify the role of changing climate, land-use, pesticide, and soil variables in structuring contemporary genomic landscapes. We will then use these models to project the degree of local adaptation and invasion of future pest weed populations under various land-use, land-management, and climate change scenarios. Ultimately, we hope such a framework will provide a powerful approach to forecasting genomic responses to ongoing, human-mediated environmental change across landscapes.

303T Cross-species analysis reveal the quantitative relationship between context-dependent mutation spectrum and genome composition

Yijie Wang, Ziyue Gao Department of Genetics, University of Pennsylvania

Mutations are the ultimate source of genetic variation, shaping genome composition over evolutionary time. However, the quantitative relationship between mutation spectra and genome composition remains poorly understood.

We modeled evolution of the genome multi-nucleotide composoition using a transition matrix derived from context-dependent mutation spectrum. With this model, we found that the ratio of genome abundance between two trinucleotide contexts correlates with the corresponding ratio of mutation rates (P_{XYZ} / $P_{XY'Z}$ ≈ $\mu_{XY'Z>XYZ}$ / $\mu_{XYZ>XY'Z}$), which means that genome composition can serve as a proxy for the underlying mutation spectrum.

Using observed genome compositions in plants, we observed significant but weaker correlations, suggesting the influence of additional evolutionary forces. We found the predicted trinucleotide proportions correlate stronger with the observed values in genomic regions with low recombination rates; incorporating differential fixation probabilities for strong-to-weak or weak-to-strong mutations in the transition matrix also substantially improved model performance. These findings indicate that GC-biased gene conversion (gBGC) plays a critical role in shaping genome composition in addition to mutation spectrum. Notably, compared across species, some ratios of trinucleotide frequencies correlate strongly with corresponding ratios of mutation rates inferred from polymorphisms.

Principal component analysis of 64 trinucleotide proportions across coding, genic noncoding, and intergenic regions in 69 plant species revealed the greatest compositional variation in the intergenic regions, indicating differences in the mutation spectrum and/or gBGC. Differences in GC content and CpG depletion primarily drive this variation, with Poaceae species exhibiting particularly high GC content and less CpG depletion. Finally, phylogenetic regression of trinucleotide ratios on gene copy variations identified two candidate genes that may modify mutation spectra in plants.

Taken together, these results establish genome composition as an informative proxy for the underlying mutation spectra, while highlighting gBGC as a key evolutionary force that modulates this relationship. This proxy may enable the discovery of novel modifiers of mutation spectra without dependence on polymorphism data.

304T Patterns of diversity in lifecycles with extended haploid phase: A coalescent approach

Marcy Uyenoyama[1], Jason Mohabir[2] [1]Biology, Duke University, [2]Duke University

Many organisms (e.g., fungi, algae, ferns, bryophytes) exist as haploids for much of their life cycle. Methods for the evolutionary analysis of genomic diversity in such organisms have generally been borrowed from population genomic theory developed for the animal model, which comprises an extended diploid somatic phase and a virtually instantaneous haploid phase. We present a coalescent framework tailored to the life cycle of some fungi, including Saccharomyces and Cryptococcus, which bear important implications for food production and disease in human populations.

Among the major consequences of an extended haploid phase is a reduction in the rate crossing-over. Even with some genetic exchange during mitosis, assortment of genomic regions is greatly reduced or precluded, causing the entire genome to segregate as a single bloc. While crossing-over may resume upon transition to the diploid phase, associations among even unlinked regions decay asymptotically over the course of several meiotic events. With respect to determinants of mating type in particular, the switching of genomic regions among mating types occurs at geometric rates.

To explore these and other implications of extended haploid phase, we have developed a coalescent approach to the determination of likelihoods based on observed levels of neutral diversity. Of particular interest are differences between the patterns of diversity exhibited by Cryptococcus populations with restricted diploid phase and Saccharomyces populations with greater diploid phase. Among our findings is that diversity in the haploid phase depends on mating type, even at loci unlinked to mating type.

305T Asymmetry in cross-sex cross-trait genetic covariances and the evolvability of sexual dimorphism

Jacqueline Sztepanacz, Mathieu Videlier University of Toronto

The evolution of sexual dimorphism is predicted to resolve conflict that can arise from divergent evolutionary interests between sexes, enabling each sex to reach its fitness optimum. However, most of the genome is shared between sexes, which can lead to a genetic constraint for dimorphism evolution. Most studies of intersexual genetic constraints have focused on the effect of genetic correlations, $r_{mf,}$ for single traits. However, multivariate studies of the **B**-matrix of intersexual genetic covariances suggest that sexual dimorphism may be more evolvable than inferred from r_{mf} because of the potential for indirect responses to selection from correlated traits. To comprehensively address this question, we collected and re-analyzed published estimates of **B** using a recently developed approach to quantify the evolvability of sexual monomorphism and dimorphism. We find that across the traits and species we study, the evolvability of dimorphism is lower than that of monomorphism, but also that sexually concordant and antagonistic selection are almost equally capable of producing dimorphism. We also find that asymmetry in **B** would affect the response to selection more in females than in males. Our results show that sexual dimorphism is more evolvable than studies of r_{mf} suggest and underscore that sexually antagonistic selection is not required for the evolution of sexual dimorphism.

306T Contrasting patterns of protein-coding and regulatory evolution associated with convergent phenotypes in birds

Tim Sackton Harvard Univ

Convergent evolution, where a similar trait evolves independently in unrelated lineages, provides a powerful natural experiment to study adaptation by comparing genomic signatures of rapid evolution across lineages that share a convergent trait. Birds are an increasingly valuable system for this work, given the wealth of phenotypic information available and rapidly growing genomic resources. Here, I contrast signatures of positive selection associated with two convergent phenotypes in birds: sugar-rich diets (focal species: hummingbirds, sunbirds, parrots, and honeyeaters) and reduced tarsus (hindlimb) length (focal species: penguins, swallows, kingfishers, and bulbuls). In each case, we identify signatures of selection in both genes and putative regulatory regions by comparing genomes across species with and without the convergent phenotype. Among nectar-feeding birds, we find evidence for positive selection in all sugar-specialist lineages in the transcription factor MLXIPL, and experimentally validate enhanced transcriptional response to sugar in the hummingbird MLXIPL protein. Among species with reduced relative tarsus length, we see little evidence for convergent signals of positive selection in protein-coding genes, but uncover strong convergent signals of selective shifts associated with putative regulatory elements near key developmental genes involved in hindlimb and skeletal development. Overall, these results contribute to a broader understanding of the genomic targets of selection underlying repeated phenotypic evolution.

307T Mutational diversity of a dsRNA virus evolved under increasing temperatures

Sonia Singhal, Parnian Pour Bahrami, Sujaya Jayatirtha Nilogal San Jose State University

One of the hallmarks of global climate change is the rapid rate at which it is occurring. Theoretical models predict that the rate of environmental change will affect evolutionary trajectories as the population adapts to the changing environment. A sudden environmental change, for example, represents an initial, large change in selective pressure, and is expected to result in the accumulation of a few mutations of large effect. In contrast, a gradual environmental change imposes an initially small change in selective pressure that then increases as the environment becomes more stressful. Gradual environmental change is thus expected to result in the accumulation of many mutations of small effect. We tested this theoretical model empirically by evolving replicate populations the lytic dsRNA bacteriophage phi-6 under high-temperature heat shocks that increased to a maximum of 50 deg. C, either gradually over 32 transfers or suddenly over a single transfer. We evaluated survival of the populations to high temperatures, and the appearance of new mutations over time. All populations increased their survival to the high-temperature heat shock. However, Gradual populations had a significantly higher average survival than Sudden populations on their first exposure to the highest temperature, indicating that gradual environmental change allowed populations to pre-accumulate thermostabilizing mutations. Sanger sequencing of four lineages (two Gradual, one Sudden, and one Control that did not experience heat shocks) revealed that mutations tended to arise sequentially (i.e., second mutations arise in the background of a first mutation). One Gradual lineage, however, showed turnover of genotypes, where a genotype that had fixed by Transfer 24 was then replaced by new mutants in the final 8 transfers. These results suggest that different mutations may be favored at different points in time during a gradual environmental change. We continue to sequence lineages to confirm these patterns in other replicates, and are using a deep sequencing approach with Nanopore technology to track dynamics of lower-frequency mutations.

308T Rapid Adaptation and Extinction In Synchronized Outdoor Evolution Experiments of Arabidopsis thaliana

Xing Wu[1], Tatiana Bellagio[2], Yunru Peng[3], Lucas Czech[4], Meixi Lin[3], Moises Exposito-Alonso[3] [1]Department of Integrative Biology, University of California, Berkeley, [2]Stanford University, [3]University of California, Berkeley, [4]Carnegie Science

Climate change is threatening species with extinction, and rapid evolutionary adaptation may be their only option for population rescue over short ecological timescales. However, direct observations of rapid genetic adaptation and population dynamics across climates are rare across species. To fill this gap, we conducted a replicated, globally synchronized evolution experiment with the plant Arabidopsis thaliana for 5 years in over 30 outdoor experimental gardens with distinct climates across Europe, the Levant, and North America. We performed whole-genome sequencing on ~70,000 surviving reproductive individuals and directly observed rapid and repeatable adaptation across climates. Allele frequency changes over time were parallel in experimental evolution replicates within the same climates, while they diverged across contrasting climates, with some allele frequency shifts best explained by strong selection between -46% to +60%. Screening the genome for signals of rapid climate adaptation identified a polygenic architecture with both known and novel adaptive genetic variants connected to important ecological phenotypes, including environmental stress responses, CAM5 and HEAT SHOCK FACTORs, and germination and spring flowering timing, CYTOCHROME P450s and TSF. In this experiment, accessions performed best in climates resembling their native habitats, providing evidence of climate-local adaptation. We modeled this phenomenon as a Gaussian function of climatic distance with two interpretable parameters, Wmax (peak fitness at the climatic optimum) and Vs (breadth of adaptation), revealing a generalist–specialist trade-off. Both parameters showed a genetic basis, indicating that evolutionary outcomes are, in part, encoded in the genome. Our model improved the prediction of evolutionary change by 22.1% over standard approaches. Nevertheless, evolutionary adaptation trajectories were not always predictable and varied among sites, suggesting limits and potential tipping points beyond which adaptation fails. Together, these results suggest that rapid climate adaptation is possible, but understanding its limits across species will be crucial for biodiversity forecasting.

309T A Coalescent Framework to Model Pathogen Evolution as a Metapopulation

James Crescenzi, Alex McAvoy, Parul Johri Genetics, University of North Carolina at Chapel Hill

Many pathogenic and microbial species undergo complex life cycles wherein they experience drastic bottlenecks during invasion into a host and rapid exponential growth within hosts. The effect of such a complex population history on patterns of genetic variation is not currently understood. We employ a coalescent framework to represent a pathogen population as a metapopulation where each deme represents a host and incorporate the effects of transmission and recurrent bottlenecks. We obtain analytical expressions for pairwise time to coalescence in the metapopulation and expected differentiation between hosts. We find that recurrent bottlenecks rescale the coalescent process within hosts to the size of the bottleneck, reducing within-host variation. We then examine the distribution of time to coalescence when migration rates are low and find that it deviates significantly from the Wright-Fisher process when the number of hosts is of similar order to the size of the bottleneck. Otherwise, the pairwise process under our model converges in distribution to the Wright-Fisher process. Simulations of the site frequency spectrum (SFS) show that the SFS under our model deviates from the Kingman coalescent (in particular, is skewed towards singletons) when the number of hosts is of similar order to the bottleneck. Our results suggest that the inference of selection or demographic history from pathogen genetic data may be systematically biased by the mechanics of pathogen reproduction, indicating a need for further theoretical work that can inform inference.

310T Population Structure, Ancestry, and Demographic History of House Mice in the Northeastern United States

Zade R Alafranji[1], Rene Clark[1], Stephen Kupchella[1], Adrienne Kasprowicz[1], Logan Lacy[1], Samantha Giancarli[1], Nina Mallalieu[1], Megan Phifer-Rixey[2] [1]Biology, Drexel University, [2]Biology, Drexel Univeristy

House mice (*Mus musculus domesticus*) have migrated globally in association with European colonists, including eastern North America where they have arrived in the last ~500 years. The details of their arrival and their source ancestry, however, remain unclear. To approach this problem, we generated moderate coverage whole genome sequencing data for mice from urban and rural localities in three major metro areas (New York City, NY, Philadelphia, PA, and Richmond, VA). Combining this data with publicly available European *M. m. domesticus* samples, we characterized the population structure of these populations expected to represent early expansion of house mice in the United States. Using both phylogenetic and demographic modelling approaches, we also investigated the ancestry of North American populations. Our data suggest that the sampled North American populations are most closely related to mice from northwestern Europe (e.g. the United Kingdom and Germany). However, NY and PA populations likely have significant contributions from other European regions while mice from Richmond, VA are sister to populations from northwestern Europe. Finally, we modelled the demographic history of each locality with top models for both urban and rural populations suggesting that Richmond diverges from the other two localities first and the timing of divergence is consistent with the hypothesis of arrival with European colonizers. In summary, wild house mice on the East Coast of North America have a complex ancestral history which likely has been driven by human migration.

311T Uncovering Intraspecific Variation in the Repetitive Architecture of the Drosophila melanogaster Y Chromosome

Adolfo A Delgado, Mahul Chakraborty Biology, Texas A&M University

The Drosophila melanogaster Y chromosome is a paradigm of assembly complexity due to its dense satellite DNA and repetitive architecture. By employing a repeat-aware merging strategy and reference-guided scaffolding, we generated highly contiguous Y assemblies for the reference strain ISO1 and two isogenic strains A3 and A4. These assemblies added ~2.5 Mbp of novel sequence and quadrupled contiguity relative to the current reference, providing a high-resolution framework to investigate intra-specific structural variation. We find dramatic strain-specific shifts in satellite and ampliconic composition despite a high total repeat burden (85.3%–93.8%) across all strains. Notably, the expansion of the AATAC satellite in A3 and A4 underscores the rapid turnover of repetitive DNA even between members of the same species. Our results further establish the Y chromosome as an active "genomic sink" that recruits and maintains genes from across the genome. While we observed a conserved core of complete copy gene translocations shared by all three strains, we also uncovered significant strain-specific acquisitions. The A4 Y chromosome uniquely captures partial copies of the autosomal gene pAbp. In contrast, the ISO-1 Y chromosome lacks unique gene acquisitions, positioning it as a structural subset of the more diverse A3 and A4 haplotypes. Unique to our study, all three strains share a translocation of CG40635, a CK2 regulator linked to the Stellate genes on the X chromosome. Finally, we improved the resolution of the Y-linked SuSte and rDNA ampliconic arrays, situating these historically fragmented loci within large, contiguous scaffolds near their expected cytological contexts. While SuSte clusters exhibit distinct copy-number variation across lineages, they maintain high sequence homogeneity and a shared structure. Similarly, read-depth analysis of the rDNA arrays shows substantial scaling in total sequence length between A3 and A4, showcasing expansions and contractions of ampliconic arrays between strains. Targeted assembly of rDNA units reveals that while array size varies dramatically between strains, the coding consensus remains strikingly conserved, with structural variation concentrated almost exclusively within the repetitive intergenic spacers. These findings illustrate a Y chromosome in constant flux, where the rapid structural turnover of repetitive and ampliconic sequences coexists with the stringent conservation of essential coding units.

312T Inferring unknown spatial locations with genome-wide genealogies

Alexander Lewanski[1], Gideon Bradburd[2] [1]Integrative Biology, Michigan State University, [2]Ecology and Evolutionary biology, University of Michigan

Many organisms disperse and reproduce in a spatially localized manner, which results in the widely observed pattern of isolation-by-distance. The spatially structured nature of many populations can be leveraged, using large, georeferenced genetic datasets, to infer the geographic location of a genotyped individual of unknown provenance. Geographic assignment using genetic data can be a powerful tool for identifying the sources of epidemiological outbreaks, tracking the spread of invasive species, or localizing poaching activity. Here, we introduce an approach that uses an ancestral recombination graph to infer the origins of individuals with unknown locations. We extend this approach to infer not just one, but a set of spatial origins for an individual, which is relevant to situations where an individual has recent ancestors from multiple, distant locations across the species' range and could be used to identify those ancestor locations. This genealogy-based location assignment method will have relevance for numerous topics across ecology, evolution, conservation, epidemiology, and forensics where location information for an unknown sample is necessary.

313T The Extent of Sparsity In The Genotype-Phenotype-Fitness Map

Reza Ghelich[1], Mohammad H Donyavi[2], Kara Schmidlin[1,3], Kerry Geiler-Samerotte[1] [1]School of Life Science, Arizona State University, Biodesign Center for Mechanisms of Evolutio, [2]School of Biology, Arizona State University, Biodesign Center for Mechanisms of Evolutio, [3]Arizona State University

For decades, geneticists have debated the extent of pleiotropy: Does a single mutation affect all traits simultaneously (as in Fisher's Geometric Model), or is cellular biology organized into semi-independent modules? While selection for modularity is a theoretical staple to explain how organisms evolve without breaking existing functions, empirical evidence quantifying pleiotropy in the genotype-phenotype-fitness map remains rare. We address this by applying Sparse Structure Discovery (SSD), an interpretable, parsimony-based computational method, to high-dimensional datasets where the relative fitness of hundreds of adaptive yeast mutants was measured across many diverse environments. SSD decomposes variation in fitness into a small number of inferred phenotypic components or modules, then trims weaker connections between genotype and phenotypic modules to maximize model accuracy while minimizing pleiotropy. Here, we study how the inferred degree of pleiotropy depends on data composition, including the number of samples, genetic backgrounds, environments, and the nature of the selective pressure. We find that pleiotropy varies with evolutionary context. In particular, mutants evolved under glucose limitation exhibit sparse effects, each influencing only one or two inferred phenotypic modules. However, mutants evolved under drug selection are substantially more pleiotropic and affect many phenotypic modules. Together, these results may suggest that long-term, recurrent selective pressures, such as glucose limitation, shape cellular organization into semi-independent modules, such that mutations can affect one module without broadly perturbing others. In contrast, rare or evolutionarily novel pressures, such as drug exposure, may lack such pre-existing modular structure, leading mutations to propagate system-wide effects and manifest as increased pleiotropy.

314T Differential adaptation and genic divergence related to drought tolerance among rattlesnakes within a phylogenetic framework

Austin Biddy[1], Griffin McDaniels[1], Derek Benson[2], Kenro Kusumi[2], Dale DeNardo[2], Greer Dolby[1] [1]University of Alabama at Birmingham, [2]Arizona State University

Taxa have adapted to arid environments by using behavioral, physiological, and/or molecular mechanisms to reduce dehydration. The Sonoran Desert is a good natural laboratory for investigating how these mechanisms individually or in combination facilitate survival, because the desert exhibits a west-east precipitation gradient: the west receives winter rain and the east receives precipitation in the winter and summer. Thus, western species endure longer droughts and scarcer water resources overall. Within a phylogenetic framework, we studied the behavioral, physiological, and molecular mechanisms underlying drought adaptation across the Sonoran Desert in *Crotalus* rattlesnakes comprising three co-occurring lineage pairs with varying genetic divergences: *C. atrox* (between populations), *C. cerastes* (between subspecies), *C. tigris* and *C. pyrrhus* (between species). We hypothesized that lineage pairs with less divergence would rely more on behavioral or epigenetic mechanisms, whereas more diverged lineage pairs would additionally have protein coding adaptations to drought. We analyzed 109 wild individuals across the gradient, generating whole genome sequence (WGS), transcription (RNA-seq), chromatin accessibility (ATAC-seq), whole genome methylation, and physiological data. Preliminary PCA results from ATAC and WGS data reflect west/east differentiation, and pairwise F_{ST} values follow known phylogenetic distances. Across seasons, western species have low evaporative water loss and eastern species have a wider range in osmolality. As expected, only the sister species comparison had strong coding sequence changes. Based on McDonald-Kreitman and dN/dS, these loci were in genes involved in chromatin remodeling/transcription (*Polr3d*, *Kansl3*), skin barrier (*Stra6*, *Alox12b*), metabolism (*Ncor1*, *Scap*), and circadian rhythm (*Hcrtr2*, *Plch2*). Gene expression data from skin tissue show differential expression in chromatin remodelers, hinting at genetic and transcriptional modifications underlying epigenetic mechanisms of drought tolerance, though work is ongoing. Epidermal barriers may relate to water loss prevention through the skin. Differences in precipitation seasonality likely cue behavior differences, suggested by selection in circadian regulator genes. So far, we show increased divergence in transcription and genic adaptation with population divergence, underscoring there are multiple mechanisms and tactics *Crotalus* rattlesnakes use to persist in arid environments.

315T Modeling Early Neural Tube Morphogenesis in Organoids as a Comparative Framework for Human Brain Evolution

Breanna Takacs, Soojin Yi University of California Santa Barbara

The human brain is remarkably expanded relative to other primates, a difference thought to reflect evolutionary changes in early neurodevelopment. A proposed driver of this expansion is an increase in neuroepithelial founder cell populations during early development, potentially linked to regulatory differences across species. These cells arise from the neural tube during neurulation, a morphogenetic process that establishes the earliest organization of the central nervous system. Neurulation follows a broadly conserved sequence of events across vertebrates. However, how variation in neural tube morphogenesis across primate lineages coincides with the establishment of gene regulatory and epigenetic states remains unclear, in part due to limited access to great ape embryos. Here, we present the first organoid model that faithfully recapitulates chimpanzee neural tube morphogenesis, enabling direct morphological and molecular comparisons of neurulation between humans and our closest evolutionary relatives. This platform provides a new opportunity to examine early morphogenesis alongside emerging gene regulatory and epigenetic states, and how their divergence across primate lineages may contribute to human brain evolution.

316T Are inversion mechanisms stable across mammalian genomes?

Lucia N Ramirez, Rafael F Guerrero Department of Biological Sciences, North Carolina State University

Chromosomal inversions are important drivers of genome evolution, influencing recombination rate, gene expression, and gene flow. Recent work has shown inversions are caused by at least three different mechanisms, including nonallelic homologous recombination (NAHR), non-homologous end joining (NHEJ), and replication-based errors. The evolutionary implications of these mechanistic differences are still unknown, including how they correlate to other important inversion properties (like size), or how their relative distributions vary across different taxa. Here, we attempt to infer the mechanistic origins of inversions fixed between closely related species, using a data set of publicly available mammal genomes. We test for associations between mechanism type and inversion size, age, and TE content, and quantify phylogenetic signal in the relative abundance of each mutation type. We expect to find that NAHR-derived inversions are larger (which would make them more likely to evolve by capturing coadapted genes), and have a higher proportion of repetitive elements, while NHEJ-derived inversions will be smaller and lack TE enrichment. With this work we hope to bridge the mechanistic and comparative view of inversions, and lay the foundation for more precise characterization of these mutations across the tree of life.

317T Insertion of an invading retrovirus regulates a novel color trait in swordtail fish

Nadia Haghani, Molly Schumer Stanford University

For over a century, evolutionary biologists have been motivated to understand the mechanisms through which organisms adapt to their environments. Coloration and pigmentation are remarkably variable within and between species and can serve as an important window into the mechanisms of adaptation. We mapped the genetic basis of a newly described iridescence trait in swordtail fish to a single locus. Individuals with this trait appear to sparkle as they move through the water. We find that the trait is driven by the recent endogenization of a retrovirus that inserted near the gene alkal2a. This insertion is associated with changes in the chromatin landscape, upregulation of alkal2a, and accumulation of iridescent cells that adhere to the scales. Rather than causing diseases, our results demonstrate that invading endogenous retroviruses can also regulate novel trait variation in the host. Moreover, we find that this coloration trait may act as an important signal in interactions between fish and their predators in the natural environment. Strikingly, another closely related species from this region in Mexico has evolved a phenotypically similar iridescence trait through distinct genetic and developmental mechanisms. We hypothesize that comparison of this trait across species may yield exciting insights into the genetic mechanisms underlying the evolution of novel traits.

318T Quantifying the Genome-Wide Effect of Polygenic Selection over a Single Generation

Gabriele M Sgarlata, Graham Coop Department of Evolution and Ecology, University of California, Davis

Genomic variation is shaped by neutral and selective evolutionary forces such as genetic drift and selection. While their relative contribution is actively debated, how short-term evolutionary change plays out remains relatively unaddressed. Most genetic evidence of short-term selection has been based on the identification of major allele frequency (AF) changes at few loci with large selective advantage. Yet most traits are polygenic, where short-term selection response is shaped by thousands of loci with small AF changes that are difficult to distinguish from drift. Here, we quantify the genome-wide effect of polygenic selection over one generation using the signal mediated by the random association [linkage disequilibrium (LD)] of loci with a large set of selected loci across the genome. We rely on the idea that loci in stronger LD with selected loci show greater variance in AF change than expected under drift. We derive expressions relating variation in LD among loci to the variance in AF change due to linked selection and drift, and leverage a mathematical framework that quantifies their relative contribution to a single generation of AF change. To demonstrate our approach, we decompose the genome-wide variance in AF change in the UK Biobank using AF changes predicted from the fitness proxy of the number of children, showing that there is a small percentage of AF change due to linked selection even within a generation. This framework would allow to investigate the short-term, genome-wide effects of polygenic selection across a wide range of species.

319T Does polygenic adaptation leave a molecular trace?

Dandan Peng[1], Matthew W Pennell[2], Michael D Edge[1], Joshua G Schraiber[1] [1]University of Southern California, [2]Cornell University

Integrating short-term and long-term evolutionary processes into a unified framework has been a central goal in evolutionary biology. While theory suggests there is no fundamental gap between these processes, with macroevolution viewed as the result of accumulated microevolutionary changes over long periods of time, contradictory patterns between these two scales are still observed. To investigate the connection between molecular signatures and phenotypic divergence, we simulated the evolution of a polygenic trait under stabilizing selection. We tracked the synonymous and nonsynonymous mutations in a structured chromosome under varying selection strength and shift magnitudes in phenotypic optima. Our simulations reveal that under neutrality, the genomic patterns are dominated by negative selection. When a phenotypic shift is introduced, a detectable burst of adaptive substitutions occurs only when the shift it extreme. Furthermore, this molecular signal is transient, as the population adapts to the new optimum, the substitution rate rapidly returns to the normal level. These findings demonstrate that for polygenic traits, substantial phenotypic adaptation can occur with minimal lasting signatures of accelerated molecular evolution, challenging straightforward interpretations of divergence data.

320T An *FT* paralog causes photoperiod divergence between closely related *Mimulus*

Logan Scott[1], Sam Mantel[2], Andrea Sweigart[1] [1]Genetics, University of Georgia, [2]Botany & Plant Pathology, Purdue

In plants, flowering time is often adapted to local environmental conditions and, when shifted, can directly induce temporal reproductive isolation. *Mimulus guttatus* and *Mimulus nasutus* are closely related yellow monkeyflowers that often co-occur in secondary contact across western North America. *Mimulus guttatus* is an outcrossing species with flowers that bloom from late spring into summer under longer days, while *M. nasutus* is a self-fertilizing species with small flowers that bloom in the spring under shorter days. Previous work showed that divergence in critical photoperiod between these species is controlled by two major loci; here, we identify *FLOWERING LOCUS T(FT)-like* as one of the causal genes. FT-like acts as a suppressor of flowering in these species: the gene is expressed in *M. guttatus* under short days when it remains vegetative, but it is only lowly expressed in *M. nasutus*, which flowers readily under the same conditions. Additionally, using *Agrobacterium*-mediated transformation, we show that overexpressing the *M. guttatus* allele of *FT-like* in *M. nasutus* suppresses flowering entirely. We are currently testing whether CRISPR-mediated disruptions to *FT-like* in *M. guttatus* de-repress flowering under short days, resulting in a flowering phenotype similar to that of *M. nasutus*. In crop species, the *FT* gene family responds to environmental cues to tightly regulate flowering. Our work provides strong evidence that divergence in the function of *FT* paralogs maintains species barriers in natural settings by creating distinct photoperiodic responses.

321T Contribution of locally adapted variation to adaptive potential in experimental cages of *Drosophila melanogaster*

Jamie C. Freeman, Emmah Daanen, Joy Zhang, John E Pool Genetics, University of Wisconsin-Madison

Genetic diversity is central to the ability of a population to adapt to its environment, but neutral genetic diversity does not always have a simple relationship with fitness. Local adaptation of populations to distinct environments provides a potential reservoir of fitness relevant variation. Experimental cages comprised of *Drosophila melanogaster* from Zambia, highland Ethiopia, and France were founded with equal numbers of inbred lines from one, two, or all three of the source populations. In separate experiments, these 18 cages were selected over 10 generations against two distinct selection pressures: survival on a toxic food source dosed with a detergent and survival in a wounding assay. Survival and population size metrics were measured over the 10 generations of selection. For both experiments, the cages founded with lines from a single geographic source had smaller population sizes after 10 generations of selection than cages founded with two or three geographic sources.

322T Genome-Wide Alignments Reveal Turnover And Retention Of De Novo Genes In Brassicaceae

Adekola O Owoyemi[1], Sierra Sanders[2], Catherine Chaison[2], Claudio Casola[1] [1]Ecology & Conservation Biology, Texas A and M University, [2]Texas A and M University

De novo genes originate from previously noncoding DNA and represent a major source of lineage-specific genetic novelty. However, their study has been hampered by high false-positive rates and limited ability to identify reliable enablers substitutions. We developed a novel computational pipeline that minimizes undetected homology errors and annotation artifacts and rely on novel genome-wide multiple alignments generated with Progressive Cactus. We applied this framework to a comparative genomic analysis of 21 Brassicaceae species to investigate patterns of de novo gene emergence and retention. Focusing on *Arabidopsis thaliana*, we identified over 1,000 candidate de novo genes, including approximately 600 de novo-like genes lacking detectable synteny and 305 *A. thaliana*-specific de novo genes supported by enabling mutations in syntenic noncoding regions. We also identified 141 older de novo genes shared across multiple species, constituting the first systematic set of retained de novo genes in plants. Our results indicate substantial turnover, consistent with other eukaryotes. Gene and protein features, including length, codon adaptation index, and intrinsic disorder, were similar between de novo and de novo-like genes, suggesting similar origination mechanisms and evolutionary trajectories. Older de novo genes exhibited similar features of young genes, indicating that de novo gene retention might be associated with characteristics present at birth. Unexpectedly, de novo genes encode significantly fewer transmembrane proteins than conserved genes, contrasting with yeast and human systems. Together, our study establishes a robust framework for investigating de novo gene evolution in plants and reveals lineage-specific evolutionary trajectories.

323T Impact of *I* Element Activation on Meiotic Recombination

Diane Nguyen, Justin Blumenstiel Kansas University

Transposable elements (TEs) are mobile genetic elements that influence genome evolution by contributing to genome size variation and genetic variation that can be harmful but also fuel adaptation. Transposition can also pose a threat to genome stability by causing chromosomal breaks. In meiosis, programmed double-stranded breaks are repaired through homologous recombination and are the substrates of meiotic crossing over. The *I* element in *Drosophila melanogaster* is of particular interest because it creates non-programmed DNA breaks during meiosis through its transpositional activity.

The *I* element is a 5.4 kb non-LTR retrotransposon of the LINE family and related to human LINE-1 elements. It has experienced waves of activity in *D. melanogaster* and now occurs widely, with its activity regulated by maternally inherited piRNAs. When these piRNAs are lacking, as in *I-R* hybrid dysgenesis, transposition is reactivated in the germline, generating double-stranded breaks (DSBs), a severe form of DNA damage requiring repair through homologous recombination or non-homologous end joining.

Because meiotic recombination itself is a DNA repair process initiated by programmed DSBs, *I* element–induced breaks may disrupt the balance and spatial regulation of crossing over. This project investigates how TE-derived DSBs influence the formation and distribution of meiotic crossovers in dysgenic versus non-dysgenic females. Specifically, it asks how an element active during meiosis modifies crossover frequency and interference, and whether transposon-derived breaks can act as alternative recombination substrates.

Initial data suggest differences in crossover patterning between dysgenic and control females, hinting that TE activation may increase interference or shift recombination landscapes. This study aims to uncover how TE-induced DNA damage interacts with meiotic processes, revealing links between transposon activity, DNA repair, and the evolution of recombination.

324T Genome wide segregation distortion reveals the complexity of reproductive isolation in recently diverged *Mimulus* species

Natalie M. Gonzalez Genetics, University of Georgia

Hybrid seed lethality is a powerful and common postzygotic barrier among closely related flowering plant species, yet its genetic and evolutionary causes remain poorly understood in wild systems. Here I investigate the genetic basis of hybrid seed lethality in closely related species of *Mimulus* (Yellow Monkeyflower) using mapping populations derived from reciprocal backcrosses. With this crossing design, I have discovered patterns of transmission ratio distortion consistent with parent of origin effects, a hallmark of F1 hybrid seed lethality. Additionally, I find evidence of cytonuclear incompatibilities and F1 sterility, and I identify putative pollen viability QTL. These results reveal that even in closely related species of *Mimulus*, multiple genetic incompatibilities can contribute to strong postzygotic reproductive isolation.

325T Genomic insights into Lewontin's Paradox: investigating the discrepancy between population sizes and genetic diversity in waterfow

Jeffrey L. Peters Biological Sciences, Wright State University

A fundamental prediction in population genetics is that neutral genetic diversity will be positively correlated with population size. Yet across taxa, empirical patterns consistently violate this expectation: species differ in population size by many orders of magnitude, but exhibit only modest variation in genetic diversity. Here, I investigate this long-standing discrepancy—Lewontin's Paradox—using a large-scale, comparative dataset of >3,000 loci sampled from 41 species of globally distributed ducks. Among these species, census populations sizes vary by more than 40,000-fold whereas nucleotide diversity only varies 20-fold. Using coalescent analyses and Approximate Bayesian Computation, I demonstrate that demographic processes, such as fluctuations in *N*, divergence times, and interspecific hybridization, have had a measurable contribution to, but cannot fully explain, the observed discord. Multiple lines of evidence support a prominent role of selection in contributing to this discord—the correlation between population size and genetic diversity was weaker for (1) coding exons compared to noncoding introns and intergenic regions, (2) conserved regions of the genome compared to regions with high diversity, and (3) loci linked to the Z sex-chromosome, which is more often exposed to selection in the hemizygous state, than autosomal loci. Furthermore, the heterogeneity observed among marker types was significantly greater for species with large *N*, which is consistent with the expectation that selection is more effective in larger population sizes. Overall, this large-scale, comparative dataset yielded quantitative evidence that both neutral and nonneutral processes have contributed to Lewontin's Paradox.

326T Leveraging a system of Poecilia hybrids to study trait genetic architecture and speciation

Sophia Haase Cox[1], Tristram Dodge[2], Kelsie Hunnicutt[2], Rhea Sood[2], Theresa Gunn[2], John J Baczenas[2], Stepfanie Aguillon[3], Manfred Schartl[4,5], David Reznick[6], Molly Schumer[2] [1]Biology, Stanford University, [2]Stanford University, [3]UCLA, [4]University of Würzburg, [5]Texas State University, [6]UC Riverside

Poecilia reticulata, the common guppy, is well known for its extreme intraspecific variation in male coloration and life history traits. Extensive research on *P. reticulata* populations in northern Trinidad has shown that male pigmentation and a suite of life history phenotypes rapidly and repeatably evolve in response to predation regime. While most ecological research has been centered in Trinidad, *P. reticulata* has a broad natural range across northern South America. In northern Venezuela, *P. reticulata* co-occurs with sister species *P. wingei*. *P. reticulata* and *P. wingei* can produce viable offspring in lab crosses and are thought to hybridize in the wild, but the strength of reproductive isolation between species and potential negative consequences in hybrids are unknown. We plan to leverage the hybridization of these species to investigate evidence of hybrid incompatibilities and perform QTL mapping of pigmentation and life history traits that differ between the two species. By producing crosses between various strains of *P. wingei* and *P. reticulata*, we aim to characterize the extent and genetic basis of reproductive isolation, and identify if pigmentation or life history traits of known ecological importance contribute to incompatibilities. As a first step, we used Oxford Nanopore Technology (ONT) long read sequencing to generate a high-quality reference genome assembly of *Poecilia wingei.* This research will provide insight into whether rapidly evolving traits under high selection within a species can also serve as important contributors to the formation of reproductive isolation between diverging lineages.

327T The counteracting effects of hitchhiking and repulsion of deleterious alleles during a selective sweep

Maya Weissman[1], Nandita R. Garud[1,2], Kirk E. Lohmueller[1,2] [1]Ecology and Evolutionary Biology, UCLA, [2]Human Genetics, UCLA

Deleterious mutations are abundant in nature. As a consequence, when an adaptive mutation arises, it will likely land on a chromosome with several deleterious mutations. While deleterious mutations may hitchhike with the adaptive variant, potentially leading to elevated linkage disequilibrium (LD), deleterious mutations also experience repulsion, or Hill-Robertson interference. Both hitchhiking and repulsion shape patterns of LD, but their interaction during a selective sweep is still not fully understood. Here, we investigate patterns of LD during a sweep using simulations across a wide range of parameters, including strength of selection. We find that LD will be higher between intermediate frequency nonsynonymous variants compared to synonymous variants because initially rare deleterious nonsynonymous variants private to the haplotype bearing the adaptive allele can only reach high frequency due to hitchhiking. Synonymous variants, on the other hand, can reach high frequency due to neutral drift. However, further away from the sweep, the effect of hitchhiking cannot counteract the effect of closeby deleterious variants repelling each other. As a result, LD is depressed among nonsynonymous variants relative to synonymous variants at short distances between deleterious variants. The combined effects of hitchhiking and repulsion can result in a characteristic “criss-crossing of LD”, or non-uniform patterns of LD among nonsynonymous vs synonymous sites depending on the distances between deleterious variants. We observe these patterns in empirical sequencing data from the human gut microbiome, validating that criss-crossing of LD is a characteristic of real-world sweeps. Together, these results highlight a complex dynamic between hitchhiking vs repulsion of deleterious alleles during a selective sweep, revealing novel signatures that can be leveraged for increased power in selective sweep detection.

328T Evaluating Robustness and Parameter Sensitivity in ROH Inference

Raya Esplin[1,2], Tessa Ferrari[3], Jazlyn Mooney[3], Ellie Armstrong[2] [1]San Diego State University, [2]University of California-Riverside, [3]University of Southern California

In small, isolated populations, including those managed for conservation or captive breeding, genomic estimates of inbreeding are widely used as a proxy for genetic health. Runs of homozygosity (ROH) allow researchers to identify recent and historical inbreeding, however ROH detection tools can produce discordant results. These discrepancies arise from differences in underlying algorithms, assumptions about heterozygosity and linkage disequilibrium, and sensitivity to user-defined parameter choices. Here, we systematically evaluate how ROH caller choice and parameterization influence inferred inbreeding. We consolidate popular ROH callers (PLINK, BCF/VCFtools, GARLIC, ROHan, and RZooROH) into a unified analytical pipeline and apply them to simulated genomes with known ROH distributions. We establish baseline performance by running each method with minimal filtering. Then, we tune parameters with simulated data to improve concordance with known ROH and assess performance across callers. Finally, we apply the calibrated pipeline to empirical genomic data. We observe substantial variation in the number, length, and genomic distribution of ROH inferred under minimal filtering. Thus, our results indicate a strong sensitivity to parameter choice across all callers and underscore the critical need for filtering. While parameter tuning improves agreement with simulated ROH, performance remains caller-dependent. Based on these results, we provide filter recommendations and advocate the use of multiple ROH callers to improve inbreeding inference. Our findings clarify sources of uncertainty in ROH-based metrics and inform best practices for genomic inbreeding assessment in small and isolated populations.

329T Examining the features of diet underlying signatures of convergent molecular evolution in mammals

Wynn K Meyer[1], Michael Tene[1], Kathleen Foley[1,2], Amelia Harralson[1], Alexander Seaver[1], Derrick Um[3], Harshal Wasnik[1]
[1]Biological Sciences, Lehigh University, [2]University of Iowa, [3]Lehigh University

Since their presumed insectivorous origin, mammals have adapted to a wide array of diets, varying in the content of macromolecules, nutrients, and potential toxins. In turn, selective pressures related to dietary changes have influenced mammalian metabolism, behavior, sensory perception, and immune responses. Many features of mammalian diet exhibit convergent evolution, with multiple independent lineages evolving similar dietary phenotypes. This makes diet a strong candidate for phylogenetic genotype-to-phenotype approaches, in which we identify genes that have repeatedly experienced changes in selective pressure associated with changes in diet. We recently applied such approaches to identify protein-coding genes whose evolutionary rates covary with mammalian diets defined in several different ways, including trophic level categories (herbivore, omnivore, carnivore) and quantitative representations of animal versus plant content. For these analyses we leveraged large genome-wide datasets from hundreds of mammals generated by the Zoonomia consortium, as well as information about mammalian dietary strategies from the EltonTraits database. Our results highlight several lipid and amino acid metabolism genes and pathways that show strong and robust signatures of association between molecular evolutionary rate and diet, regardless of how it is defined. We also see some signatures that differ between mammalian lineages specializing on invertebrate food sources and those that consume primarily vertebrates. For example, vertebrate-eating mammals appear to experience stronger purifying selection on genes related to RNA virus immunity and steroid hormone metabolism than invertebrate-eating mammals, suggesting that vertebrate diets pose challenges related to the pathogens and hormones ingested along with lipids and proteins. Using additional public diet phenotype resources, we further explore which quantitative features of diet are most strongly associated with molecular evolutionary rates for the top genes from our initial analyses. Our analyses provide a deeper understanding of how different aspects of this suite of complex phenotypes influence selective pressures across mammalian lineages. They also highlight how subtle differences in phenotype definition can affect outcomes of phylogenetic genotype-to-phenotype associative studies.

330T The spud, the bad, and the ugly: polyploidy, domestication, and deleterious alleles

Sapphire Coronejo, Laura M Shannon University of Minnesota

Simulations suggest that autopolyploidy results in relaxed selection on individual alleles allowing for the accumulation of deleterious variants. Section *Petota* with over 100 species ranging from diploid to hexaploid including cultivated potato, presents an ideal system for investigating the relationship between ploidy and deleterious alleles. Specifically, we compare deleterious alleles in *Solanum candolleanum*, a wild diploid, *S. tuberosum ssp andigena*, a cultivated diploid, and commercial cultivated US potato from *S. tuberosum ssp tuberosum*. Historically, deleterious alleles have been detected using a combination of evolutionary rate profiling and functional annotation. New large language learning models such as Plant Caduceus, offer alternate strategies for prediction of deleterious alleles. We compared Plant Caduceus v2 to more traditional approaches and found that Plant Caduceus indicated an excess of fixed deleterious alleles in the wild species, which was an unexpected result. We hypothesized that this was due to a lack of representation of *Solanaceae* genomes in the training model. To test this we added solanaceous plants to the 65 species included in the Plant Caduceus model. We produced a series of fine-tuned models with the addition of: (1) *S. lycopersicum*, (2) *S. candolleanum*, (3) *S. lycopersicum* and *S. melongena*, (4) *S. lycopersium* and *S. candolleanum*, (5) S. *candolleanum* and *S. melongena* and (6) *S. lycopersicum*, *S. melongena*, *Capsicum annum*, and *Nicotiana tabacum*. Both the first and second model improved our ability to detect deleterious variants in *S. tuberosum tuberosum* as measured by the distribution of zero-shot scores and the comparison of minor allele frequencies for synonymous and non-synonymous variants. We uncovered a range of overlapping variants in wild, cultivated, diploid, and polyploid potato.

331T BLInG: Enabling Population Genomic and Evolutionary Studies of Unicellular Organisms Through High-Throughput Single-Cell DNA Sequencing

Parker Crossland, Kara Schmidlin, Leandra Brettner, Kerry Geiler-Samerotte Arizona State University

Unicellular organisms are essential for understanding major evolutionary transitions, including the origin of eukaryotes, emergence of multicellularity, evolution of sex, and complex symbioses. Recent discoveries of novel eukaryotic lineages—including uncultured protists serving as sister groups to multicellular lineages—have transformed our understanding of genome evolution, speciation, and adaptation. However, many phylogenetically pivotal organisms remain impossible to culture, creating barriers to genomic characterization and population-level studies. While single-cell transcriptomics has advanced rapidly, single-cell genomics has lagged behind, limiting investigations of population structure, evolutionary dynamics, and fitness landscapes in these critical systems.

We present BLInG (Barcode Ligation for Intracellular Genomics), a novel single-cell DNA sequencing method that eliminates the need for physical cell isolation—a fundamental constraint of existing technologies that is incompatible with fragile protists, environmental assemblages, and symbiotic consortia. BLInG employs a paradigm shift: rather than separating cells into wells or droplets, we perform whole-genome amplification inside intact cells within mixed populations, then use combinatorial split-pool barcoding to assign unique molecular identifiers to each cell's DNA. This approach aims to enable simultaneous sequencing of thousands of cells while maintaining single-cell resolution and haplotype phasing, potentially increasing throughput and reducing costs compared to isolation-based methods.

BLInG's ability to preserve haplotype information could make it a valuable tool for studying epistasis, fitness effects of mutation combinations, and the genetic architecture of adaptation in experimental evolution systems. The method has potential to facilitate population genomic analyses of previously inaccessible lineages, enabling investigation of population history and demography, contemporary evolution in natural populations, host-microbiome interactions, and comparative genomics across major evolutionary transitions. By enabling analysis of unculturable organisms, BLInG could accelerate discovery of genetic innovations underlying evolutionary transitions, variations in genetic codes, chromosomal dynamics, genome evolution, and the origins of cellular complexity across the tree of life.

332T Genomic differentiation between northern and southern California voles in regions of low recombination

Isaac J Linn[1], Chris J Conroy[1], Michelle M Davila[2], Eric D Enbody[3,4], Lydia L Smith[1], Rauri CK Bowie[1], Michael W Nachman[1] [1]Museum of Vertebrate Zoology, UC Berkeley, [2]UC Berkeley, [3]Department of Computational Biology, Cornell University, [4]Department of Biomolecular Engineering, UC Santa Cruz

Genomic regions of low recombination are expected to display higher levels of differentiation between natural populations under several different theoretical models. Despite this expectation, the relationship between recombination rate and differentiation has been examined in only a few empirical systems. The widespread California vole (*Microtus californicus*) is composed of two clades which are isolated by hybrid male sterility. Using a high-quality reference genome and whole-genome sequence data from 95 individuals that were generated as a part of the California Conservation Genomics project, we find that northern and southern clades are genetically distinct, but with some evidence of admixture. We also find that across the genome, Fst is higher on the X chromosome and in genomic regions of low recombination. However, Dxy is similar in these comparisons, suggesting that patterns of differentiation are driven less by differences in gene flow than by the effects of selection at linked sites. Comparisons between the X chromosome and the autosomes, as well as the distribution of allele frequencies in regions of low recombination, suggest that observed patterns are more consistent with genetic hitchhiking than with background selection.

333T Ancestry-Specific Genetic Risk Factors for Severe COVID-19 in an Italian Cohort

Marisol Fermin Flores, Alessandro Lisi, Alexandria Koedel, Ryan Johnson, Michael C Campbell Human and Evolutionary Biology, University of Southern California

SARS-CoV-2, the virus that causes COVID-19, spread to almost every region of the world, infecting millions and resulting in substantial global mortality. Italy emerged as one of the earliest and most severely affected epicenters of the pandemic, experiencing intense transmission and tens of thousands of deaths during the initial wave of infection. While epidemiological and demographic factors contributed to disease burden, the role of host genomic variation in COVID-19 susceptibility and severity remains incompletely understood. To investigate the genomic basis of COVID-19–related clinical outcomes, we analyzed genome-wide genetic data and clinical phenotypes from 1,141 laboratory-confirmed COVID-19–positive individuals in Italy, spanning in age from 18 to 99 years. More explicitly, we first characterized population structure using complementary ancestry inference approaches, followed by genome-wide association analyses of quantitative traits associated with COVID-19 severity. Based on our analyses, we found that individuals in our dataset exhibited mainly European ancestry. However, a subset possessed ancestry originating from the Middle East, Asia, sub-Saharan Africa, and the Americas, reflecting Italy's complex demographic history which has been shaped by migration and admixture. Leveraging these genetic patterns, we identified multiple suggestive and genome-wide significant variants of primarily European origin associated with key clinical biomarkers of severe COVID-19, including neutrophil-to-lymphocyte ratio, fibrinogen, and D-dimer, known indicators of systemic inflammation and coagulopathy that drive severe disease. Furthermore, the effect sizes (β) for the most significantly associated variants ranged from -0.44 to 0.39, indicating modest but biologically meaningful contributions to inter-individual variation in phenotypic outcomes. Together, these results provide new insight into the genetic architecture underlying COVID-19 severity, highlighting ancestry-specific genetic factors that may influence inflammatory and thrombotic responses to SARS-CoV-2 infection. Additionally, this study underscores the importance of integrating population genetic structure into genomic analyses of infectious disease outcomes.

334T A pedigree study in rhesus macaques reveals a large fraction of early embryonic germline mutations

Anastasia Stolyarova[1], Milind Jagota[1], Ronald Bontrop[2], Ivanela Kondova[2], Tomas Marques-Bonet[3,4,5,6], Yun S Song[7], Jeffrey D Wall[8,9], Molly Przeworski[1] [1]Columbia University, [2]Biomedical Primate Research Centre, [3]Institute of Evolutionary Biology (UPF-CSIC), PRBB, [4]Catalan Institution of Research and Advanced Studies (ICREA), [5]CNAG-CRG, Centre for Genomic Regulation (CRG), [6]Institut Català de Paleontologia Miquel Crusafont, [7]University of California, [8]Oregon Health and Science University, [9]California Academy of Sciences

Germline mutations arise in a succession of developmental stages, including in early embryogenesis, within a cell lineage that gives rise to the germline, and during gametogenesis. Disentangling the relative contributions of different stages remains challenging. All else being equal, the contribution of early embryonic mutations should be higher in species that reproduce earlier.

With this consideration in mind, we analyzed mGAP, a repository of genomic data from multi-generational cohorts of rhesus macaques, a primate species that reproduces after a decade on average. Based on patterns of allele sharing, we identified 437 parent–offspring trios. Looking for variants present in blood samples from the offspring but absent in both parents, we found 8,921 germline de novo mutations, the largest such set in a non-human primate. As in humans, the majority (73%) of mutations accumulate in the paternal germline, at a similar rate of approximately 1.3 mutations per year.

Strikingly, we also detected a large number of DNMs (6,135) that show allelic imbalance in the offspring blood sample (i.e., are carried in <35% of reads). We show, using read-based phasing and transmission patterns across multi-generational pedigrees, that a substantial fraction (at least 40%) of such variants are early embryogenic, rather than reflecting somatic mutations in blood or sequencing artifacts. Unlike mutations originating in the parental germline, these early embryonic mutations do not accumulate with parental age and exhibit no significant paternal bias. Thus, our results indicate that early embryogenesis contributes substantially to the rhesus macaque germline. Although direct comparisons across studies are difficult, this contribution appears more pronounced than that reported in humans using analogous methods, a finding that would be consistent with expectations for a species with a younger age of reproduction.

335T Novel gene content and meiotic drive in *Mimulus guttatus*

Evan Stark-Dykema[1], Findley Finseth[2], Lila Fishman[3] [1]Department of Biological Sciences, University of Montana, [2]Claremont Mckenna college, [3]University of Montana

Centromeres mediate faithful chromosome segregation during cell division yet are paradoxically diverse. The selfish centromere model explains the rapid evolution of both centromere satellite sequence and associated kinetochore proteins by positing centromeres as selfish elements that distort asymmetric female meiosis to gain a transmission advantage. Thus like many other meiotic drivers, selfish centromeres are predicted to accumulate linked enhancers. Here, we integrate experimental and population genomic analyses to investigate linked enhancers of meiotic drive in *Mimulus guttatus* (yellow monkeyflower). A meiotic driver (*D*) consisting of a unique set of extra D-only genes (EDG) is present in our focal population, yet the function and evolutionary history of these genes is unknown. Using a rare, shortened drive haplotype from a neighboring population that lacks the EDG region, we test if these genes are modifiers of drive and explore their evolutionary history. Experimental crosses reveal that the shortened haplotype exhibits a ~5% reduction in transmission distortion, consistent with the presence of an enhancer in the EDG region. Population genomic analyses further identify candidate enhancer genes and show that the EDG region arose through multiple insertion and tandem duplication events involving genes from throughout the genome. Taken together these analyses are the first evidence of linked enhancers in this system and expand our understanding regarding the evolutionary history of meiotic drive in *Mimulus*.

336T The Distribution of Superarchaic Admixture across the Human Genome

Alan R Rogers[1], Colin M Brand[2], Touhidal Islam[3], Timothy H Webster[3] [1]Anthropology, University of Utah, [2]University of California, San Francisco, [3]University of Utah

We introduce a new method for identifying genomic regions that are rich in introgressed DNA. The method does not search for introgressed haplotypes on individual chromosomes. For this reason, it is impervious to differences among populations in the length distributions of introgressed fragments. The method works with or without a genome from the donor population. When such a genome is available, however, the method has greater statistical power.

We apply the method to four previously-documented episodes of admixture: (1) from Neanderthals into modern Eurasians, (2) from a distantly related "superarchaic" population into Denisovans, (3) from this same superarchaic into Neanderthal-Denisovan ancestors, and (4) from a different superarchaic population into the ancestors of modern humans.

Our estimates confirm that Neanderthal admixture is generally low in regions that previous studies have identified as "deserts"--genomic regions lacking archaic admixture. However, our method also identifies "oases"--small regions of inflated admixture--within these deserts.

For all four episodes of admixture, we show that selection has opposed admixture near exons and regulatory elements.

337T Energy production capacity increases following genome doubling

Hailey Harkins[1], Breana Silvis[2], Evita Chee[1], Joel Sharbrough[1] [1]University of California, Santa Barbara, [2]New Mexico Institute of Mining and Technology

The multi-subunit enzymes that carry out photosynthesis are jointly encoded by the nucleus and the chloroplasts, and disturbing the stoichiometric balance between separately encoded subunits has the potential to affect the assembly dynamics, abundance, and activity of plastid-nuclear enzyme complexes. Whole-genome duplication events (WGDs) are expected to alter the gene dosage balance of plastid-nuclear enzyme complexes, which has implications for photosynthetic performance. Plants respond to WGDs by increasing the number of chloroplast genomes per cell, which broadly appears to restore the stoichiometry between compartments. Here, we evaluated the cellular and tissue-level consequences of this restoration in diploid vs. polyploid *Arabidopsis* by estimating the ratio of the chloroplast:nuclear genomes (and transcripts) using quantitative PCR (and RT-qPCR), quantifying and measuring chloroplast number and area in leaf tissue mesophyll using light microscopy, and measuring photosynthetic traits using a MultispeQ fluorometer. We tested for relationships between the genomic, tissue, and organismal level traits using a mixed linear modeling approach, and found that chloroplast genome copy number per cell and the number of chloroplasts per gram of tissue, but not their size, scaled with ploidy, leading us to hypothesize that polyploids exhibit enhanced photosynthetic capacity compared to diploids.

338T Comparing patterns of speciation in depth-separated species pairs of rockfishes

Jordan Pilcher[1,2], Liz Alter[3] [1]Department of Marine Science, CSU Monterey Bay, [2]Moss Landing Marine Laboratories, [3]Department of Biology, Agriculture, and Chemistry, CSU Monterey Bay

Speciation is the evolutionary process responsible for generating biodiversity, which is critical for ecosystem structure and function. Investigating the genomic, ecological, and geological factors involved in speciation helps to better understand historic and presentday patterns of biodiversity. Rockfishes (genus *Sebastes*) are a remarkably diverse clade with wide-ranging phenotypes and ecology that provide a unique opportunity to investigate and compare multiple replicates of genomic speciation. Using *Sebastes* as a model clade, this project identifies genomic patterns of speciation across environmental and temporal gradients in depth-separated sister species pairs. We performed low coverage (~2X) whole genome sequencing for ~15 individuals of 10 species comprising 5 sister species pairs, as well as ~4 individuals of species representing candidate sister taxa to each pair. We evaluated genetic population structure for each species, identified regions of genomic divergence between species pairs, and reconstructed the demographic history, including divergence times, of the species pairs. We also assessed the function of regions of genomic divergence and evaluated these patterns based on historic palaeoceanographic conditions, providing insight into the selective forces acting on the species pairs during the process of speciation. Our findings provide insight into the genomic mechanisms of sympatric speciation and the selective environmental pressures that influence speciation.

339T Evolutionary History of California Threespine Stickleback

Caitlin Wise[1], David Ferranti[1], Matthew Farnitano[1], Diana Rennison[2] [1]Ecology, Behavior, and Evolution, University of California, San Diego, [2]University of California, San Diego

Historical and environmental processes play a central role in shaping genomic variation across populations and landscapes. The threespine stickleback (*Gasterosteus aculeatus*), an anadromous fish species found across the Northern Hemisphere, provides a powerful system for examining these processes. Following the Last Glacial Maximum (LGM), marine stickleback repeatedly colonized newly formed freshwater habitats, leading to extensive adaptive divergence. As a result, most genomic work has focused on young postglacial populations in the Pacific Northwest, where patterns of genetic divergence are well-characterized. However, the broad geographic distribution of stickleback provides an opportunity to examine populations with potentially older and more complex evolutionary histories. California remained largely unglaciated during the LGM and may have supported long-term population persistence rather than recent postglacial colonization. California stickleback inhabit bar-built estuaries, lagoons, and lower river reaches that experience pronounced temporal variation in salinity, hydrology, and ocean connectivity. These habitats shift seasonally between brackish and freshwater conditions, creating dynamic and fluctuating selective environments that differ fundamentally from relatively stable postglacial lake systems. Despite this ecological complexity and the potential for a distinct evolutionary history, patterns of population structure and genomic divergence in California stickleback have not been characterized. Here, we conducted a landscape-level, genomic survey of California stickleback populations to characterize population structure, genetic variation, and genomic divergence across heterogeneous habitats. We evaluate whether patterns of divergence in California populations resemble classical postglacial freshwater adaptation or instead reflect alternative evolutionary trajectories shaped by long-term persistence in dynamic environments. Together, these results provide a framework for better understanding the evolutionary history of California stickleback populations.

340T The role of immune system incompatibilities in the evolution of isolating barriers within a tree adaptive radiation

Hossein Madhani[1], Elizabeth Stacy[1], Jae Choi[2], Maryam Hadi[3] [1]School of Life Sciences, University of Nevada, Las Vegas, [2]Department of Ecology & Evolutionary Biology, University of Kansas, [3]University of Nevada, Las Vegas

Intrinsic postzygotic isolating barriers are thought to play a critical role during the early stages of speciation with ongoing gene flow, yet their genetic and physiological bases remain poorly understood in natural systems. Hybrid necrosis (HN), a form of intrinsic hybrid incompatibility caused by negative epistatic interactions among immune system components, has been documented primarily in model plants. Here, we present ongoing work investigating the genetic, hormonal, and genomic basis of hybrid necrosis within the adaptive radiation, Hawaiian *Metrosideros*, a dominant tree lineage spanning multiple stages of ecological divergence and speciation across the Hawaiian Islands. We phenotyped more than 3,000 seedlings from 28 full-sib families derived from controlled crosses between sister varieties of *Metrosideros polymorpha*. Seedlings were classified into discrete phenotypic categories ranging from normal growth to severe stunting and necrosis. Segregation patterns within families were consistent with simple one- and two-locus Mendelian models, demonstrating that hybrid necrosis is genetically determined and segregates predictably, generating substantial intrinsic reproductive isolation even between recently diverged, sympatric varieties. To test whether the necrotic phenotype reflects immune-system misregulation, we conducted a targeted hormone-panel analysis comparing normal and necrotic seedlings. Necrotic seedlings showed significant elevation of jasmonic acid (JA) and its biosynthetic precursor, 12-oxo-phytodienoic acid (OPDA), implicating JA-dependent immune pathways that are most commonly associated with defense against herbivory and necrotrophic attackers. To place these findings in an evolutionary context, we integrated segregation and hormone data with population-genomic and introgression analyses of resistance (R) genes across the Hawaiian *Metrosideros* radiation. Genome-wide scans revealed extensive clustering, elevated nucleotide diversity, and signatures of selection in NLR-type resistance genes. Notably, several candidate R-gene regions showed reduced introgression between HN-producing populations despite high genome-wide gene flow, suggesting that immune genes may act as barrier loci during divergence. RNA sequencing of purported hybrid-necrotic and normal seedlings is underway and expected to show constitutive upregulation of immune-system genes in the former. Combined, these results are likely to demonstrate immune-system misregulation—particularly involving JA-dependent pathways—as a driver of intrinsic postzygotic isolation during adaptive radiation in trees.

341T Cell-type-resolved regulatory evolution during marine–freshwater divergence in stickleback

Zachary Tobias[1], Benjamin Gastfriend[1], Matthew Farnitano[1], Martin Tresguerres[1,2], Richard Daneman[1], Diana Rennison[1] [1]University of California - San Diego, [2]Scripps Institution of Oceanography

Evolutionary adaptation often involves not only changes to a gene's function, but also to where, when, and under which environmental contexts the gene is expressed. Functional genomic studies have revealed striking examples of how selection on cis-regulatory elements can mediate adaptive shifts in ecologically relevant phenotypes. However, identifying the specific cellular contexts in which gene expression remodeling occurs has proven difficult, as most approaches rely on bulk tissues or simplified experimental systems. Recent advances in single-cell sequencing methodologies now provide the opportunity to resolve cell-type-specific regulatory landscapes and investigate how regulatory programs are reshaped during environmental adaptation.

Here, we present an integrative analysis that combines paired single-nucleus RNA-seq and ATAC-seq (10x Genomics Multiome) with population genomic data to link putatively adaptive genetic variants to cell-type-specific cis-regulatory elements. By assaying gene expression and chromatin accessibility in individual cells from whole brains of both juvenile and adult threespine stickleback (*Gasterosteus aculeatus*), we are able to generate a single-cell regulatory atlas across development. Because marine–freshwater divergence in stickleback involves repeated shifts in behavior and sensory ecology, the brain provides a compelling system in which to study adaptive regulatory evolution. Using this atlas, we aim to investigate how adaptive divergence during colonization of new habitats may proceed through regulatory reprogramming in the brain. Widespread and repeated transitions from the ancestral marine form of *G. aculeatus* into varied freshwater environments offer an excellent opportunity to examine core, repeated targets of cis-regulatory evolution versus more idiosyncratic responses. By integrating population genomics with single-cell regulatory maps, we assess how adaptive regulatory divergence is distributed across cell types, regulatory element classes, and repeated evolutionary transitions.

342T An exapted transposase suppresses the *Mu*-suppressible lesion-forming Uroporphyrinogen III Synthase mutant allele

Rajdeep Khangura[1], Xinyan Zhang[2], Amanpreet Kaur[3], Damon Lisch[3], Brian Dilkes[4] [1]Plant and Agroecosystem Sciences, University of Wisconsin-Madison, [2]Chinese Academy of Agricultural Sciences, [3]Purdue University, [4]Biochemistry, Purdue University

Porphyrins are tetrapyrrole molecules that are required for metabolic processes and adaptative responses to environmental cues. Porphyrins include energy-transferring molecules such as siroheme, cobalamin, heme, bilins, and chlorophylls. Three enzymes in porphyrin biosynthesis are completely conserved across living things: porphobilinogen synthase, hydroxymethylbilane synthase, and uroporphyrinogen III synthase. We identified two semi-dominant mutant alleles at the *lesion28* (*les28*) locus in maize that are encoded by uroporphyrinogen III synthase. The *Les28-1* and *Les28-2* alleles carry *Mutator* transposable element insertions in the first intron and 5' UTR resulting in light-dependent punctate lesion formation in leaves. The *Les28-1/Les28-2* seedlings were yellow seedling-lethal, consistent with other loss-of-function alleles in early steps of the porphyrin pathway. Phenotypic expression of *Les28-1/+* and *Les28-2/+* was modulated by the epigenetic state of their *Mutator* transposon insertions, a phenomenon called *Mu*-suppressibility. Silencing of *Mutator*transposition activity suppressed lesion formation in *Les28-1/+* and *Les28-2/+*. We carried out a Genome Wide Association Study of F1 families segregating for *Les28-1* to identify natural variants that modify mutant phenotype expression. Alleles linked to genes in the porphyrin pathway had small effect on lesion severity but included modification in cis by alleles at the *les28* gene itself. The most signification association with lesion severity was encoded by an exapted MURA-like transposase which has been inherited as a gene since before the split of *Tripsacum* and *Zea* and is functionally polymorphic in maize. Putative functional alleles present in cultivated maize are also found in *Zea mays* spp indicating that these alleles predate domestication. This work demonstrates a new aspect of *Mutator* biology and the regulation of *Mu*-suppressible alleles by an exapted transposon-derived trans-regulator.

343T High-throughput analysis of the eco-evolutionary dynamics over 75,000 generations in the E. coli long-term evolution experiment using a novel barcoding system

Jenya Belousova[1], Joao Ascensao[1], Tatiana Ruiz-Bedoya[1], Misha Gupta[1], Jack Edwards[1], Shreyas Pai[1], Adam Deutschbauer[2], Jeffrey Barrick[3], Michael Desai[1] [1]Harvard University, [2]UC Berkeley, [3]Michigan State University

Experimental evolution provides unprecedented opportunities to study the dynamics of microbial adaptation and ecological diversification. The Escherichia coli Long Term Evolution Experiment (LTEE), ongoing for over 75,000 generations, has revealed long-term coexistence between clades of the same population likely due to complex ecological structure in many LTEE populations. Some of these ecological interactions have been characterized, however, the eco-evolutionary mechanisms of the remaining majority are unclear.

In this study, we employ a high-resolution barcoding system to dissect the eco-evolutionary trajectories of all 12 LTEE populations. We have constructed a library of barcoded clones isolated from these populations across the longest available time course of the experiment. Using competitive fitness assays for these clones, we aim to investigate their ecological interactions by measuring frequency-dependent effects on fitness. This will enable us to characterize the spectrum of ecological possibilities explored by LTEE populations and how it shifts over time.

This approach leverages high-throughput barcode-based phenotyping to systematically characterize ecological structure of populations undergoing long-term microbial adaptation, providing new insight into the dynamics of ecological differentiation.

344T The polygenic basis of ecogeographic isolation maintains species distinction despite ongoing introgression

Julia G Harencar[1], Kathleen M Kay[2] [1]Evolution and Ecology, UC Davis, [2]Ecology and Evolutionary Biology, UC Santa Cruz

Evidence of introgression between well-defined species is rapidly accumulating, raising the question: how are species boundaries maintained in the face of gene flow? Previous work has identified genomic regions with reduced recombination that harbor key differentiating loci. But what happens when divergence is polygenic – when many loci spread across the genome contribute to differentiation? Here, we combine analysis of the genetic basis of divergent habitat adaptation with population genomic analyses of introgression to understand how species remain isolated despite ongoing gene flow. We focus on two species from a recent, rapid radiation of Neotropical understory plants — *Costus villosissimus* and *C. allenii* — that are primarily reproductively isolated by divergent habitat adaptation. We map quantitative trait loci (QTL) underlying key adaptive differences, quantify both recent and historical introgression, and compare the genetic basis of divergent adaptation to genome-wide patterns of divergence and introgression. This combined top-down and bottom-up approach, applied to a system with well-characterized reproductive isolating barriers, clarifies the genetic underpinnings of polygenic species boundary maintenance despite introgression.

345T Unveiling the genetic basis of water stress resilience in *Fragaria vesca* through high-resolution genome-wide association mapping

Sebastian Arenas, Johan Stenberg, Carolina Diller Swedish University of Agricultural Sciences, SLU, Sweden, Swedish University of Agricultural Sciences

Understanding the genetic basis of water stress resilience is crucial for improving crop performance under changing climates. We studied 220 genotypes of *Fragaria vesca* sampled across a broad latitudinal gradient from Turkey to northern Scandinavia. Genotypes were clonally propagated in nurseries and grown under controlled conditions for 8 months, achieving uniform development (~15 cm height) before treatment application. From these, 160 genotypes were selected and exposed to either well-watered or water-limited conditions, with the latter irrigated only once per week. All plants were grown in a highly controlled drip irrigation system, ensuring identical water volumes per plant and minimizing environmental variation. Phenotypic measurements were collected throughout the summer, capturing growth and physiological traits such as plant height, width, volume, leaf number, leaf area, specific leaf area, PSII efficiency, stomatal conductance, and reproductive output. Significant differences among genotypes were observed for nearly all traits, highlighting strong genetic variation in response to water stress. High-density whole-genome sequencing (~1 million high-quality SNPs) aligned to the latest *F. vesca* reference genome enabled genome-wide association analyses. This approach identifies loci and candidate genes associated with drought resilience, elucidating the underlying genetic architecture of adaptive traits. Our results demonstrate that water stress response is highly heritable and polygenic, with distinct genomic regions influencing key physiological and growth traits. This study integrates phenotypic and genomic data to reveal mechanisms of drought adaptation in *F. vesca*, providing a foundation for breeding programs aimed at improving stress resilience in strawberry and related species. The combination of extensive genotypic diversity, controlled experimental conditions, and high-resolution genomics offers unprecedented insight into the genetic determinants of plant performance under water-limited environments.

Keywords: *Fragaria vesca*, drought tolerance, genome-wide association study, clonal replicates, high-throughput phenotyping, water stress

346T Evolutionary Flexibility of *Daphnia* Ribosome

Man Lin[1], Edgar Ferrer[2], Po-Lin Chiu[1], Michael Lynch[1] [1]Center for Mechanisms of Evolution, Arizona State University, [2]TCLAS School of Molecular Sciences, Arizona State University

Ribosomes are highly conserved in overall function, yet many ribosomal proteins, rRNA regions, and lineage-specific structural elements undergo substantial divergence. Previous studies have solved three-dimensional ribosome structures from many species and used them to reconstruct the evolutionary history of the ribosome. However, the mechanisms that allow lineage-specific sequence divergence without compromising ribosome function remain poorly understood. Here we show the structure of the *D. pulex* cytoribosome using single-particle cryogenic electron microscopy (cryo-EM), and our consensus cryo-EM density map allowed atomic modeling of both ribosomal subunits together with associated tRNAs, mRNA, and nascent peptides captured during translation. The structure reveals a conserved eukaryotic core with lineage-specific expansions in rRNA and ribosomal proteins, and comparative analyses within *Daphnia* and among *Daphnia*, *Drosophila*, and human ribosomes show that these features contribute to structural divergence while preserving the overall architecture of the ribosome. By integrating structural comparisons with population-genetic analyses of ribosomal protein genes and rRNAs, we link patterns of sequence variation to specific structural features and dynamic regions of the ribosome, establishing a framework for testing compensatory evolution in a highly constrained macromolecular complex.

Keywords: ribosomes, population genetics, structural evolution

347T Not All Roads Lead to Red: The genetics of sexual dichromatism in the northern cardinal (Cardinals cardinalis)

Miranda J Wade[1], Sara Patton[2], Elizabeth C Scheer[3], Rebecca E Koch[4], Yufeng Zhang[5], Matthew B Toomey[6,7], Geoffrey E Hill[2], Simon YW Sin[1] [1]School of Biological Sciences, The University of Hong Kong, [2]Department of Biological Sciences, Auburn, [3]University of Tulsa, [4]Department of Biology, University of Wisconsin-Stevens Point, [5]College of Health Sciences, University of Memphis, [6]Department of Integrative Biology, Michigan State University, [7]Department of Biological Science, University of Tulsa

Sexual dichromatism, characterized by sex-specific differences in coloration, is widespread among birds and often involves carotenoid-based pigmentation. Despite extensive research on the social and ecological environments favoring sexual dichromatism, the molecular mechanisms underlying its development and evolution remain largely unexplored. In this study, we investigated the genetic and molecular processes driving sexual dichromatism in the red ketocarotenoid-based plumage coloration of northern cardinals (Cardinalis cardinalis). We quantified carotenoid concentrations in plasma and feather follicles, confirmed that homologs of CYP2J19, BDH1L, and TTC39B catalyze ketocarotenoid metabolism, and performed gene expression analyses across tissues. Males showed significantly higher plasma and feather ketocarotenoid concentrations along with upregulated CYP2J19 and TTC39B expression in liver and feather tissues and upregulated carotenoid transport gene expression in gut and feather follicles. Females exhibited a dramatic upregulation of BCO2, facilitating carotenoid degradation and attenuating red pigmentation. Additionally, sex-biased expression of hormonal regulators such as HSD17B4 and ZNF131 in the feather follicle suggests hormonal modulation influences dichromatism. These findings indicate that sex-specific regulation of carotenoid processing genes underpins the vivid red coloration in males and the drab phenotype in females, likely maintained by a balance between natural and sexual selection, with mechanisms of sexual antagonism driving divergent gene expression. Our work advances understanding of the molecular basis of avian sexual dimorphism, highlighting key genetic pathways involved in carotenoid-based coloration and providing a foundation for further research into the evolution of sexually dichromatic traits.

348T Living with a killer: how coevolved *Saccharomyces cerevisiae* become killer toxin resistant

EmilyClare Baker[1], Angelina Chan[2], Magdalena Pieczynska[3], Arjan de Visser[4], Gavin Sherlock[2], Michelle Hays[1]
[1]Department of Human Genetics, University of Michigan Medical School, [2]Genetics Department, Stanford University, [3]Institute of Genetics and Animal Biotechnology Polish Academy of Sciences, [4]Laboratory of Genetics, Wageningen University

Killers abound among *Saccharomyces cerevisiae*. Toxic peptides provide an advantage against susceptible neighbors, but under some conditions their metabolic cost can result in rapid loss. A seemingly simple pairwise conflict, a complex genetic system is often involved. In many strains a dual viral system is required for killing; the killer toxin/antidote encoding satellite and a virus it utilizes for survival and replication. How the genomes in this system evolve can provide insights into fungal intracellular defense, fungal-virus interactions, and toxin evasion strategies in fungi; all compelling places to approach the discovery of new anti-fungal strategies.

In an experimentally co-evolved population of susceptible and killer yeast, with the killers containing a virally encoded K1 type toxin, the susceptible yeast evolved resistance to toxin over 500 generations. To determine the molecular basis of evolved toxin resistance we performed bulk segregant analysis of resistant isolates. A missense variant of unknown function, which had become fixed in the population, was identified in the gene *SSK1*. Our genetic analyses confirm this mutation is necessary and sufficient for K1 toxin resistance and acts in a genetically dominant manner. *SSK1* is an intermediate regulatory signal transducer of the HOG osmoregularity pathway and part of a two-component system. Metazoans lack such systems, which have been implicated in fungal adaptation under other conditions, making them appealing anti-fungal drug targets.

The K1 toxin is an ionophore that causes cell death through loss of membrane integrity, but it also interacts with components of the cell wall. To determine how mutant *SSK1* confers K1 toxin resistance we challenged mutant *SSK1* strains with stressors targeting the membrane and cell wall components. Preliminary results suggest *SSK1* increases resistance to a compound targeting the membrane, but decreases resistance to beta-glucan and chitin targeting dyes. These results suggest *SSK1* engages in crosstalk with the cell wall integrity pathway and that resistance comes at a cost to cell wall defenses. Tests of other cell wall and membrane targeting drugs and transcriptional profiling will shed further light on how mutant *SSK1* alters the cell wall and plasma membrane. In addition to being part of a signaling system not found in animals, *SSK1s* vulnerability to tradeoffs in evolving resistance to toxins suggest its potential as a target for anti-fungal treatments.

349T Genomic signatures of spatially heterogeneous selection in evolving *Pseudomonas fluorescens* populations

Chimaroke Onyeaghala, Susan Bailey biology, Clarkson University

Understanding how spatial structure influences evolutionary dynamics is critical for predicting microbial adaptation in heterogeneous environments. We evolved replicate populations of Pseudomonas fluorescens for 21 days on spatially structured agar plates in which glucose was localized at the center, creating a resource-rich patch surrounded by nutrient-poor regions, in contrast to a uniform environment with an even distribution of glucose. Populations were sampled from both the center and edge of the plates, and whole-genome sequencing was used to characterize evolutionary changes.

Across 16 evolved subpopulations, we identified over 1,100 mutations, the majority occurring at low frequencies, indicating that most genomic variation remained polymorphic at the end of the experiment. Statistical analyses revealed that spatial location strongly influenced evolutionary outcomes. Mutation category composition differed significantly between center and edge populations, and nonsynonymous-to-synonymous mutation ratios varied by spatial location, suggesting differences in selection intensity and/ or type across regions of the plate.

Analysis of parallel evolution revealed that genomic similarity among replicate populations was also structured by spatial location. In particular, a large proportion of high-frequency nonsynonymous mutations occurred in the flagellar regulator gene fleN, primarily in center populations, indicating strong and repeated selection on motility-related functions.

To complement genomic analyses, we performed phenotypic assays comparing evolved and ancestral strains. Growth rate assays revealed measurable differences between evolved and ancestral populations, while competition experiments were conducted to test for local adaptation between edge populations and the ancestral strain.

Together, these results suggest that spatial position within structured environments can generate heterogeneous selective pressures that shape both genomic and phenotypic evolution. Our findings highlight the importance of spatial context in determining adaptive trajectories in microbial populations.

350T Comparative Mutational Fitness Landscapes of SARS-CoV-2 Spike Domains: Diverting De Novo Immunogen Design from the RBM to the Fusion Peptide

Anna Fong Sacred Heart Preparatory

A central challenge in predicting viral evolution is distinguishing between domains under adaptive positive selection and those restricted by severe epistatic constraints. Current vaccine strategies predominantly target the Receptor Binding Motif (RBM), a region characterized by rapid antigenic drift and immune evasion. To quantify the evolutionary constraints across the SARS-CoV-2 spike protein, we utilized ESM-2, a 650M-parameter protein language model, to perform a comparative zero-shot Deep Mutational Scan (DMS) of the S1 RBM and the S2 Fusion Peptide (FP).

By calculating the Log-Likelihood Ratio (LLR) for the entire 1-step and 2-step mutational neighborhoods, our ESM-2 analysis revealed starkly contrasting evolutionary landscapes. The RBM exhibited high mutational tolerance (frequent positive LLRs), confirming its role as an evolutionary playground driven by positive selection. Conversely, the canonical FP residues 816–827 (SFIEDLLFNKVT) presented a profound "fitness valley" (mean LLR < -4.8). The hydrophobic LLF motif (821–823) in particular demonstrated intense purifying selection, where any sequence deviation incurs a catastrophic predicted fitness cost due to its structural role in membrane insertion.

Driven by these quantitative genetic insights, we abandoned the mutable RBM and selected the invariant FP as the target for de novo immunogen design. Using a computational pipeline of RFdiffusion and ProteinMPNN, we engineered a minimalist protein scaffold optimized for this evolutionarily locked motif. The lead candidate achieves extreme structural mimicry (RMSD < 0.3 Å to PDB: 6VXX) and sub-nanomolar binding affinity, while successfully clearing a homology-based safety audit to prevent human musculoskeletal mimicry. This work establishes a predictive, genetics-first framework for achieving variant-proof immunity by targeting motifs where viral escape is evolutionarily prohibited.

351T The role of structural variants in *Formica* ant supergenes

Zul Alam[1], Jessica Purcell[2], Alan Brelsford[3] [1]Evolution, Ecology, and Organismal Biology, University of California, Riverside, [2]Entomology, University of California, Riverside, [3]University of California, Riverside

Supergenes, or regions of suppressed recombination containing multiple linked functional elements, underlie complex polymorphisms across a wide variety of organisms. In the ant genus *Formica*, an ancient autosomal supergene on chromosome 3 underlies colony queen number. This supergene is tightly linked to a queen-size supergene on chromosome 9 in *Formica cinerea*. The near perfect linkage of supergenes on different chromosomes suggested physical linkage of the supergenes through a chromosomal fusion or reciprocal translocation. It was however recently shown that the supergenes assort independently in eggs and only exhibit linkage in adults, suggesting epistatic linkage rather than physical linkage between the supergenes.

In this study, we examined the genetic architecture of the queen-number and queen-size supergenes in two additional species, *Formica argentea* and *Formica marcida*, using long-read sequencing. We found that homologous inversions do indeed underlie the suppression of recombination on chromosome 9 in both species. Since *F. argentea* and *F. marcida* also show linkage between supergenes, albeit weaker than *F. cinerea*, we also searched for structural variants which may underlie the linkage of chromosome 3 and 9. We found evidence in both species for a fusion or reciprocal translocation. Given linkage is higher in *F. cinerea* than *F. marcida* and *F. argentea*, one might expect physical linkage in *F. cinerea* but not in *F. marcida* or *F. argentea*, yet we find the opposite. We then searched for evidence of degeneration in the region of suppressed recombination on chromosome 9 and found increased repeat content on the derived haplotype of *F. marcida* but not *F. argentea*. We also aim to assess degeneration due to SNPs and inversion breakpoints disrupting genes. Overall, we find evidence that autosomal supergene fusion/reciprocal translocations across chromosomes can be selected for due to likely epistatic interactions in a process analogous to fusions leading to neo-sex chromosome formation.

352T cis-regulatory variation in the RNA binding protein *bruno* is modulates the sterility effects of DNA transposition

Lorissa Saiz, Noelle Ramos, Erin Kelleher Biology and Biochemistry, University of Houston

Transposable elements (TEs) are genetic parasites that increase in copy number in the germlines of organisms to guarantee transmission to offspring. While TEs are predominantly vertically transmitted, they occasionally invade new host genomes through horizontal transfer. Horizontal transfer events are often followed by a burst of transposition, leading to fitness costs in the new host. Natural variation in the new host with regards to the fitness costs of transposition could therefore be of major significance following TE invasion.

P-elements are DNA transposons that were horizontally transferred to *Drosophila melanogaster* in the 1950s. In naive hosts *P*-element transposition causes a sterility syndrome called hybrid dysgenesis. Through QTL mapping we localized natural variation in dysgenic sterility to the *bruno* locus, which encodes an RNA binding protein that acts as a translational regulator throughout oogenesis. *bruno* is a highly-conserved mega locus (140 Kb), with 24 exons and at least 6 promoters, that encodes for multiple protein isoforms with varied tissue-specific functions. Through fine-scale and deletion mapping, I show that natural variation in bruno corresponds to cis-regulatory elements that impact the expression of the ovarian Bruno isoform. Furthermore, there are likely multiple causative variants that reside in regions of differing chromatin accessibility between sterile and fertile alleles. Our work highlights the complexity of regulatory variation in pleiotropic genes that code for alternative transcripts, and paves the way for evaluating the recent history of adaptive regulatory variation.

353T The selective sweeps underlying the origin of a hummingbird-adapted *Penstemon* species

Aidan Short[1], Samson Acoca Pidolle[2], Lena Hileman[2], John Kelly[2], Carolyn Wessinger[3] [1]University of South Carolina, [2]University of Kansas, [3]Biological Sciences, University of South Carolina

Adaptation by natural selection generates distinct combinations of traits that function well together. The maintenance of adaptive trait combinations in the face of gene flow depends on the strength and nature of selection acting on the underlying genetic loci. Floral pollination syndromes are complex adaptations that exemplify trait combinations adaptive for particular pollinators. The North American wildflower genus *Penstemon* exhibits a remarkable pattern of repeated evolutionary shifts from ancestral bee syndrome flowers to hummingbird syndrome flowers. These evolutionary shifts have occurred relatively recently, leading to closely related pairs of species that differ in floral syndrome. Using a focal species pair with alternate floral syndromes, we investigated the genomic basis of floral syndrome divergence, finding only narrow genomic regions sprinkled throughout the genome confer species differences. These narrow regions show evidence of recent selective sweeps in the hummingbird syndrome species. We infer that divergent floral syndromes are maintained through strong selection acting on combinations of traits, in combination with assortative mating conferred by distinct pollinator preferences. To test this notion, we performed population genomic simulations incorporating pollinator-mediated selection and gene flow between populations in a metapopulation framework. We find this model can generate genomic patterns that resemble our empirical dataset.

354T Variable barriers to gene flow across a field cricket hybrid zone

Erica L Larson[1], Thomas J Firneno[2], Alberto Aleman[1], Hannah Eckert[1], Alyson Emery[1], Isaac Hudson Foy[1], Martin Garlovsky[1], Scott Melander[1], Mykaela Tanino-Springsteen[1], Robin Tinghitella[1] [1]Biological Sciences, University of Denver, [2]Juniata College

We often think of the traits that define species boundaries as constants – fixed characteristics of species that act as barriers to gene exchange with other lineages. Yet even in the early stages of divergence, species are often isolated by multiple barriers that vary across different ecological contexts, population history and genetic variation, or interactions with other species. The extent and consequences of variation in the strength of barriers is still largely unknown, but it is likely that the nature of species boundaries is much more complex than we find from comparisons of only a few populations. We are conducting a geographically comprehensive survey of multiple reproductive barriers and gene flow in an expansive hybrid zone between the field crickets, *G. pennsylvanicus* and *G. firmus*. We report new results that evaluate temporal, behavioral, and fertilization barriers, as well as hybrid inviabilty, across replicate transects of the hybrid zone. We present preliminary results that evaluate patterns of introgression across contacts and QTL mapping of barrier traits.

355T Population genomics of Barred surfperch (*Amphistichus argenteus*), a viviparous fish associated with sandy environments along the California and Baja California coasts

Allyson C Sawkins, Giacomo Bernardi Ecology and Evolutionary Biology, UC Santa Cruz

Dispersal limitation and environmental heterogeneity can interact to generate pronounced population structure and promote divergence in coastal marine fishes. The barred surfperch (*Amphistichus argenteus*), a viviparous species lacking a pelagic larval stage, provides an opportunity to examine how life history and habitat context shape genomic variation across a dynamic coastline. We use medium-coverage whole-genome resequencing of 256 individuals to assess population structure, divergence, and signatures of local adaptation across a latitudinal gradient from central California to Baja California, Mexico, including an offshore island. We identify four genetic groups—three coastal groups separated at major biogeographic breaks (Big Sur and the Palos Verdes Peninsula) and one island group. Differentiation among mainland populations is low to moderate (FST = 0.04–0.14), whereas island–mainland divergence is higher (FST = 0.12–0.20). Coastal populations exhibit isolation-by-distance, and we detect loci showing signatures of selection. Together, these results suggest that divergence among coastal populations reflects spatially variable effective gene flow, where dispersal is present but constrained by habitat discontinuity and post-dispersal selection. This research provides insights into the processes driving divergence in marine vertebrates, particularly those without dispersive larvae, and offers a baseline for monitoring genetic responses to environmental change.

356T Quantifying anthropogenic effects on maize genetic diversity

Samantha Snodgrass[1], Forrest Li[1,2], Luke Sparreo[3], Santiago Medina-Muñoz[4], Sowmya Mambakkam[1], Mitra Menon[1,5], Claudia Perez[1], Andrés Moreno-Estrada[4], Daniel Runcie[2], Graham Coop[1,6], Jeffrey Ross-Ibarra[1,6,7] [1]Evolution and Ecology, University of California Davis, [2]Plant Sciences, University of California Davis, [3]Center for Biodiversity and Evolution, New York Botanical Garden, [4]Advanced Genomics Unit (UGA), Centro de Investigación y Estudios Avanzados del IPN (Cinvestav), [5]Colossal Biosciences, [6]Center for Population Biology, University of California Davis, [7]Genome Center, University of California Davis

When describing genetic diversity, environment and wild relative introgression often capture major axes of variation. Yet, for domesticates that rely upon humans for survival, anthropogenic effects often remain unquantified. Maize genetic diversity is unstructured, weakly correlated with geography, and admixture with a highland wild relative drives the largest axis of variation. Given maize depends completely upon humans for long-term persistence and dispersal along with the deep cultural connections with maize agriculture in the Americas, we hypothesize human variation may capture additional variation beyond environment and wild relative introgression. Using publicly available GBS and passport data from around 2,000 traditional maize varieties and 515 whole genome sequences of humans from the Mexican Biobank, we quantify the anthropogenic effect on maize genetic diversity in the Americas. Pairing maize and human samples, Procrustes correlations between maize and human genetic diversity was equivalent to maize and geography. The limited population structure within maize points to the effective mixing of maize by human trade across large geographic distances. Within the limited population structure, no environmental or anthropogenic parameters capture a large proportion of variation genome-wide. Thus, long-distance trade may have played a large role in creating existing patterns of genetic diversity found in traditional varieties.

357T Local adaptation in *Mimulus guttatus*

Genevieve Triplett, Sarah Uher, Jacob Mull, Katherine Toll University of South Carolina

Spatial variation in chemical and physical attributes of soils are important drivers of plant population differentiation. Dramatic transitions between “normal” and atypical soils (limestone, shale, gypsum, and serpentine) can generate strong divergent selection, resulting in the evolution of local adaptation. Prior work demonstrated that populations of a common wildflower species, *Mimulus guttatus*, were locally adapted to serpentine soils. However, fitness trade-offs across soils were not detected, despite the low frequency of serpentine tolerance off of serpentine soils. To test for fitness trade-offs, identify phenotypes contributing to fitness, and determine the genetic basis of adaptation, we used large-scale field reciprocal transplant experiments with local serpentine and non-serpentine genotypes and hybrids. Consistent with prior work, we found strong viability selection against non-serpentine genotypes on serpentine soils. Serpentine tolerance was dominant, segregation ratios in hybrids were consistent with a major locus underlying survival on serpentine, and cytoplasmic genomes did not contribute to local adaptation. Contrary to prior studies, we detected fitness trade-offs across soils, although selection against serpentine genotypes on non-serpentine soil varied by site. The onset of flowering differed between parental genotypes and soils, and selection on flowering time in hybrids differed between soils, potentially contributing to the maintenance of differentiation.

358T Natural variation reveals hidden divergence in the evolution of a polyphenism

Stephen M Dreyer, Erik J. Ragsdale EEB, Indiana University

Polyphenism, the ability of an organism to produce discrete alternative phenotypes in response to environmental conditions, exemplifies the non-linear relationship among genes, the traits they influence, and the environment. While several genes that regulate polyphenism have been identified in some systems, it is mostly unknown which loci harbor intraspecific variation for polyphenism. To identify cryptic loci associated with polyphenism and natural evolutionary divergence, we queried the nematode *Pristionchus pacificus*, which exhibits resource polyphenism, including the development of alternative feeding-structure morphologies. To do this, we created a panel of recombinant inbred lines derived from *P. pacificus* isolates that exhibit similar morph-bias ratios under common laboratory conditions. We found that the resulting recombinants included morph-frequency phenotypes outside the parental range, indicating cryptic divergence between strains. Quantitative trait loci (QTL) analysis subsequently revealed three loci of large effect that influence morph production: two on Chromosome I, and one on Chromosome X that is epistatic over the autosomal loci. The X-linked QTL overlaps with a region known to contain the polyphenism switch gene *eud-1* and is consistent with recent work that similarly found this region to harbor variation for the polyphenism. However, functional validation by CRISPR/Cas9-driven allelic replacement revealed that the causal allele in our panel is different from that previously identified, indicating that this locus has been targeted independently by evolution among different isolates of *P. pacificus*. Using CRISPR/Cas9-directed recombination in a near-isogenic line, we have developed a panel of additional recombinants within this region to fine map the causal locus to a 3-kb interval that excludes the gene body and known regulatory regions of *eud-1*. In summary, we describe a diverging polyphenism’s genetic architecture, which consists of multiple QTL that together hide genetic variation by epistasis, and we show that a narrow, X-linked locus has been a convergent target of polyphenism evolution.

359T Inference of polygenic selection on thermal performance variation in a marine invertebrate

Alejandro Calderon, Daniel Wuitchik, Lawrence Uricchio Biology, Tufts University

Thermal performance variation appears to have a polygenic basis in many organisms. With global temperatures expected to increase rapidly in the next century, many species may experience range reductions as temperatures exceed thermal performance limits, unless they evolve to tolerate increased temperatures. We developed a new polygenic selection inference method that pairs experimental thermal performance data with a sib-based phenotype-genotype associations to infer the strength of selection acting on thermal performance variation. Our approach overcomes some limitations of prior methods, by enabling the estimation of rare allele contributions to heritability and reducing the potential for stratification biases. We apply the method to data from froward simulations, showing that it is able to infer both the strength of selection and the fraction of heritability explained by rare alleles, parameters that may be critically related to the potential rate of adaptation under global change. We then use simulations to explore how these parameters relate to adaptation rates under global change in situations when adaptation is constrained by pleiotropy or seasonality, and apply our method to a pilot dataset from the Atlantic Slipper Snail.

360W Finding order in evolution: how fitness reveals hidden biological modules

Kara Schmidlin, Mohammad Donyavi, Leandra Brettner, Kerrry Geiler-Samerotte Arizona State University

Understanding the phenotypic mechanisms by which mutations affect fitness remains a central challenge in evolutionary biology. Genotype-to-fitness maps are increasingly accessible through high-throughput evolution experiments. But the phenotypes that mediate these effects, and thus determine how fitness consequences change across environments or genetic backgrounds, often remain unknown. Without identifying these phenotypes, predicting evolutionary outcomes across contexts remains fundamentally limited.

I addressed this challenge by using massively parallel lineage-tracing experimental evolution to generate ~1,000 independently evolved drug-resistant yeast mutants. I measured how their fitness effects trade off across 12 environments and clustered mutants by their genotype-by-environment (GxE) fitness responses. Remarkably, these fitness landscapes collapsed into just six characteristic response types, consistent with long-standing theoretical expectations that phenotypic space must be modular and pleiotropy limited for evolution to proceed efficiently. However, because these clusters were interpreted as reflecting shared underlying phenotypes based solely on similar fitness responses across environments, rather than on direct phenotypic measurements, it remained unclear whether they corresponded to true biological modules.

I directly tested this inference using cutting edge single-cell RNA sequencing on a subset of mutants spanning distinct GxE fitness clusters. We find that mutants from different fitness clusters exhibit strikingly distinct transcriptional states, involving largely non-overlapping gene sets and divergent responses to drug exposure. Further, mutants with known fitness deficits in specific drug environments exhibit distinct stress-response signatures. These results demonstrate that GxE fitness-based clustering recovers underlying molecular phenotypes and that phenotypic space is indeed modular at the level of gene expression.

Together, these findings show that fitness landscapes can be used not only to quantify evolutionary outcomes, but to generate testable predictions about underlying cellular phenotypes and transcriptional programs. By validating phenotypic inferences drawn solely from fitness responses, this work demonstrates how evolutionary experiments can be used as a powerful lens for dissecting cellular organization. More broadly, these results provide empirical resolution to long-standing debates about the modularity of biological systems, showing that phenotypic space is structured into discrete, biologically meaningful modules with distinct molecular signatures. While not eliminating evolutionary complexity, this modular organization suggests that evolutionary responses may be more structured—and thus more interpretable across contexts—than previously appreciated.

361W Deconstructing empirical fitness seascapes across scales of granularity

Swathi Nachiar Manivannan[1], Brandon Ogbunu[1,2] [1]Ecology and Evolutionary Biology, Yale University, [2]Santa Fe Institute

The fitness landscape metaphor remains resonant in evolutionary theory and has facilitated the birth of newer concepts---like the fitness seascape---that consider the role of environmental context in shaping the dynamics of evolution. Since the emergence of the fitness seascape, it has appeared in several studies that examine how different and fluctuating environments shape evolutionary outcomes. Despite a growing interest in these topics, we lack comprehensive examinations of the role of environmental context in shaping features of fitness seascapes. In this study, we address this gap by deconstructing empirical fitness seascapes across scales of granularity: mutational steps, loci, locus interactions, alleles, trajectories, and entire seascapes. For each, we examine how environmental context influences qualitative and quantitative aspects of seascapes, and find that they change appreciably, with patterns that are specific to individual systems of study. In summary, we reflect on the implications of the seascape metaphor with respect to the incorporation of environmental effects into theoretical population genetics, for understanding how the environment shapes evolution in disease systems, and for contemporary bioengineering excursions.

362W Ascertainment biases empirical fitness landscapes

Maryn O Carlson[1,2], Caelan Brooks[3], Michael M Desai[3,4] [1]NSF-Simons National Institute for Theory and Mathematics in Biology, [2]Department of Statistics, The University of Chicago, [3]Department of Physics, Harvard University, [4]Department of Organismic and Evolutionary Biology, Harvard University

How epistatic are fitness landscapes? That is, to what extent do interactions between loci shape the genotype-to-fitness map? High-throughput genotyping and phenotyping techniques now enable investigators to estimate genotype-to-phenotype maps at scale. A common experimental approach is to generate all possible 2^K sequence intermediates between an ancestral and an evolved sequence differing by K mutations. Summary statistics, including the variance explained by a regression model incorporating interactions up to a given order, the number of local maxima, and the number of accessible paths, are often used to characterize the amount of epistasis in an empirical landscape. We study the degree to which the ascertainment scheme biases these statistics relative to those admitted by the global fitness landscape in several well-studied landscape models. We show, for example, that selecting mutations which occurred along an adaptive trajectory inflates the variance explained by lower order epistasis, reduces the number of local maxima, and increases the number of accessible paths relative to a non-ascertained fitness landscape. This theoretical analysis sets the stage for more careful interpretation of empirical landscapes and experimental design.

363W Global epistasis constrains evolution across environments.

Jhelam Nitin Deshpande, Sarah Ardell, Sergey Kryazhimskiy Department of Ecology, Behavior and Evolution, University of California, San Diego

Predicting how populations adapting to one (“home”) environment lose or gain fitness in other (“non-home”) environments is a fundamental question in evolutionary biology with implications for many practical problems such as antibiotic resistance, climate change and emergence of novel pathogens. Yet, very little is known about the structure of the joint distribution of fitness effects (JDFE) of new mutations, which dictates the dynamics of collateral evolution. Recent studies have found that many mutations exhibit “global epistasis”, meaning that their fitness effects are negatively linearly related to the fitness of the background genotype into which they are introduced. In particular, our previous work (Ardell et al, Science 2024) found that transposon-insertion mutations in yeast Saccharomyces cerevisiae exhibit global epistasis across multiple environments with largely invariant slopes. Here, we examine correlations between the effects of these mutations across environments and investigate how global epistasis constrains the structure of the JDFE and the resulting collateral evolution. Specifically, we first characterize empirical JDFEs of transposon-insertion mutations in yeast and find that they exhibit a variety of shapes and correlation structures. We then develop a statistical model for cross-environmental global epistasis which allows us to explain some properties of these empirical JDFEs. Finally, we examine the collateral effects of adaptation in a home environment on non-home fitness in our model. Most importantly, our model predicts that, as populations adapt at home, their fitness in a large class of non-home environments converges to a finite value determined by the structure of global epistasis. Our model provides the first empirically grounded and testable prediction for the shape of JDFEs. If this model is correct and sufficiently general, it would enable predictions of evolutionary dynamics in temporally and spatially variable environments.

364W Interpreting empirical fitness landscapes under a combinatorial fitness-landscape model

Shoshana S Elgart, Noah A Rosenberg Biology, Stanford University

In investigating the empirical properties of fitness landscapes, evolution experiments have studied the number of paths that are possible on the binary hypercube between an initial genotype and a high-fitness endpoint. We examine mathematical properties of a simple random fitness landscape model on the binary hypercube in light of an experiment in which a high-fitness genotype required mutations at 5 biallelic loci to occur, and 18 of 5!=120 possible sequences of the 5 mutations were able to reach the high-fitness endpoint (Weinreich et al., *Science* 312: 111-114, 2006). In a well-studied simple model of random landscapes on the binary hypercube, fitnesses are randomly assigned to the 2^n possible genotypes at n biallelic loci. Under this model, we obtain new mathematical results describing the number of landscapes on which precisely p paths lead from a starting point to an endpoint. In simulations designed to mimic the motivating 5-mutation experiment, we find that the probability is 1.06% that at least 18 of 120 paths reach the high-fitness genotype. Although the number of mutational paths to the optimum was initially viewed as small, we find that across possible fitness functions, it is relatively large. We perform computations to interpret a variety of subsequent similar experiments, finding that the number of mutational paths to the optimum is often unexpectedly large in relation to the predictions of the model. The analysis contributes to the understanding of fitness landscape experiments involving biallelic loci.

365T The genetic architecture of an ancestral neofunctionalization

Jose Fabricio Lopez Hernandez[1], Jaeda E Patton[1,2], Santiago Herrera Alvarez[1,3], Brian P Metzger[1,4], Joseph W Thornton[1,2,5] [1]Ecology and Evolution, University of Chicago, [2]Committee on Genetics, Genomics and Systems Biology, University of Chicago, [3]Developmental Biology Unit, European Molecular Biology Laboratory, [4]Department of Biological Sciences, Purdue, [5]Human Genetics, University of Chicago

Neofunctionalization after gene duplication is an important mechanism for functional diversification. To explain the evolution of functional transitions, we need to distinguish between single and combinatory substitutions that were critical to evolve the new function. Here we investigate the functional evolution of the ancestral ketoesteroid receptor, AncSR2 (DNA binding site SRE), the product of a gene duplication and neofunctionalization from AncSR1 (DNA binding site ERE). We used a reference-free analysis to access the genetic architecture by quantifying the effects of all the single (additive) and combined (epistatic) effects of ancestral and derived substitutions, and their ability to bind ERE and SRE.

The genetic architecture of the AncSR2 new function was determined by 19 out of 34 historical substitutions, while the remaining had virtually neutral effects on function. Of the 19 substitutions, only five changed the DNA-binding specificity to SRE. The rest of the substitutions affected binding in ERE and SRE equally. We also found that the change in the effect of a substitution is caused by the epistatic interactions of the new substitution with the protein background, and with the different DNA binding sequence. Three derived substitutions, with large positive epistatic effects, together with other three substitutions, that increase SRE specificity, were sufficient to acquire specific-SRE binding. These observations show how 56% of substitutions had a historical relevance in the evolution of a new gene while only 17% can be sufficient to confer a completely new function.

To evaluate the genetic mechanisms and all possible evolutionary scenarios that caused the SRE-specificity of AncSR2, we built a directed network of variants connected by single amino acid substitutions informed by the genetic code. We found that connectivity is larger between functional variants, reducing the possibility of non-functionalization. By evolving from AncSR1 (ERE) to AncSR2 (SRE) we observed that SRE-specificity evolved mostly from promiscuous intermediates that accumulated combinations of the five substitutions associated SRE-specificity. The phenotypic robustness, of ERE binding, delayed the appearance of the fully SRE-specific phenotype when evolved from AncSR1. This work highlights how the genetic architecture enables the gaining of new functions and how evolutionary processes, constrained by the genetic code, create different paths in the evolution of new functions.

366T The distribution of fitness effects of new mutations in regulatory regions of *D. melanogaster*

Austin Daigle[1], Jacob Marsh[2], Andrew Kay[3], Parul Johri[1,2] [1]Genetics, University of North Caronlina at Chapel Hill, [2]Biology, University of North Caronlina at Chapel Hill, [3]University of North Caronlina at Chapel Hill

Although non-coding regions play important roles in gene regulation and contribute to individual fitness, the precise distribution of fitness effects (DFE) of new mutations in these regions remains poorly understood. Here, we carefully compile experimentally validated non-coding regulatory regions in the *Drosophila melanogaster* genome and identify putatively neutral sites near them. Incorporating a realistic genomic architecture that mimics the placement of the regulatory and coding regions, as well as a realistic heterogeneity in recombination and mutation rates across the genome, we use forward-in-time simulations to assess the power and accuracy of population genetic approaches that infer the DFE of new mutations in these regions. While the parameters of DFEs primarily comprising moderately and strongly deleterious mutations are estimated accurately, those of a DFE comprising mostly mildly deleterious mutations are misinferred, leading to an underestimation of the mildly deleterious class. Applying these insights to empirical data from three geographically distinct African *D. melanogaster* populations, we estimate the DFE of functionally important non-coding regions. We find that a large fraction of new mutations in these regions are moderately deleterious, as opposed to strongly deleterious in coding regions, and the fraction of beneficial substitutions in regulatory regions was estimated to range from approximately 0.25 to 0.45, lower than that estimated in coding regions (~0.5). Overall, our results suggest that non-coding regions contribute a majority of new deleterious mutations and beneficial substitutions in *D. melanogaster* populations.. By incorporating both the genomic distribution and the inferred DFE of non-coding regions into a model of background selection, we demonstrate that the effects of background selection across the genome are more accurately captured than a model with coding regions alone, highlighting the importance of considering selection on non-coding regions when interpreting patterns of genomic variation.

367T Beyond Misfolding: How Cellular Context Reshapes a Deep Mutational Fitness Landscape

Ashley N Brown[1,2], Natalie Quan[1,2], Kerry Geiler-Samerotte[3,4] [1]School of Life Sciences, Arizona state university, [2]Center for Mechanisms of Evolution, Arizona State University, [3]School of Life Sciences, Arizona State University, [4]Center for Mechanisms of Evolution; School of Life Sciences, Arizona State University

Mutations can reduce fitness even when a protein's native function is irrelevant. Such costs are referred to as "collateral" and arise purely from how a mutant protein behaves in the cell, through toxic aggregation, aberrant interactions, or the disruption of energetic balance. In a deep mutational scan of yellow fluorescent protein (YFP) expressed in *Saccharomyces cerevisiae*, many variants show fitness defects that are not explained by the toxicity of misfolded proteins alone. If not misfolding, then what cellular limitation explains these fitness defects?

We address this using a deep mutational scan of YFP variants, where fitness differences reflect collateral effects by design. We quantify variant fitness through pooled competitions across environments that probe distinct proteostasis routes and cellular demands. We vary expression load and temperature to tune protein burden, and we apply pathway-targeted perturbations to the proteasome, vacuolar proteolysis, Hsp90, and membrane composition, each with matched vehicle controls.

By comparing each variant's response across environments, we aim to identify classes of mutations whose collateral costs are selectively buffered or exacerbated in specific cellular contexts. This work maps genotype-by-environment structure in collateral fitness effects, implicating mechanisms that convert molecular damage into evolutionary burden.

368T How Cancer Cells Undergoing a Treatment Course Evolve to Survive Multiple Mass Extinctions

Adam Cheong[1], Mark Cheong[2], Paul Fong[3] [1]Proof School, [2]UC Berkeley, [3]Duke University

The vast diversity of genetic and phenotypic characteristics observed in human cancers, according to Hanahan (2026), can be understood in the context of how cancers arise via multistep pathways of tumorigenesis that lead to the formation of primary tumors, which subsequently evolve and progress due to selective advantage in the face of multifaceted barriers to continuing proliferative expansion, resulting in metastasis and adaptive resistance.

Here we elucidate the adversarial viewpoint: a *press-pulse* framework borrowed from paleobiology (Arens & West, 2008) for modeling therapeutic intervention targeted at the primary tumor, whose microenvironment undergoes *terraforming* and subsequently becomes unfit for cancer cells proliferation, leading to multiple *mass extinctions* over the course of cancer treatment, finally resulting in remission as the residual cancer cells turn quiescent.

The population dynamics driven by tumor growth vs. dose response is captured in our proposed model, as the cancer cells population follows a Gompertzian growth curve after metabolic reprogramming but shrinks in response to drug treatment. Tumoral evolution within our framework offers mechanistic insights into the structure of a shifting fitness landscape characterized by varying residual regrowth and dynamic dose responses, analogous to how Lotka-Volterra equations describe the population dynamics of snowshoe hares and the lynx that feed upon them.

We propose to construct a simulation testbed based upon this key insight from population dynamics for model validation. To shrink a tumor over multiple treatment cycles, dose responses are modeled in our testbed as diffusion of antibody-drug conjugates (Madhusoodanan, 2024) across blood vessel walls to cancer cells at various tumor depths, driven by concentration gradient according to Fick's laws.

We illustrate how residual subpopulations of cancer cells evolve to survive multiple mass extinctions originated from a prescribed course of antibody-drug conjugate targeting the primary tumor, potentially laying the seeds for future metastasis or adaptive resistance in the unfortunate event of a relapse.

References:

Hanahan, D. (2026). Hallmarks of cancer—Then and now, and beyond. *Cell, 189*.
Arens, N. C., & West, I. D. (2008). Press-pulse: a general theory of mass extinction? *Paleobiology, 34*(4), 456–471.
Madhusoodanan, J. (2024, March 1). These cancers were beyond treatment—but might not be anymore. *Scientific American.*

369T Temperature Seasonality and Water Vapor Pressure as Selective Drivers of Local Adaptation in the Woodland Strawberry

Oluwafemi Alaba, Daniel Koenig Department of Botany and Plant Sciences, University of California, Riverside

Landscape genomics is an approach for understanding the genetic basis of local adaptation in wild crop species. In this study, we used the European woodland strawberry (Fragaria vesca) as a model system to determine the selective drivers of local adaptation and identify adaptive loci in 179 geo-referenced individuals spanning seven distinct geographical regions. By integrating multiple genotype-environment association methods, 95 climate-adaptive loci potentially involved in local adaptation were identified. Our results revealed temperature seasonality (bio_4), isothermality (bio_3), wind speed, and water vapor pressure as the most significant environmental factors driving adaptive processes. The top adaptive candidate genes identified, associated with plant development and stress responses, include Ethylene Responsive Factor (ERF) and Heat Shock Proteins (HSPs). Genetic vulnerability analysis under projected climate scenarios indicated that Fragaria vesca populations in Iceland, along with those in other northwestern regions, are expected to face a risk of future maladaptation. Overall, these results deepen our understanding of local adaptation and highlight candidate adaptive loci that can be used to enhance climateresilient breeding and crop management.

370W Unusual population frequencies of transposable element in *Drosophila Affinis* and *Drosophila Algonquin*

Mark Xu[1], Grace Y Lee[2], Yuheng Huang[1], Robert L Unckless[3], Taylor Conway[3], Ryan R Bracewell[4] [1]University of California, Irvine, [2]Ecology and Evolutionary Biology, University of California, Irvine, [3]The University of Kansas, [4]Indiana University

Transposable elements (TEs) are DNA sequences that can self-replicate, jump, and spread in an organism's genome. Their activity leads to disruptive mutations and structural variations, and they play a major role in genome evolution. Studies in *Drosophila melanogaster* and many other *Drosophila* species have consistently found that most TE insertions have low population frequencies. This highly skewed frequency spectrum of TEs is usually explained as a result of strong selection against their deleterious fitness impacts. However, a study in the 1980s found surprisingly high population frequencies for the majority of TEs in two *Drosophila* species, *D. affinis* and *D. algonquin*, suggesting possible distinct TE evolutionary dynamics from what has been observed in other species. Yet, this old Southern-blot-based discovery has not been re-examined with modern sequencing technologies, and both species currently lack high-quality TE annotations. To address these questions, we used EarlGrey and manual curation to annotate TEs in the long-read-based genome assemblies of both species. We also developed a novel computational pipeline that involves local assembly to distinguish the presence and absence of TE insertions in putative heterozygous individuals. Results from 24 resequenced, wild *D. affinis* iso-female lines reveal that TEs indeed have high frequencies in the species, with over 70% of TE insertions fixed in the population. We are extending our analysis to *D. algonquin* and other closely related species in order to identify when this high-TE-frequency characteristic emerged on the phylogenetic tree. Our identification of unusual TE population frequencies in *D. affinis* and other yet-to-be-identified *Drosophila* species will serve as powerful new models for studying the evolutionary dynamics of TEs. In particular, many hypotheses regarding the impact of effective population size and epigenetic silencing of TEs can be tested to determine the mechanisms driving the distinct differences in TE population frequencies between *D. affinis* and other species.

371W Expression variability following small-scale duplication facilitates gene retention

Haoran Cai[1], Dave Des Marais[2] [1]Department of Ecology and Evolutionary Biology, University of California Los Angeles, [2]Civil and Department of Environmental Engineering, MIT

Genome analyses reveal that gene duplication in eukaryotes is pervasive, providing a primary source for the emergence of new genes. However, the mechanisms influencing the probability of early duplicate retention and the emergence of functional biases---such as the enrichment of tandem duplicates in environmental responses---remain unclear. Here, to better understand the mechanisms and factors determining gene retention, we study a frequently overlooked molecular feature---expression variability---as measured by within-line expression variation, termed variability. We demonstrate that, on average, genes with duplicates exhibit higher expression variability than singletons. Furthermore, small-scale duplications (SSDs) and whole-genome duplications (WGDs) display contrasting functional outcomes and time-dependent profiles in expression variability. These findings suggest a potential overarching mechanism that facilitates gene expression divergence, functional gains of environmental responses, and duplicate retention following SSDs.

372W The Ghost of Hubby and Lewontin: Navigating the Uncharted Wilds of the Drosophila Genome

Harsh Girish Shukla, Grace Yuh Chwen Lee, J.J. Emerson Ecology and Evolutionary Biology, University of California, Irvine

In 1966, Hubby and Lewontin articulated four requirements for assaying genetic variation: detect allelic differences in single individuals, distinguish substitutions at different loci, resolve most allelic variants, and sample the genome without bias. Protein gel electrophoresis satisfied these criteria for coding sequences, revolutionizing population genetics. Six decades on, even long-read sequencing fails to meet these requirements for highly repetitive regions encoding essential functions: centromeres, telomeres, rDNA, histones, and the Y chromosome. These regions remain the uncharted wilds of the genome, their variation systematically undersampled and functional consequences largely unknown.

I present research addressing this barrier through two interconnected projects. First, I systematically characterize genetic variation across recalcitrant repetitive regions in Drosophila melanogaster by generating reference-quality assemblies from natural populations. Platform comparisons reveal critical biases: PacBio HiFi excels at transposable element-derived beta-heterochromatin and macrosatellites but fails on simple satellite arrays due to severe sequencing bias against abundant repeats like AAGAG. Integrating Oxford Nanopore R10 data overcomes HiFi limitations, achieving near-telomere-to-telomere assemblies. This work establishes technical foundations for finally meeting Hubby and Lewontin's requirements in genomic dark matter, enabling unbiased surveys of variation in functionally critical loci.

Second, leveraging these assemblies, I investigate functional consequences of variation in repetitive regions by examining how polymorphic transposable elements shape 3D genome architecture. I developed novel Hi-C analyses enabling allelic comparisons of spatial interactions with pericentromeric heterochromatin. Nearly 40% of euchromatic regions harboring TEs show enhanced proximity to heterochromatin compared to TE-free homologous alleles. Critically, TEs involved in heterochromatin interactions reduce adjacent gene expression.

By developing methods to properly characterize variation in genomic dark matter, we can finally address fundamental questions about genome evolution that have remained in the uncharted wilds since Hubby and Lewontin first articulated their vision.

373W Revealing Ancestral Functions of Human Duplicated Gene Families through Large-Scale Zebrafish Screening

Eleni Katsougia, Ashley N. Ou, Shriya R. Kota, Megan Y. Dennis Genome Center, University of California Davis

Gene duplication is a fundamental driver of evolutionary innovation, providing genetic material that facilitates functional divergence and increased organismal complexity. Human-specific duplicated (HSD) genes have been implicated in brain development and are hypothesized to contribute to neuroanatomical and cognitive traits that distinguish modern humans from other primates. However, evolutionary interpretation remains limited because the ancestral functions of most duplicated genes are unknown, and functional characterization has largely focused on a small number of paralogs. A major challenge is the absence of scalable in vivo systems capable of interrogating gene function across many families in parallel. While mammalian models and organoid systems have yielded important insights, they remain relatively low throughput. Zebrafish provide a powerful comparative and experimental system due to their rapid development, external fertilization, and high genetic conservation with humans (~75% of human genes have zebrafish orthologs), allowing functional analysis of conserved ancestral genes that predate human-specific duplication events. Here, we perform large-scale functional screening of HSD gene families using mosaic G0 CRISPR-Cas9 knockouts ("crispants") to enable rapid phenotypic assessment. Using high-content imaging and automated feature extraction, we quantify morphological traits—including head size, interocular distance, and body length—during key stages of neuronal differentiation and completion of neurogenesis. Preliminary data from 15 duplicated gene families show that multiple genes exert significant effects on diverse morphological traits, providing in vivo functional evidence linking these gene families to conserved developmental roles. Ongoing work using brain imaging and transcriptomic analyses aims to determine how these morphometric changes influence neurodevelopment in larvae. Together, these findings establish a scalable evolutionary framework for connecting gene duplication to developmental phenotypes and offer insight into how ancestral gene functions may have contributed to the emergence of derived traits and disease susceptibility in the human lineage.

374W HspA1B leads to a poor prognosis of Triple Negative Breast Carcinoma via promoting Tfh cell infiltration

Jian He Shanghai Jiao Tong University School of Medicine

Background: Triple-negative breast cancer (TNBC) presents a critical clinical challenge due to its aggressive nature and limited therapeutic options. The tumor microenvironment (TME), particularly immune cell infiltration, is a key determinant of progression and response. Heat shock protein A1B (HSPA1B) is implicated in tumorigenesis, yet its role in shaping the TNBC immune landscape remains unclear.

Methods: We conducted a multi-omics investigation integrating pan-cancer bioinformatics (TCGA, METABRIC), single-cell transcriptomic analysis of patient samples (GEO), tissue microarray (TMA) analysis of 32 TNBC patients, and functional studies. In vitro and in vivo models using CRISPR-Cas9-mediated HSPA1B/Hspa1b knockout in MDA-MB-231 and 4T1 cells assessed tumor phenotypes. Single-cell RNA sequencing (scRNA-seq) of patient samples and murine tumors elucidated immune microenvironmental changes.

Results: HSPA1B was significantly overexpressed in TNBC and correlated with poor patient prognosis. Knockout of HSPA1B suppressed tumor cell proliferation, migration, and invasion in vitro, and attenuated tumor growth in vivo. TMA analysis of TNBC patients first identified an inverse relationship between high HSPA1B expression and low helper T cell (Th) infiltration, which was associated with worse survival. Subsequent deconvolution analysis of single-cell transcriptomic data (GSE176078) refined this observation, revealing a specific negative correlation between HSPA1B expression and tumor-infiltration of follicular helper T (Tfh) cells. scRNA-seq of knockout murine tumors confirmed a significant expansion of Tfh cells and a reprogramming of CD4+ T cell differentiation trajectories away from immunosuppressive Treg/Th2 fates and towards anti-tumor Tfh/Th1 fates.

Conclusion: Our study identifies HSPA1B as a novel prognostic biomarker and an oncoprotein in TNBC. It promotes tumorigenesis through cell-intrinsic mechanisms, driving proliferation, migration, and invasion, while simultaneously fostering an immunosuppressive tumor microenvironment by subverting anti-tumor Tfh cell differentiation. Targeting HSPA1B presents a promising strategy to simultaneously impair tumor growth and counteract immune evasion.

375W Capturing a TE Burst in Real-Time: *mPing* Dynamics and Structural Variation in Rice RIL Populations

Nathan Mathieu UC Riverside

Transposable element (TE) bursts are characterized by the rapid amplification of TE copies in a host genome. Although many of these ancient evolutionary events have been well documented, understanding their direct impact on the genome requires observing a burst in real time. *mPing* is the first active miniature inverted transposable element (MITE) identified and the first active Class 2 element found in rice. Genomic analysis of 3,000 domesticated rice accessions has shown that the mPing burst is restricted to only 4 specific accessions within the temperate japonica subspecies of *Oryza sativa*. To model the spread and dynamics of the burst, three recombinant inbred line (RIL) populations were created through crosses of the 4 bursting rice accessions. The resulting ~700 RILs across the three populations capture a continuous range of burst activity and newly formed variation that is not shared with the parents. With the regular use of long read sequencing, we now see that variation among individuals cannot be represented with a single collapsed reference genome, and as new pangenome references are created, these can serve as tools to help identify variation in newly sequenced individuals. Analysis of the bursting accessions revealed patterns of *mPing*-associated complex structural variation (SV) providing evidence of a TE mediated mechanism generating variation underlying many pangenomes.

376W Domestication as an evolutionary driver of telomere length in angiosperm plants

Linh K Nguyen, Jae Y Choi Ecology and Evolutionary Biology, University of Kansas

Telomere is a nucleoprotein complex that protects chromosome ends from genome instability and damage. Length of the telomere can vary significantly between species, but what function and evolutionary processes are maintaining this variation has remained largely elusive. While in animals, longevity and senescence have explained telomere length differences between species, in plants, there is no hypothesis to explain the length variation. In this study, we attempted to address this question by analyzing decades of long read sequencing data generated for angiosperm plants. We developed Topsicle, a novel computational method that estimates telomere length from long read sequencing data using k-mer distribution profiles. Topsicle can analyze organisms with any telomere repeat sequence, does not require a reference genome, and accommodates error rates in long read sequencing data for estimating telomere length. It has comparable results to direct telomere length measurements and achieved high accuracy when examining telomere length in *A. thaliana*, maize, rice, and human data, demonstrating its capacity to investigate telomere length dynamics across species. We applied Topsicle on 134 angiosperm plant species with 872 libraries that spanned 22 orders. Telomere length varied significantly between species, ranging from 800 bp to 18370 bp, and comparing between plant orders, the median telomere length was lowest for Brassicales and highest for Vitales order. We then investigated life-history traits (e.g., annual versus perennial, woodiness, latitude versus longitude, and other traits) that could explain telomere length variation in angiosperms. Interestingly, domestication was a noticeable trait associated with telomere length, where multiple species in eudicots and monocots showed significant length differences between the wild progenitor and domesticated plant species. The direction of difference was idiosyncratic and not dependent on phylogeny, suggesting additional life-history traits were involved in explaining the telomere length variation. In the end, our results suggest human-mediated agroecology may have selected for telomere length differences in domesticated plants.

377W Shared signatures of gene expression in *I-R* dysgenesis and ovaries with excess DNA damage

Ekta Mohanty, Justin Blumenstiel Ecology and Evolutionary Biology, University of Kansas

Transposable elements (TEs) act as genetic parasites, often causing DNA damage and potentially impacting genome stability. In *Drosophila melanogaster*, *I-R* hybrid dysgenesis (HD) is induced when males carrying active *I*-element retrotransposons are mated with females that lack them. Instead of being activated during early development, the *I*-element is exclusively activated in the germline of adult female progeny. Due to this critical difference, HD in the *I-R* system causes a failure in female gametogenesis which results in low hatch rate. We wanted to understand the mechanism underlying failure to hatch in *I-R* dysgenesis. In our efforts to identify the mechanism in dysgenic ovaries, we sought to determine whether there are shared signatures with ovaries in which DNA damage is known to cause a failure to hatch. Our RNA-seq analysis of dysgenic ovaries revealed two major classes of response- mitochondrial and ribosomal. Similarly, RNA-seq analysis of ovaries from *Spn-A (Rad51)* knockdown ovaries showed a shared response. To further investigate the overrepresented mitochondrial response as a consequence of DNA damage, we plan to knock out *p53* and *chk2* to assess their roles in this process. These findings will contribute to understanding the link between DNA damage and cellular stress responses.

378W Integrating long-read assemblies and ancestry graphs to map Neanderthal-introgressed structural variation in humans

Charikleia Karageorgiou[1], Thomas J. Hitchcock[2], Omer Gokcumen[3], Leo Speidel[2] [1]Department of Biological Sciences, University at Buffalo, [2]RIKEN Center for Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS), [3]University at Buffalo

Structural variation (SVs) constitutes a major yet historically undercharacterized component of human genomic diversity with potentially large phenotypic effects. Despite their importance, the evolutionary trajectories of SVs have remained challenging to reconstruct, owing to limitations of current population genetic methodology that mainly leverage single-nucleotide differences and due to the limited resolution of short-read sequencing for calling such variants in modern and ancient DNA. Here, we leverage Neanderthal introgression as a window into the evolutionary consequences of complex structural variation. We integrate short-read-based ancestry inference with long-read genome assemblies to systematically characterize structurally introgressed variation and characterise how structural variants emerge, are maintained in a population, and have shaped patterns of Neanderthal introgression in modern humans. Using Ghostbuster, we first delineated Neanderthal-introgressed haplotypes in unrelated individuals from the 1000 Genomes Project. We then performed structural variant discovery on 578 phased long-read assemblies from the HPRC and HGSVC, followed by genotyping of inversions, duplications, deletions, and mobile element insertions (MEIs). Candidate introgressed SVs were identified by intersecting long-read-resolved variants with short-read-defined introgressed tracts in matched individuals, and subsequently evaluated using linkage disequilibrium and population frequency patterns to distinguish introgression from incomplete lineage sorting (ILS). Our preliminary analyses identify 15 candidate introgressed inversions, all relatively small, generated by nonhomologous end joining, and restricted to intergenic regions. In contrast, MEIs represent a larger source of archaic structural variation, with 79 candidate introgressed insertions and 405 additional MEIs that are potentially older and consistent with ILS. The predominant introgressed MEIs belong to the AluYb8 and AluYa5 subfamilies, consistent with their known activity in recent human evolution. Of the 79 candidate introgressed MEIs, three occur near genes encoding odorant receptors, suggesting potential sensory-related impacts. Collectively, these results reveal striking differences among SV classes in their persistence following archaic introgression and underscore the power of combining long-read assemblies with ancestral recombination graph methods to resolve the evolutionary dynamics of structural variation.

379W T2T genomes of *Caenorhabditis nigoni* and *Caenorhabditis briggsae* reveals extensive loss of satellite DNA associated with self-fertilization

Ryan Pellow[1], Tejus Kotikalapudi[2], Ofer Rog[1] [1]University of Utah, [2]University of Southern Mississippi

The two closely related *Caenorhabditis* nematode species, *C. nigoni* and *C. briggsae*, are commonly used to study the evolution of reproductive modes in animals, with the self-fertile *C. briggsae* and outcrossing *C. nigoni* sharing a common ancestor ~3.5 million years ago. Earlier genome assemblies of these species revealed gene loss associated with selfing and proposed that gene loss can be adaptive and promote genome shrinkage. However, the incomplete *C. nigoni* reference genome limited most comparative analyses to genic regions. Here, we leveraged long-read sequencing to generate a telomere-to-telomere (T2T) assembly for the *C. nigoni* strain JU1422 and the *C. briggsae* strain AF16. This new 139Mb *C. nigoni* genome resolved 57 gaps and 149 unassigned scaffolds from the previous genome assembly. Comparison with the 107Mb T2T *C. briggsae* genome reveals that the major driver of genome content differences are deletions to satellite DNA arrays. Interestingly, many of the differences are on the *C. nigoni* X chromosome, which is >13Mb larger than in the previous assembly. The transition to selfing was thus accompanied by a 37% reduction in the size of the sex chromosome compared to 16-21% shrinkage of the autosomes. Surprisingly, the X chromosome also harbors a second 45S rDNA array that is absent in *C. briggsae*. Our analysis reveals that obligatory outcrossing may play a major role in the maintenance of satellite DNA arrays.

380W The tortured past of young polymorphic sex chromosomes revealed through multiple de novo genome assemblies of the mountain pine beetle

German Lagunas-Robles[1], Barbara Bentz[2], Scott Geib[3], Ryan Bracewell[1] [1]Indiana University Bloomington, [2]USDA FS RMRS, [3]USDA ARS

Neo-sex chromosomes provide a powerful system for studying the early stages of sex chromosome evolution and the genomic mechanisms that may contribute to reproductive isolation. Using PacBio long-read HiFi sequencing, Hi-C scaffolding, and sex specific transcriptomic data, we generated six chromosome-level assemblies (male and female from three populations) of the mountain pine beetle (*Dendroctonus ponderosae*), a species known to harbor three partially reproductively isolated neo-Y haplogroups. These assemblies reveal that the large neo-X and neo-Y chromosomes formed through sequential fusions of the ancestral X with three autosomes, with recombination cessation occurring at ~8.6, ~6.3, and ~4.3 MYA for each event. Comparative analyses show that while neo-X chromosomes remain largely collinear across populations, neo-Ys exhibit dramatic structural divergence, with 900–1,200 inverted segments per haplogroup and only ~65% of sequence able to be aligned to the neo-X. Repeat analyses demonstrate moderate TE accumulation on the neo-Y, particularly LTR elements, and gene mapping analyses reveal extensive degeneration: ~62% of neo-Y genes exhibit gene loss, fragmentation, or disruptive mutations. All populations retain a single pseudoautosomal region (PAR), though PAR size and gene content vary due to neo-Y specific rearrangements. Across neo-Ys, 27 genes are uniquely missing in the Western haplogroup, including previously identified candidates implicated in hybrid male sterility. Broader comparisons among neo-Ys show widespread structural variation, population specific patterns of degeneration, and limited gene family expansions. Together, these results provide the first full characterization of neo-sex chromosome evolution in *D. ponderosae*, revealing rapid, lineage specific neo-Y degeneration and highlighting the potential for sex chromosome divergence to contribute to emerging reproductive incompatibilities within a single species.

381W Gene expression evolution across primate cardiomyocyte differentiation

John D Hurley, Alyssa R Bogar, Jose A Gutierrez, Michelle C Ward University of Texas Medical Branch

The hourglass model of development posits a phylotypic stage in embryonic development where conservation of morphology and gene expression is observed across diverse vertebrate species. This model has not been investigated across primates for ethical and technical reasons. With the advent of induced pluripotent stem cell (iPSC) technology, developmental processes can now be modelled across humans and chimpanzees. To determine the level of conservation of gene expression across primate development, we differentiated iPSCs from seven humans and seven chimpanzees into cardiomyocytes. We collected cells at five differentiation stages – iPSCs (Day 0), mesoderm (Day 2), cardiac mesoderm (Day 5), early cardiomyocytes (Day 15) and late cardiomyocytes (Day 30) to profile gene expression. Stage-specific protein marker expression is highest in the corresponding stage for both species, and there is no difference in the proportion of cells representing each stage across species. Global gene expression data across 70 samples identifies differentiation stage as the largest contributor to expression variation, followed by species. Jointly modeling the data reveals 11 gene expression trajectories that are largely conserved across species. When considering each differentiation day independently, we identify thousands of differentially expressed genes between species with the fewest observed at the mesoderm and cardiac mesoderm stages. These stages also show the lowest absolute species effect sizes. Pairwise correlations of effect sizes between temporally-adjacent days show highest inter-species correlations between days 2 and 5 and days 5 and 15. Our primate data are consistent with the hourglass model of development with most inter-species conservation at mid-developmental stages.

382W Heritability of germline mutagenesis in 40 large three- and four-generation pedigrees

Michael E Goldberg, Alexis C. Garretson, Camila Goclowski, Thomas A. Sasani, Hannah C. Happ, Julia Ostrander, Lynn B. Jorde, Deborah W. Neklason, Aaron R. Quinlan Department of Human Genetics, University of Utah

Germline mutations are the basis for genetic disease but also underly all heritable phenotypic variation on which evolution can act. Estimating the mutation rate, or the expected number of *de novo* mutations (DNMs) per site, genome, and generation, is therefore critical for modeling disease burden and selection. Mutation rate is a polygenic trait, affected by both genetic and environmental factors that modulate DNA damage, repair, and replication pathways. In the human germline, non-heritable traits such as parental sex and age strongly predict mutation rate; few studies, however, have been able to identify specific genetic loci that commonly affect mutation rate or measure its heritability. Nevertheless, higher germline mutation rate is associated with earlier mortality, hinting at shared architecture between mutagenesis and health.

Here, we present results from the first phase of the Gametes Through Generations (GTG) project, which comprises new whole genome sequencing of >1000 individuals from 22 four-generation and 18 three-generation CEPH/Utah pedigrees. Each family's third and fourth generations include a median of eight and four children per couple, respectively. Prior studies of mutation rate heritability in humans have been indirect and limited to single-nucleotide mutations observed in mother, father, and child "trios". In contrast, the large GTG pedigrees allow us to measure germline mutation in hundreds of individuals, eliminate false positives, and assign mutations to a parent-of-origin.

Because germline mutagenesis is a low count Poisson process, its inherently low signal-to-noise ratio clouds inference of heritability, especially in trios. Indeed, we find much higher power to detect nonzero heritability in GTG using simulated data and test traits. The large number of children per couple also allows us to measure repeatability, a statistic that marks an upper bound for its heritability, scales negatively with shot noise, and has never been inferred for mutation rates. We find that, even in GTG, the high variance in mutation rate contributes to a low repeatability of 0.15 and 0.51 for maternal and paternal mutation rates, respectively. Thus, detection of nonzero maternal mutation rate heritability may be impossible given ours and other modern datasets.

Accordingly, initial results using GTG DNMs show low to zero heritability for both maternal and paternal mutation rate. While preliminary, these findings indicate that the GTG dataset provides novel resolution critical for accurately estimating the heritability of germline mutagenesis and, as we will present, related genomic traits. The sequencing of these pedigrees has broad implications for both molecular evolution and genomic medicine, helping quantify an individual's unique risk for mutational burden.

383W Single-fly long-read assemblies enable comparative analysis of repetitive genome regions in *Drosophila melanogaster*

Alejandra Samano, Mahul Chakraborty Biology, Texas A&M University

A major limitation of genome analysis is that many of the most structurally complex and repetitive regions - such as centromeres, satellite arrays, and other repeat-rich loci - remain inaccessible using standard short-read sequencing approaches and are often fragmented or missing even from high-quality reference assemblies. As a result, comparative analyses across strains and populations have largely excluded these regions, constraining our understanding of genome architecture and repeat evolution. Here, we use a long-read, single-fly assembly strategy to recover and compare highly repetitive genomic regions across multiple *Drosophila melanogaster* genomes. Using PacBio HiFi sequencing, we generated high-quality *de novo* assemblies from individual flies sampled from two wild tropical populations (American Samoa and Guam), along with the isogenic strain A4. All three assemblies contain contiguous sequences that fully span the centromere of chromosome 3 and extend the X chromosome deep into the 1.688 satellite array - regions that are typically unresolved or absent from *Drosophila* genome assemblies. These assemblies enable direct comparison of centromeric and satellite organization across strains without the confounding effects of pooling or reference bias. Ongoing work integrates these genomes into a pangenome framework and incorporates long-read RNA sequencing data to examine the transcriptional consequences of structural and repetitive variation. Together, this work demonstrates that single-individual long-read assemblies provide a powerful approach for comparative analysis of repetitive genome regions, opening new opportunities to study variation in genome structure and repeat content that has been largely inaccessible to evolutionary and genomics analyses.

384W Escape from X inactivation drives sex differences in gene expression

Carrie Zhu, Liaoyi Xu, Arbel Harpak UT Austin

X chromosome inactivation (XCI) partially balances gene dosage between sexes, yet many genes are expressed from the inactive X (Xi) to a variable degree. In this study, we investigate whether variation in Xi expression among genes predicts transcriptional and phenotypic consequences of X-linked variation. We find that Xi expression levels are a strong linear predictor of female-male expression differences, suggesting that other compensatory or regulatory mechanisms play a more minor role in sex differences in X-linked gene expression. Among females, we identify traits—including BMI, estradiol, and testosterone levels—for which higher Xi expression correlates with the strength of evidence for either additive or dominance effects on the trait. We hypothesize that an underappreciated mechanism could generate dominance effects of X-linked variants on a trait—specifically when the variant influences skew in X inactivation. This work establishes Xi expression as important for understanding transcriptional sex differences and physiological variation among females.

385T Mapping a naturally occurring sex-linked recessive lethal allele in the yellow fever mosquito *Aedes aegypti*

Abdulhadi Kobiowu[1], Atashi Sharma[2], Melanie Hempel[2], Qi Yumin[2], Zhijian Tu[2] [1]Virginia Tech, [2]Biochemistry, Virginia Tech

Aedes aegypti females pose a significant public health threat globally, as they transmit dengue, Zika, chikungunya, and yellow fever viruses. This becomes increasingly important as vector control is hindered by insecticide resistance. Genetic approaches for population suppression and modification benefit from tools that enable reliable female removal for the purposes of male-only release. Sex-linked recessive lethal alleles (RLAs) have been used to improve the efficiency of sex separation. RLAs have been mainly reported on the Y-like chromosome 1 that contains the male-determining locus (M locus). Here we report a naturally occurring RLAs linked to the X-like, m-bearing chromosome 1, which produced a 2:1 male-to-female sex-ratio bias during sib-mating. We generated mapping strains using classical genetics and employed marker-assisted mapping to recover informative recombinants, narrowing the candidate region by localizing crossover breakpoints relative to the genetic markers. In addition, 15 generations of sib-mating led to near-genome-wide approaching to homozygosity except at the RLA. Therefore, we applied forced-heterozygosity mapping, leveraging the expectation that viable carriers must remain heterozygous across the RLA region, to delimit a contiguous RLA-containing block. Finally, pooled whole-genome sequencing followed by genome-wide association analysis fine-mapped the RLA to ~1 Mb resolution. Two candidate genes, one of which has a *Drosophila melanogaster* ortholog with an embryonic-lethal phenotype, are currently undergoing CRISPR validation. This work provides the first evidence of naturally occurring m-linked RLAs in *A. aegypti* and demonstrates how classical population-genetic principles can enable efficient fine-mapping. Characterized m-linked RLAs offer a practical route to female removal, strengthening genetic vector control strategies. In the long term, integrating m-linked RLAs into the existing control strategies could support sustainable, targeted reductions in *A. aegypti*–borne disease risk worldwide.

386T The rapid evolution of centromeric satellite sequences in house mouse lineages

Jitendra Thakur[1], Keenan Wiggins[2] [1]Emory University, [2]Biology, Emory University

Centromeres are chromosomal sites where spindle fibers attach via the kinetochore to enable chromosome segregation. Despite conserved function, centromeric DNA sequences evolve rapidly across eukaryotes. Yeast centromeres are short, sequence-dependent regions (~100–400 bp), whereas mammalian centromeres comprise long tandem satellite repeats (up to several megabases). The centromere drive model explains this paradoxical rapid sequence evolution: centromeric variants that strengthen centromeres preferentially segregate to the egg pole during asymmetric female meiosis, while centromeric proteins subsequently evolve to suppress these deleterious effects. Because established mammalian species exhibit substantial centromeric sequence variation, identifying conserved motifs that define centromere identity remains challenging, especially given the highly repetitive nature of mammalian centromeres. To investigate how DNA sequences contribute to centromere identity, we examined centromeric variants in house mouse (*Mus musculus*) lineages that diverged 0.5 million years ago into Eastern European (*M.m. musculus*) and Western European (*M.m. domesticus*) populations. We used PacBio and Nanopore long-read sequencing to generate high-resolution assemblies of centromeric regions across several *Mus* lineages. Our findings reveal the complex organization of satellite DNA, the prevalence of Robertsonian translocations and inversions, and the diversification of specific motifs within lineage groups. Interestingly, we observed lineage-specific enrichment of transposable elements within centromeres, suggesting a dynamic interplay between repetitive elements and centromeric repeats. Our findings establish a framework for understanding centromere evolution within *Mus* lineages, providing insights into rapidly evolving genomic elements and reproductive isolation.

387T Cell-type- and locus-resolved transposable element expression during *Drosophila* spermatogenesis

Imtiyaz E Hariyani[1], Anirudh Margasahayam[1], Rongying Lu[1], Edward Oropeza-Rodriguez[1], Vivek Jayaswal[2], Jose M Ranz[1]
[1]Department of Ecology and Evolutionary Biology, University of California Irvine, [2]School of Mathematics and Statistics, The University of Sydney

Transposable elements (TEs) represent a fundamental evolutionary challenge to genome integrity, particularly in the germline where insertions can be inherited. Paradoxically, TE expression increases sharply during spermatogenesis, raising questions about the interplay between TE activity, genome defense mechanisms, and large-scale chromatin remodeling during development. In *Drosophila melanogaster*, prior work has shown that this burst of TEs coincides with downregulation of piRNA pathway genes and activation of *Y*-chromosome fertility factors. However, because TE expression has largely been analyzed at the family level, it remains unclear which individual TE loci drive the meiotic burst, whether activation reflects a global loss of repression or is restricted to specific genomic compartments, and how TE activity differs between germline and somatic lineages. Here, we address these questions using single-nucleus RNA sequencing data from the *D. melanogaster* testis together with a strain-matched assembly to enable locus-resolved analyses. Our findings suggest that the meiotic TE expression burst in spermatocytes is driven predominantly by retrotransposons. In contrast, somatic cell types—primarily cyst cells—show elevated expression of DNA transposons, revealing distinct TE expression niches across germline and somatic lineages. In spermatocytes, expressed TEs are disproportionately enriched in pericentromeric heterochromatin and are located farther away from nearby genes compared to non-expressed TEs. Consistent with this organization, retrotransposon expression is strongly anticorrelated with core piRNA pathway gene expression, with *papi* emerging as a candidate gene that may regulate TE activity in the male germline. Together, our results indicate that the meiotic TE burst reflects stage-specific vulnerabilities in heterochromatin-dependent silencing that arise during chromatin remodeling. By resolving TE activity at the level of individual loci and cell lineages, our study reframes the germline TE burst as a heterochromatin-enriched, multi-chromosomal phenomenon, rather than a *Y*-restricted anomaly, demonstrating that TE activation during spermatogenesis is structured by developmental stage, cell type, chromosomal context, and TE class.

388T Convergent evolution of genetic sex determination among species of stickleback fishes

Michael White, Matthew Treaster, Jenny McCann Genetics, University of Georgia

The evolution of sex chromosomes begins with the acquisition of a primary sex determination gene. Although sex differentiation is a remarkably conserved process among vertebrates, the primary sex determination gene that initiates sex differentiation can rapidly evolve across taxa. Stickleback fishes are an excellent model group to understand how newly evolved primary sex determination genes gain this function. The threespine stickleback (*Gasterosteaus aculeatus*) evolved a Y chromosome approximately 22 million years ago. Within the oldest region of the Y chromosome is a copy of anti-Müllerian hormone *(Amh)*, which duplicated and translocated from autosome eight. We have shown the Y-specific duplication of *Amh (Amhy)* is both necessary and sufficient for male development using transgenesis and gene editing approaches. We are now comparing sex determination in the threespine stickleback with a closely related species, the brook stickleback (*Culaea inconstans),* which has an independently evolved Y chromosome at the early stages of differentiation. Population genomics work in this species suggested there was a convergent duplication of *Amh*. To verify this, we have completed a genome assembly and indeed discovered a 7 kb insertion on the new Y chromosome from autosome eight that contains a copy of *Amh*. In contrast to the threespine stickleback, the brook stickleback *Amhy* only has a handful of mutations that differentiate it from the ancestral autosomal paralog, indicating it evolved much more recently. Despite its recent origin, *Amhy* has already evolved multiple unique isoforms that are not detected from the ancestral autosomal locus or from the convergently evolved *Amhy* in the threespine stickleback. Both species will therefore offer unique insights into the different ways *Amhy* regulation can evolve to neofunctionalize as a primary sex determination gene.

389T The mutational and epigenomic landscape in replicated tumor growth: insights from a mouse model of HER2-Low breast cancer

Asad Ahmad, Alex Samano, Kayla Carter, Tapasree Sarkar, Mahul Chakraborty Biology, Texas A&M University

Genomic variation within tumors reflects the combined effects of mutation, selection, and large-scale genome restructuring, producing complex allele frequency landscapes. HER2-low breast cancers, defined by low HER2 expression in the absence of gene amplification, provide a model for studying how structural variation and epigenetics shape genomic heterogeneity. The goal of this study is to define how structural and epigenomic alterations contribute to tumor evolution in HER2-low disease. To capture this at genome-wide resolution, we used Oxford Nanopore long-read sequencing to characterize both structural variation and DNA methylation from the same reads, generating an integrated map of genomic and epigenomic alterations in replicated tumor and normal tissues. Reads were aligned to the reference genome to identify somatic structural variants using Severus and infer CpG methylation using modkit, enabling the detection of both structural mutations and differential methylation between tumor and normal samples. We identified 11,077 differentially methylated regions that contribute to the enrichment of developmental regulators involved in pattern specification, stem cell differentiation, and cell fate determination, including Shh, Hox family genes, Tbx2, Foxc2, and Sox family transcription factors. Structural variant analysis revealed 240 deletions, 193 insertions, one duplication, one inversion, and 42 break-end junctions, 81% of which localized to a 36 Mb region on chromosome 16, with deletions and insertions comprising the majority of events. Analysis of variant allele frequencies (VAF) revealed a broad, multimodal distribution with a dominant peak near 0.5, particularly enriched among deletions and insertions, consistent with a mixture of clonal and subclonal somatic events. Lower-frequency break-ends were also observed, along with a subset of somatic variants with VAF = 1 concentrated on chromosome 16, suggesting large-scale loss of heterozygosity. Together, these results highlight structural and epigenomic remodeling as a defining feature of HER2-low tumor evolution and provide insight into the mechanisms driving genomic instability in this breast cancer subtype.

390T Multiple independent origins of neo-sex chromosomes in *Dendroctonus* bark beetles

Gabrielle Coffing[1], Medhavi Verma[2], Harshita Sharma[1], Scott Geib[3], Ryan Bracewell[1] [1]Department of Biology, Indiana University, [2]Indiana University, [3]Tropical Pest Genetics and Molecular Biology Research Unit, USDA-ARS

Sex chromosomes have repeatedly evolved across the tree of life. Although the general processes that lead to highly heteromorphic X-Y (or Z-W) chromosomes are well established, less is known about the tempo of Y chromosome degeneration following recombination suppression. The early stages of Y degeneration can be studied in species with neo-sex chromosomes, as neo-Ys are initially gene rich due to fusions with one or more autosomes. Dendroctonus bark beetles are an emerging system for studying Y chromosome evolution and degeneration as they are known to have neo-XY chromosomes in multiple lineages. However, the history of sex chromosome to autosome fusions and the degree of Y degeneration in these species has been limited to karyotypic data. Here, we have assembled and annotated chromosome-level genome assemblies for eight Dendroctonus species, comprising sixteen total genomes (male and female from each species) using a combination of PacBio long-reads, Hi-C, and RNA-seq. With the addition of two publicly available genomes, we analyzed three species with putative ancestral chromosomal configurations (14AA + XY) and five species with neo-sex chromosomes (5AA + neo XY, 6AA + neo XY, & 11AA + neo XY). Using comparative analyses, we found there have been three independently formed neo-sex chromosomes in Dendroctonus. By estimating divergence times using gametologs, we show that autosomal fusions to the ancestral X occurred at three distinct evolutionary timepoints, resulting in neo-sex chromosomes of three different ages: young, intermediate, and old. The three species that share the 'old' neo-sex chromosomes have neo-Ys that have experienced extensive gene loss and proliferation of LTR elements. In contrast, a single species with the most recently formed neo-sex chromosome exhibits recombination suppression over only half of the neo-XY pair, with the non-recombining portion of the neo-Y showing TE proliferation. Overall, we find that the Dendroctonus genus is a promising system to explore neo-Y chromosome degeneration at different evolutionary timescales.

391T Gene expression of neo-sex chromosomes in hybrid mountain pine beetle

German Lagunas-Robles[1], Gabrielle Coffing[1], Matt Hansen[2], Jim Vandygriff[2], Medhavi Verma[1], Barbara Bentz[2], Ryan Bracewell[1] [1]Department of Biology, Indiana University, [2]USDA Forest Service Rocky Mountain Research Station

Sex chromosomes are known to play a disproportionately large role in speciation and the evolution of reproductive incompatibilities. Most studies have traditionally focused on understanding the X chromosome since the Y is often highly degenerated, challenging to study, and is not easily implicated in incompatibilities. Young sex chromosomes, so called neo-sex chromosomes, provide an opportunity to study the evolution of reproductive incompatibilities as neo-Y degeneration is an ongoing process. To explore the dynamics of gene regulation and reproductive isolation in a neo-sex chromosome system, we explored the young neo-sex chromosomes of the mountain pine beetle (MPB) *Dendroctonus ponderosae*. MPB is comprised of three Y-haplogroups that vary in the degree of postzygotic reproductive incompatibilities in genetic crosses. Additionally, our recent work showed there were ~60 missing genes from any given neo-Y, although in most cases, the neo-X copy is still present in all populations. In this study, we first generated PacBio long-read Iso-Seq to refine previous neo-Y annotations and help generate a complete list of missing or pseudogenized genes on the different neo-Ys. We then performed reciprocal crosses between populations that 1) have different neo-Ys and 2) show some level of reproductive isolation in the form of hybrid sterility. We generated RNA-seq data (heads and testes) from hybrid males and the pure populations. We then 1) characterized gene dosage, 2) explored aberrant expression in hybrids, and 3) evaluated misregulation of dosage compensated genes in hybrids.

392T Global Patterns of Genetic and Structural Variation in Wild and Cultivated Hops

Alexandra McElwee-Adame, Trevor Mugoya, Priyanshi Shah, Arun Sethuraman Biology, San Diego State University

Humulus lupulus, commonly known as hop, is a cash crop native to Eurasia and North America, with domesticated and selectively bred cultivars now grown across the world. Hops display several agronomically important properties applicable to commercial products in the pharmaceutical and commercial industry, however, they are most extensively used in brewing beer. Commercially, over 250 cultivars (cultivated varieties) have been developed worldwide, each displaying differences in bitter acid/essential oil compositions, benchmarks that play key roles in the brewing process. Through crop improvement efforts, several new cultivars have been developed in recent years, each selected for differences in agronomic traits such as growth, yield, seasonality, disease/pest tolerance as well as abiotic and biotic stress tolerance. Previous genomic work in hops have led to the assembly of several genome-level assemblies, opening the door for pangenomic studies in hops to investigate global patterns of domestication. Here we generated whole genome sequences from 192 wild and cultivated USDA accessions, and developed a Snakemake workflow to orchestrate sequence quality-based filtering, the NVIDIA Parabricks germline pipeline, and subsequent SNP variant calling and processing on the TIDE GPU cluster at San Diego State University. We catalog ~200 million SNPs across globally diverse hops, to generate the first hop population genomic database. Additionally, we implemented a structural variant (SV) calling pipeline across 40 global cultivars to quantify small to large SVs in hop genomes. Our analyses catalog >850k translocations, >172k deletions, 34k duplications, 300 insertions, and >94k inversions across globally structured cultivars (K=4 subpopulations, global Fst ~ 0.05). This study opens doors for future hop research and crop improvement in a warmer world.

393T The dynamic evolution of 5S rDNA

Brandon Fagen[1], Bethannie Noel Caparoon[1], Kobe Koren[1], SaraH Zanders[1,2] [1]The Stowers Institute for Medical Research, [2]2Department of Molecular and Integrative Physiology, University of Kansas Medical Center

Essential genes often maintain high sequence conservation as well as stable synteny over evolutionary time. However, there are striking exceptions. The 5S rRNA genes (5S rDNA) encode an essential component of ribosomes, yet they exhibit strong diversity in their genomic organization and copy number. Individual 5S rDNA genes can be dispersed throughout the genome or exist in long tandem arrays. To explore how such variation arises and is tolerated, we are characterizing 5S rDNA evolution in fungi. Through phylogenetic analysis of 389 species, we found that 5S rDNA genes have switched from arrayed to dispersed organization (or vice versa) at least 22 independent times. This organization appears to affect evolution as dispersed 5S rDNA genes have much higher levels of sequence variability than arrayed copies. This is consistent with cyclic birth-death evolution of dispersed genes, where new copies are generated while other copies degrade, analogous to transposable elements. The mechanisms underlying the gene duplications are not clear, but the birth-death rate of dispersed 5S rDNA genes varies by up to 27-fold between clades. The variability in copy number and prevalence of pseudogenes suggested that dispersed 5S rDNA copies might be in excess in some lineages. To test this, we knocked out 7 of 34 5S rDNA copies in *Schizosaccharomyces pombe*. We observe no fitness effects, supporting a model in which excess copies may decrease the selective constraints on individual 5S rDNA genes. These data shed light on the evolutionary complexity of multicopy gene families and how their organization can affect their evolution.

394T The distinct roles of genome, methylation, transcription, and translation on protein expression in *Arabidopsis thaliana* resolve the Central Dogma's information flow

Ziming Zhong[1], Mark Bailey[2], YongIn Kim[3], Nazanin P Afsharyan[2], Briony Parker[2], Louise Arathoon[1], Xiaowei Li[2], Chelsea A Rundle[2], Andrew Behrens[4], Danny Nedialkova[5], Gancho Slavov[6], Keywan HassaniPak[2], Kathryn S Lilley[3], Frederica L Theodoulou[2], Richard Mott[1] [1]Genetics, Evolution and the Environment, UCL, [2]Plant Sciences and the Bioeconomy, Rothamsted Research, [3]Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, [4]Mechanisms of Protein Biogenesis, Max Planck Institute of Biochemistry, [5]Department of Bioscience, TUM School of Natural Sciences, Technical University of Munich, [6]Institute of Biological Environmental and Rural Sciences, Aberystwyth University

We investigated the flow of genetic information from DNA to RNA to protein as described by the Central Dogma in molecular biology, to determine the impact of intermediate genomic levels on plant protein expression. We performed genomic profiling of rosette leaves in two *Arabidopsis* accessions, Col-0 and Can-0, and assemble their genomes using long reads and chromatin interaction data. We measured gene and protein expression in biological replicates grown in a controlled environment, also measuring CpG methylation, ribosome-associated transcript levels, and tRNA abundance.

Each omic level is highly reproducible between biological replicates and between accessions despite their ~1% sequence divergence; the single best predictor of any level in one accession is the corresponding level in the other. Within each accession, gene codon frequencies accurately model both mRNA and protein expression. The effects of a codon on mRNA and protein expression are highly correlated but independent of genome-wide codon frequencies or tRNA levels which instead match genome-wide amino acid frequencies. Ribosome-associated transcripts closely track mRNA levels.

DNA codon frequencies and mRNA expression levels are the main predictors of protein abundance. In the absence of environmental perturbation neither gene-body methylation, tRNA abundance nor ribosome-associated transcript levels add appreciable information. The impact of constitutive gene-body methylation is mostly explained by gene codon composition. tRNA abundance tracks overall amino acid demand. However, genetic differences between accessions associate with differential gene-body methylation by inflating differential expression variation. Our data show that the dogma holds only if both sequence and abundance information in mRNA are considered.

395T Chromosome-scale recombination landscapes govern de novo gene evolution in humans

Claudio Casola[1], Matthew A Marano[2] [1]Ecology and Conservation Biology, Texas A&M University, [2]Ecology and Evolutionary Biology, Texas A&M University

De novo gene birth is increasingly recognized as a major source of evolutionary novelty, yet the forces shaping the emergence and persistence of de novo genes remain poorly understood. It has been proposed that many young de novo genes are slightly deleterious and therefore more likely to persist in genomic regions where genetic drift is strong, such as low-recombination regions associated with reduced effective population size (N_e). Here, we tested this hypothesis using the GENCODE catalog of 7,264 human small open reading frames (sORFs) encoding microproteins, of which more than 4,000 evolved de novo. We focused on the two most prevalent sORF categories containing substantial numbers of both de novo and non-de novo cases: long noncoding ORFs (lncORFs) and upstream ORFs (uORFs). After controlling for chromosome size, GC content, and gene density, we found that de novo lncORFs preferentially accumulate on long, low-recombination chromosomes, whereas de novo uORFs are relatively enriched on short, high-recombination chromosomes. Notably, chromosomes 4, 13, and 18, which combine low recombination with low gene density, harbor disproportionately high ratios of de novo to non-de novo lncORFs. Furthermore, we observed a similar enrichment of de novo lncORFs, but not uORFs, in the X-linked recombination desert recently described in mammals. Together, these findings support the hypothesis that many de novo lncORFs represent slightly deleterious novel genes that persist preferentially under conditions of reduced selective efficiency. Our results reveal how chromosome-scale variation in recombination and linked selection shapes the evolutionary trajectories of emergent coding sequences in the human genome.

396T Assembly and characterization of chromosome W in monarch butterfly (*Danaus plexippus*)

Martina Dalikova, James R Walters Ecology & Evolutionary Biology, University of Kansas

Degenerated sex chromosomes have a precedent of being notoriously difficult to assemble. Recent advances in long-read sequencing are catalyzing increases in the number and quality of W and Y assemblies, providing novel insights concerning the content and evolution of these elusive chromosomes. Here we report a novel genome assembly for the monarch butterfly (*Danaus plexippus*), focusing on the W chromosome. This species harbors a neo-Z chromosome, arising from the fusion of the ancestral Z with an autosome. Previous cytogenetic analyses indicated a similarly large and bipartite W chromosome, suggesting the possibility of a comparable neo-W, but much ambiguity remains concerning the sequence and history of the monarch W chromosome. We generated PacBio HiFi reads with Hi-C data from females to support *de novo* assemblies of maternal and paternal haplotypes using Trio binning. This produced chromosomal-level scaffolds for the Z and all autosomes. Approximately 14 Mbp of W-linked contigs from the maternal genome were identified based on male-to-female coverage and sex-specific k-mers. The W chromosome scaffold length was around 10 Mbp, thus leaving about one-third of this chromosome in unscaffolded contigs. We used this new assembly to investigate the gene and repeat content as well as population genetic diversity of the monarch W chromosome. The W is highly repetitive and contains very few protein coding genes, which mainly arose through retroposition or ectopic recombination from other chromosomes. Exons of these W-linked copies retain high sequence identity compared to their Z-linked and autosomal counterparts despite substantial divergence in introns and intergenic regions, suggesting strong stabilizing selection on protein coding regions for these W genes. At least some W genes appear to be expressed in female tissues, especially ovaries. The prevalent repetitive content of the W chromosome is formed by transposable elements (TEs) from the LINE and LTR retrotransposon groups. Surprisingly, the TEs on the W chromosome have strikingly lower divergence compared to the rest of the genome, which we suggest results from gene conversion arising from TEs occurring in tandem arrays. Population genetic analyses revealed nucleotide diversity is about ten times lower on the W chromosome compared to autosomes. Finally, despite this novel W assembly, strong evidence for or against a neo-W chromosome origin remains elusive.

397T Genome-wide selection on transposable elements in maize

Beibei Liu[1], Regina A Fairbanks[1], Candice N Hirsch[2], Manisha Munasinghe[2], Nate Pope[3], Jeffrey Ross-Ibarra[1] [1]University of California, Davis, [2]University of Minnesota, [3]University of Oregon

While most evolutionary research has focused on single nucleotide polymorphisms (SNPs), transposable elements (TEs) represent a major but understudied source of mutations that can influence organismal fitness. Previous studies on TEs often overlook the mechanisms and rates of transposition, rely on short-read sequencing that limits TE detection, or focus on small genomes such as Arabidopsis or Drosophila. In this study, we leveraged high-quality, long-read genome assemblies from 26 maize inbreds to investigate natural selection on TEs. We developed a novel and interpretable method, ΦSFS, which incorporates TE age and improves resolution for detecting selection. Using this approach, we identified key factors influencing selection on TEs: (1) the distance to the nearest gene, (2) the pre-insertion DNA methylation level at the insertion site, and (3) intrinsic TE characteristics, including copy number and expression level. This work represents the first application of long-read genome assemblies to study TE selection in a major crop species with a typical plant genome size. Our ΦSFS method offers a broadly applicable framework for detecting selection on TEs, and the factors uncovered provide new insights into the evolutionary dynamics and trade-offs between TEs and host genes.

398T Extensive genome evolution distinguishes maize within a stable tribe of grasses

Michelle Stitzer Cornell University

Over the last 20 million years, the Andropogoneae tribe of grasses has evolved to dominate 17% of global land area. Domestication of these grasses in the last 10,000 years has yielded our most productive crops, including maize, sugarcane, and sorghum. The majority of Andropogoneae species, including maize, show a history of polyploidy – a condition that offers the evolutionary advantage of multiple gene copies yet poses challenges to basic cellular processes, gene expression, and epigenetic regulation. To date, understanding the genomic consequences of polyploidy has been limited by the sparse sampling of groups of taxa with multiple polyploidy events. Here, we present 33 chromosome-scale genome assemblies from 27 species, including all diploid teosinte species and subspecies. These genomes capture 15 independent polyploid formation events, including the shared whole genome duplication between *Zea* and sister genus *Tripsacum*. In maize, the after-effects of polyploidy have been widely studied, showing reduced chromosome number, transposable element (TE) expansions, and biased fractionation of duplicate genes. While we observe these patterns within the genus *Zea*, 12 other polyploidy events deviate significantly. Those tetraploids and hexaploids retain 40 or 60 chromosomes, have only stochastic TE amplifications, and maintain nearly complete complements of duplicate genes. We hypothesize this lack of genomic response to polyploidy arises from differences in the evolutionary paths to re-establishing diploid genetic inheritance. In most Andropogoneae taxa, polyploidy provides multiple copies that buffer genetic load, whereas in maize and other paleopolyploid taxa, reduced chromosome complements and selective retention of duplicate genes can link together novel adaptive combinations. In total, these genomes provide a powerful backdrop to better understand maize diversity and the evolutionary context of maize genes and alleles.

399W Antimicrobial activity characterization of a novel anticancer peptide

Ahmed Ibrahim Molecular Biology, The American University in Cairo

Cancer is one of the most dangerous diseases in the world, especially with the increasing death rates over the years. Traditional treatment approaches used for cancer treatment, such as chemotherapy, radiotherapy, and surgery, are still not very effective in combating the disease, in addition to cancer cells acquiring resistance against some of these treatment approaches. A novel approach for treating cancer is anticancer peptides which target cancer cells specifically inhibiting their proliferation and migration while avoiding healthy cells. Anticancer peptides have higher efficiency and specificity compared to conventional treatment approaches and cancer cells are less likely to develop resistance against them. The objective of this research is the characterization of the antimicrobial activity of a novel anticancer peptide to assess its safety on the microbiological level as a potential cancer drug.

The peptide was obtained from metagenomics data retrieved from samples from the Red Sea and was found to possess anticancer activity. The antimicrobial activity analysis was done using two approaches, which are computational analysis of the peptide using bioinformatics tools and wet lab analysis. Viable cell count assay, minimum inhibitory concentration assay, disk diffusion assay, and kill time assay were conducted on gram-positive *Staphylococcus aureus* and gram-negative *Escherichia coli* bacterial strains, as representatives of the human microbiome, to assess the kinetics of the effect of different concentrations of the peptide on bacterial cells.

The computational analysis showed that the peptide has an α-helix secondary structure and that the most homologous sequence to the peptide is a homeodomain protein in *Pisaster ochraceus* (sea star). In addition, the antibacterial activity of the peptide predicted computationally was estimated to be between 20% and 30% while it had no detected antifungal nor antiviral activity. The results of the lab experiments showed that the IC50 concentration of the peptide, which is 100 µg/ml, had a minimal antimicrobial effect against both *S. aureus* and *E. coli,* which is not significant. The average inhibition rate of the peptide was 14.7% on *S. aureus* and 9% on *E. coli* with both not having a statistically significant difference compared to the controls.

According to the results obtained, it can be concluded that the novel anticancer peptide tested in this research can be considered safe on the microbiological level and will not disrupt the human microbiome. Thus, further investigation can be done on the peptide to use it for cancer treatment.

400W APOE ε4 risk allele accelerates episodic memory and mental status decline in women faster than men

Yanyan Wu, Tetine L Sentell Public Health Sciences, University of Hawaii at Manoa

Introduction

The *APOE ε4* allele is associated with an increased risk of dementia and cognitive decline, particularly with age. In addition to genetic factors, gender, race/ethnicity and lifestyle may also contribute to these outcomes. Studies have found that women with *APOE ε4* have a higher risk of dementia and faster decline in verbal memory decline compared to men, but these findings are not consistent across all research. Moreover, men and women have different trajectories in the two domains of cognition: episodic memory and mental status.

Method

This study analyzed 13 waves of biennial longitudinal data collected between 1996 and 2020 from the Health and Retirement Study, including 18,988 genotyped participants aged 50 or older. Episodic memory was assessed using both immediate and delayed word recall tasks, while mental status was evaluated through serial 7s and backward counting from 20, with higher scores indicating better performance on both tasks. Both measures were scaled to 0 to 100. Mixed-effects models with three-way interactions between cubic-splines of age, sex and *APOE ε4* were used to capture the trajectory of cognitive declines, adjusting for education, poverty level, race/ethnicity and cohort. Interpretation focuses on the predicted cognition between ages 60 and 80, to minimize survival bias and potential errors at the spline boundaries.

Results

Among non-*APOE ε4* carriers at age 60, women had an average of 56 on episodic memory, which was 6 points higher than men, whereas men had a higher average mental status score (83)--5 points higher than women. The rates of cognitive decline were similar between both men and women, leaving a constant gender-gap. Having two copies of APOE ε4 risk alleles was associated with a 1 to 3 points lower score on both episodic memory and mental status for both men and women at age 60.

APOE ε4 carriers experienced faster declines in the two measures than non-*APOE ε4* carriers ($p<0.001$), with a higher acceleration rate among women-carriers compared to men-carriers. From age 60 to 80, women's episodic memory declined by 11, while men declined by 6 points. Despite the faster decline, women non-carriers still had approximately the same episodic memory level as men at age 80.

Similarly, women's mental status declined by 12, while men declined by 9 points. Because women's mental status was already 5 points lower than men at age 60, the gender gap in mental status gap widened, with men having 9 points higher than women at age 80 (men 66 vs. women 57, $p<0.001$).

Conclusion

Our study showed that middle-aged women had higher episodic memory but lower mental status than men; but having APOE *ε4 risk alleles* was associated with faster decline for women than men later in life for both measures. The results emphasize the need for earlier lifestyle interventions to reduce cognitive impairment and dementia burden in women, particularly *APOE ε4* carriers.

401W Large-effect drivers of gene-expression variation across individuals

Alber Aqil[1], Omer Gokcumen[2], Naoki Masuda[1] [1]Department of Computational Medicine and Bioinformatics, University of Michigan, [2]Biological Sciences, State University of New York at Buffalo

Gene expression is typically shaped by a large number of small-effect genetic and environmental factors. This results in a unimodal distribution of expression levels across individuals. However, large-effect factors can create a bimodal distribution, wherein a gene is switched on (near the upper mode) in some individuals and switched off (near the lower mode) in others. Using genomic and transcriptomic data from 943 individuals across 27 tissues, we identify 473 genes exhibiting such bimodal expression in at least one tissue, i.e., bimodal genes. 32 of these bimodal genes show universal bimodal expression across all tissues, while the remainder show tissue-specific bimodal expression. Of the universally bimodal genes, 60% are driven by single-nucleotide variants, while the rest are due to polymorphic deletions and duplications. For such large-effect variants, the three genotypes result in only two expression-level modes across individuals, suggesting that these variants have non-additive effects on expression levels. Having characterized universally bimodal genes, we next asked what drives the much larger set of tissue-specific bimodal genes. We find that tissue-specific bimodal genes tend to be on or off concordantly within individuals, consistent with regulators (e.g., hormones) acting in *trans* and controlling the expression of multiple genes simultaneously. In the vagina, we identified seven estrogen-regulated bimodal genes whose concordant switching off results in vaginal atrophy in postmenopausal females. Overall, our findings reveal a substantial role for large-effect regulatory drivers in shaping gene expression, with implications for both disease, phenotypic diversity, and evolution.

402W Ancestry-specific Frequencies of Pharmacogenetic Variation in the All of Us biobank

Jason V Vu[1], Christophe Thomassin[2], Cole Shanks[1], Alexander Ioannidis[3] [1]Genomics Institute, UC Santa Cruz, [2]Biomedical Data Science, Stanford University, [3]Biomolecular Engineering, UC Santa Cruz

98% of the world's population carries an actionable pharmacogenomic (PGx) variant, that is, a variant which affects one's reaction to drugs. Understanding the distribution of these variants across different populations is crucial to improving personalized medicine. However, the frequencies of pharmacogenetic variation differ based on ancestral origin. To address this, we ran local ancestry inference using Gnomix on the All of Us biobank on over three hundred thousand samples to uncover detailed ancestry composition and admixture. We identify the allele frequencies of known clinically significant PGx variants. Critically, we find dramatic differences in frequencies when conditioning on local ancestry, including the loss-of-function TPMT *3C allele which has a 5% frequency in African ancestry and less than 0.6% in other ancestries, and the loss-of-function NUDT15 *2 allele that has a 7% frequency in Indigenous American ancestry compared to ~3% in East Asian. Individuals homozygous for TPMT *3C are at a high risk for bone marrow suppression when taking mercaptopurine, which is used for the treatment of acute leukemia, among other diseases. Individuals with one or two NUDT15 *2 alleles have an increased risk of developing leukopenia or neutropenia when treated with mercaptopurine as well. Our results highlight the critical value of incorporating local ancestry into PGx analysis. Despite Indigenous American ancestry having the highest allele frequency for NUDT15 *2, it is often overlooked compared to ancestries like East Asian which were better represented in previous biobanks. Investigating these variants and how they are distributed across populations is essential to improving precision medicine worldwide.

403W Unraveling the contribution of mutational rate variation to observed patterns of driver mutations in human cancer

Diljeet Kaur, Ziyue Gao Genetics, University of Pennsylvania

The occurrence of driver mutations, including loss-of-function (LoF) and gain-of-function events, varies significantly across human cancer types and among tumors within the same cancer type. While this variation is often attributed to differences in the strength of positive somatic selection, mutation occurrence also differs across cancers, reflecting the sequence-dependent and tissue-specific activity of mutational processes. Here, we tested whether variation in baseline mutation probability contributes to the observed differences in driver mutation frequencies across cancer genomes.

We compiled a set of 85 putative tumor suppressor genes and identified all possible high-confidence LoF single-nucleotide variants in each gene, including stop-gain and splice-disrupting mutations. LoF variants were highly concentrated within a limited number of mutation types, with 22 and 20 types accounting for 80% of all possible stop-gain and splice-disrupting events respectively. Stop-gain mutations were disproportionately driven by C>A and C>T substitutions, whereas splice-disrupting mutations showed no strong enrichment for any specific mutation type. We then predicted the expected LoF mutation rate of each tumor suppressor gene under each COSMIC mutational signature. Tumor suppressor genes exhibited substantial heterogeneity in their susceptibility to LoF mutations across mutational signatures, with SBS10a, SBS10d, and SBS36 consistently ranking among signatures most likely to generate stop-gain mutations. We also found that, for many tumor suppressor genes, certain LoF mutations are predicted to arise almost exclusively from specific mutational signatures relative to other signatures active in the same tissue. We termed these events 'signature-specific hotspot driver mutations' and performed an initial assessment of whether their occurrence correlates with the activity of the corresponding signature.

Together, our results show that most loss-of-function single-nucleotide mutations in tumor suppressor genes arise from a limited subset of mutation types, with C>A and C>T substitutions dominating stop-gain events. We further show that mutational signatures vary substantially in their likelihood of producing driver mutations, and we identify many putative signature-specific hotspot driver mutations.

404W Measuring genetic variation in water consumption after sleep deprivation in Drosophila melanogaster

Katie M Traeger[1], Allison John[1], Keira M Cerda[1], Nicholas Collins[2], Laura K Reed[2] [1]Biological Sciences, The University of Alabama, [2]The University of Alabama

Metabolic syndrome is a clinically measurable cluster of risk factors leading to the progression of severe cardiovascular and endocrine disorders. Development of the condition is exacerbated through poor sleep, dietary choices, exercise levels, and gut dysbiosis. The extensive interactions between genetic and environmental factors underlying metabolic disorders make it difficult to determine the specific mechanisms mapping genotype to phenotype. Our research focuses on studying the relationship between extreme sleep deprivation and water consumption in Drosophila melanogaster. Using flies from the Drosophila Genetic Reference Panel (DGRP), we mechanically sleep deprived three strains that require different amounts of sleep- short sleepers (<400 minutes per day), medium sleepers (400-600 minutes per day), and long sleepers (>600 minutes per day). We hypothesize that individuals with a greater sleep deficit will likely drink more water in order to help regulate their bodily functions. The sleep deprivation period is monitored over five days through the Trikinetics Drosophila Activity Monitor where the flies are vigorously shaken for five seconds by vortex, every minute during the 12-hour dark/night period. After sleep deprivation, we measure water consumption and daytime activity relative to controls, which show genetic variation in their phenotypic response. The results will help uncover the dual effects of dehydration and sleep deprivation on the individual and reveal how flies recover from such events. With the conservation of important homeostatic pathways for sleep and hydration cycles between humans and Drosophila, we hope that our research can bring more awareness on the impacts of dehydration coupled with poor sleep habits, especially in individuals affected by metabolic syndrome. Further research can bring about identification of genes involved in recovery after periods of stress related to sleep deprivation.

405W The Genomic Legacy of Neanderthal and Denisovan Ancestry Across 50,000 Years of Evolutionary History in East Asia

Sophie K. Joseph Department of Molecular and Cell Biology, University of California, Berkeley

Archaic gene flow from Neanderthals and Denisovans has played a critical role in shaping variation in human populations. It is associated with a range of functional effects in Eurasians, including susceptibility to chronic and infectious diseases (e.g. SARS-CoV-2, Type II diabetes). In particular, East Asians have a complex history of admixture with Neanderthals and Denisovans, reflected in a unique set of population characteristics not seen outside of East Asians. For example, East Asians harbor around 20% higher Neanderthal ancestry compared to most other global populations and at least two divergent pulses of gene flow from Denisovans. Thus, East Asians are a critical missing piece in understanding the evolutionary underpinnings of present-day human variation and health globally. Using ~30,000 whole genome sequences from present-day individuals, as well as ancient genomes from across Eurasia ranging in age from 45,000–1,500 years before present, this work comprehensively investigates the last 50,000 years of East Asian evolutionary history. Our findings include generating a high-resolution map of the genome-wide distributions of archaic haplotypes in East Asians, explaining the complex distribution of Neanderthal haplotypes into East Asians in relation to Basal Eurasian ancestry found in other parts of the Eurasian continent, and dating the timing of two divergent pulses of Denisovan admixture into East Asians (including one pulse that is divergent from the sequenced Denisovan genome).

406W Identifying Germline Mutation Hotspots and Their Determinants in Human Genomes

Nathan Cramer[1], Ziyue Gao[2], Priya Moorjani[1] [1]Molecular and Cell Biology, University of California, Berkeley, [2]Cell and Molecular Biology, University of Pennsylvania

Germline mutations are the ultimate source of genetic diversity within and across species. Understanding germline mutation rates and mechanisms is essential for studies of medical genetics (to interpret the incidence of de novo and heritable diseases) and evolutionary biology (to date demographic and adaptive events). Germline mutation rates vary across the human genome at multiple resolutions, ranging from adjacent base pairs to whole chromosomes. However, the biological mechanisms underlying mutation rate variation are yet to be fully characterized. To understand the determinants of local germline mutation rates, we analyzed single nucleotide polymorphisms from 76,156 whole-genome sequences of Europeans, East Asians, and Africans in the Genome Aggregation Database (gnomAD). We focus on putatively neutral mutations by removing conserved regions and exons. We apply a wavelet-based approach to identify mutation hotspots across varying genomic resolutions. Simulations demonstrate that our method outperforms the naive windowing approach for detecting hotspots across multiple scales (100 KB - 1 MB). Application to gnomAD shows that many factors correlate with local mutation rate variation and are enriched in mutational hotspots, including replication timing, recombination rate, germline methylation, and GC content. We recover known C > G maternal hotspots and identify novel associations between scale-specific mutation hotspots and mutational determinants, such as replication timing. We then apply this approach to study mutation rate differences across human populations. Notably, we identify population-specific enrichment of TCC>TTC mutational hotspots in Europeans compared to Africans. By associating DNA motifs and genomic features enriched in the hotspots, we uncover candidate trans and cis-acting factors associated with mutational modifiers in Europeans. Together, our method provides a novel framework for identifying mutation hotspots and the genomic features and biochemical processes impacting the mutation rate landscape in humans.

407T Cross-population discovery and lung cancer risk prediction using TOPMed imputation and mixed-effect models

Yafang Li, Xiangjun Xiao, Jinyoung Byun, Younghun Han, Christopher Amos University of New Mexico

Genome-wide association studies (GWAS) of lung cancer have identified numerous susceptibility loci in lung cancer, yet their discovery and translational utility remain constrained by limited representation of diverse ancestries, reduced power for rare and low-frequency variants, and reliance on ancestry-specific analytic frameworks. Here, we imputed genotype data from the Integral lung cancer consortium using the Trans-Omics for Precision Medicine (TOPMed) reference panel and performed association analyses with the SAIGE mixed-effect model, enabling joint analysis across populations without explicit ancestry inference. We identified six novel lung cancer susceptibility loci, including three from multi-ancestry analyses and three from European-only analyses, of which two involve low-frequency variants (minor allele frequency <0.05). All six loci were independently replicated in large external cohorts from FinnGen, UK Biobank, and the Million Veteran Program, comprising 23,058 cases and 718,391 controls. In addition, we demonstrate that polygenic risk scores derived from mixed-ancestry summary statistics achieve predictive performance comparable to ancestry-specific PRS constructed using PRS-CSx. Together, our findings highlight the power of combining diverse imputation resources with mixed-effect GWAS models to uncover novel genetic risk factors and improve the portability of genetic risk prediction for lung cancer across populations.

408T Understanding the Global Applicability of Cancer GWAS with Population Genetics

Aiko Abo Dominguez[1], Sydney Bruce[1], Kate Olsen[1], Melody Yazdi[1], Rebecca Aviles Barahona[1], Christopher Kim[1], Kailin Liu[1], Niamh McManus[2], Talha Rafique[1], Yafei Wen[1], Ekene Okolo[1], William Baharian[1], Christopher Huh[1], Lan Trieu[1], Karen Lee[1], Preston Cheng[1], Misha Faruki[1], Lucia Zhang[1], Jazlyn Mooney[1] [1]University of Southern California, [2]Trinity College Dublin

Genome-wide association studies (GWAS) have identified numerous cancer risk loci, yet the extent to which associated variants are shared across global populations remains under explored. In this study, we analyze GWAS results for multiple cancer types (colorectal, ovarian, lung, and breast cancer) and examine patterns of allele frequency variation and population differentiation across worldwide populations. Cancer-associated single nucleotide polymorphisms (SNPs) were obtained from the GWAS Catalog and intersected with phased genotype data from the 1000 Genomes Project. Across cancer types, allele frequency distributions of associated variants were compared to randomly sampled non-associated variants, site frequency spectra (SFS) were summarized, and population structure was quantified using pairwise F_{ST}.

Cancer-associated variants were observed at more common frequencies in the populations in which the studies were conducted, reflecting the power and design of GWAS. Most associated SNPs exhibited moderate population differentiation, indicating that cancer risk loci are likely broadly shared across ancestries. However, notable heterogeneity emerged across cancer types, particularly for lung and breast cancer-associated variants, which showed elevated F_{ST} values relative to random SNPs. These patterns likely reflect a combination of demographic history, population structure, and GWAS study design.

Overall, our analysis demonstrates that while many cancer GWAS signals capture shared standing genetic variation, demographic history has likely driven to differences in allele frequencies across populations. These findings underscore the importance of expanding GWAS representation to improve the global applicability and equity of cancer genomics research.

409T MicroRNA-29 family: Clinically useful novel biomarkers in type 2 diabetes mellitus

Rifat Z. Ahmed[1], Jahangul Khakwani[2], Nadia Khan[1], Uzair Abbas[3] [1]Genetics, University of Karachi, [2]Liaquat National Hospital and Medical College, [3]Physiology, Dow University of Health Sciences

Precise molecular mechanism underlying Type 2 diabetes mellitus (T2DM) is still not clearly understood. MicroRNAs (miRNAs) offer a promising avenue providing insights into the intricate regulatory networks that influence the development of the disease. The miRNA-29 family, consisting of three mature members miRNA-29a, miRNA-29b and miRNA-29c, is among the most highly expressed miRNAs in the pancreas and liver of both mice and humans. It has been identified as a candidate regulator of glucose metabolism. This case control study was conducted to investigate the expression of miRNA-29 family in T2DM and explore its association with glycemic control and demographic, clinical, and lifestyle factors.

A total of 120 participants including 60 T2DM patients (HbA1c levels ≥ 6.50) and 60 non-diabetic controls (HbA1c levels < 5.70) were enrolled in this study. Demographic, clinical and lifestyle data was gathered for variables gender, age, BMI, exercise, smoking, and family history of diabetes. Expression levels of miRNA-29a, miRNA-29b and miRNA-29c were analyzed using RT-qPCR. Their fold changes were calculated and statistical analyses were performed to find out any differences in their expression level leading to association between diabetic and non-diabetic groups and among four subgroups on the basis of diabetes status with gender, age, BMI, exercise, and smoking. To find out the statistical significance at p value <0.05, Mann-Whitney U test was performed for comparison between two groups and Kruskal-Wallis test was performed for comparison among four subgroups followed by post hoc Mann-Whitney U test to determine which specific subgroups differed.

Demographic, clinical, and life style data of study participants revealed that; 60% were male and 40% were female, 36.7% were < 45 years and 63.3% were ≥ 45 years, 52.8% had family history of diabetes and 49.2% did not, 46.7% had Body Mass Index (kg/m^2) < 30 and 53.3% had ≥ 30, 28.3% used to exercise and 71.7% did not, 9.2% were smokers and 90.8% were non-smokers. Among the three miRNAs, statistically significant higher median expression level (2.46 (0.79 to 4.37)) of only miRNA-29c was observed in diabetic individuals compared to non-diabetic individuals (-0.84 (-2.81 to 2.76)). Among both genders, females showed a statistically significant higher median log fold change of only miRNA-29a (0.75 (-0.94 to 2.80)) compared to males (-0.40 (-2.36 to 1.77)). However miRNA-29c significantly differed across subgroups of age-diabetes, gender-diabetes, BMI-diabetes, smoking-diabetes, and exercise-diabetes.

The finding suggests that miRNA-29c may serve as a potential biomarker for T2DM, indicating its involvement in the disease molecular mechanism. Studies with larger sample sizes and functional assays are needed to validate these findings and understand the regulatory roles of miRNA-29 family members in glycemic control and diabetes pathogenesis.

410T Genetics of Stroke in Sickle Cell Disease Patients

Anvita Kulshrestha, Melanie E. Garrett, Marilyn J. Telen, Allison E. Ashley-Koch Duke University

Sickle cell disease (SCD) is caused by a single genetic mutation in the beta hemoglobin gene. Despite the same underlying causal mutation, patients exhibit highly heterogeneous clinical outcomes. Both hemorrhagic & ischemic strokes are common and serious complications observed among individuals with SCD. Transcranial Doppler (in children) and magnetic resonance angiography (across all ages) can detect high risk of stroke, yet they have inadequate specificity. Prophylactic chronic transfusion therapy is the main preventative measure, and identifying additional risk factors—such as genetic variants—could help target this treatment more effectively towards those truly at high risk.

To understand the genetic underpinnings of SCD-associated stroke, we performed a genome-wide association study (GWAS) with 416 adult SCD patients (with homozygous sickle or sickle beta-0 thalassemia genotypes) from the Outcome Modifying Genes in SCD (OMG-SCD) cohort. Patients with at least 1 episode of stroke in their lifetime were categorized as cases (n=70) while patients with no history of stroke were categorized as controls (n=346). We performed the GWAS using R package SAIGE, accounting for age, sex, treatment (hydroxyurea) status and 2 genomic principal components as covariates.

After linkage disequilibrium clumping, the top SNP associated with stroke was rs6971441 on chromosome 7, with a p-value of 3.46×10^{-8} (q = 0.01). This intronic SNP lies within the DGKB gene. DGKB is highly expressed in the brain and has previously been associated with neurologic disorders like developmental and epileptic encephalopathy 2. Overall, 212 variants were nominally associated with stroke ($p < 5 \times 10$□□). Among the genes mapped to the top 10 variants, many have been associated with neuronal development and degeneration.

We are currently deploying post-GWAS methods for functional annotation and pathway enrichment analysis. Our goal is to unravel the biological mechanisms and pathways involved in SCD-associated clinical outcomes and provide insights into the molecular profiles specific to severe SCD.

411T Gene regulation changes associated with a chromosome 15q13.3 deletion

Adeena Rahman, Sean P McGinty, Gulhan Kaya, Megan Y Dennis University of California Davis

Structural variants (SVs) are large genomic alterations that include deletions, duplications, inversions, and translocations. By disrupting higher-order DNA structures such as topologically-associating domains (TADs), SVs can reshape entire regions of the genome by rewiring cis-regulatory elements and affecting local gene-expression programs. Segmental duplications comprise a substantial fraction of the human genome, creating genomic hotspots which are prone to reorganization, structural variation, and mutations. Chromosome 15q13.3, a genomic hotspot, is an unstable region containing extensive repeats from the GOLGA8 gene family. These duplications promote evolutionary rearrangements and recurrent microdeletions that are strongly associated with neurodevelopmental conditions including schizophrenia, autism spectrum disorder, and epilepsy. Previous work in chr15q13.3 has identified several different structural haplotypes including multiple inversion-based haplotypes spanning from 12 Kbp to 2 Mbp. Leveraging genome assemblies from the Human Pangenome Reference Consortium (HPRC), we identified in approximately 14% of the dataset a novel ~470-kbp deletion removing 6 protein coding genes and segregating in human populations. RNA-sequencing analysis of lymphoblastoid cell lines revealed differential expression of multiple genes outside of the deletion, including *TJP1*, *NRN1*, and *RN7SL1*. Assessing TAD organization for a non-deletion-containing haplotype, we predict the deletion to remove an entire TAD and promote fusion of adjacent TADs, providing a possible explanation for the observed gene-expression changes. To better understand how three-dimensional genome organization between haplotypes with and without the deletion might dictate these expression differences, we are currently generating long-read chromosome conformation capture (CiFi) data. These analyses aim to determine whether deletion-associated TAD disruptions lead to altered cis-regulatory interactions and misregulation of genes flanking the affected region. Altogether, this provides insight into how SVs alter chromatin architecture and transcriptional outcomes at chromosome 15q13.3. More broadly, this work understands how genomic rearrangements are linked to disease-associated molecular phenotypes.

412T Translating Sequence to Structure: Defining the Biophysical Fitness Landscape of the HDAC6 Catalytic Domain

Paul Fong, Fan Yuan Biomedical Engineering, Duke University

A central challenge in quantitative genetics is connecting sequence-level evolutionary signals to the biophysical phenotypes upon which selection actually acts. Protein language models such as ESM-2 learn rich representations of sequence fitness from evolutionary data, yet they are blind to the structural and thermodynamic consequences of mutations. We present an automated computational framework that makes this mechanistic layer explicit, translating ESM-2 sequence predictions into quantitative biophysical phenotypes via high-throughput Molecular Dynamics (MD) simulations. We applied this pipeline to the histone deacetylase 6 (HDAC6) catalytic domain, an enzyme implicated in cancer and neurodegeneration. We selected four variants (H499N, L519V, I510P, N600A) that ESM-2 scores as high-likelihood substitutions. To empirically test this prediction, we simulated each variant bound to the selective inhibitor Tubastatin A, using inhibitor displacement as a precise functional probe for active-site integrity. We quantified backbone fluctuations (RMSF), non-bonded interaction energies, and structural deviations (RMSD) over extended production runs. RMSF profiles confirmed that all four variants maintain global structural integrity, consistent with ESM-2's assessment that these substitutions are broadly tolerable. RMSD and interaction energy analyses split the variants into two mechanistic classes: H499N and I510P showed ejection of the inhibitor into solvent (ligand RMSD > 10 nm, ~0.0 kJ/mol), defining these positions as near-invariant structural anchors under strong purifying selection. L519V and N600A produced aberrant association — the inhibitor departed the native pocket but established highly favorable non-native interactions elsewhere on the surface (−150 to −200 kJ/mol). These results demonstrate a clear dissociation between sequence-space fitness, as captured by ESM-2, and biophysical fitness, as revealed by MD simulation: variants that appear evolutionarily permissible at the sequence level can be functionally catastrophic at the structural level. This gap is precisely the layer of the fitness landscape that population genetic models currently lack. Our framework offers a scalable approach for bridging this divide — connecting allele-level evolutionary signals to mechanistic, spectrum-based phenotypes, and providing quantitative genetics with a physically grounded basis for modeling how selection shapes protein-coding sequence.

413W Engaging Undergraduates in Evolutionary Genomics through Biocontrol Agents at a Rural PUI: A Scaffolded Research Pipeline

Amanda Stahlke[1], Zeynep Ozsoy[1], Dan W Bean[2] [1]Biological Sciences, Colorado Mesa University, [2]Palisade Insectary, Colorado Department of Agriculture

Classical biological control agents represent natural experiments through which evolutionary hypotheses can be observed and tested in the wild, and modern sequencing strategies now allow us to interrogate these processes at the genomic scale. Regional, primarily undergraduate institutions (PUIs), particularly those in rural communities, are uniquely positioned to link critical ecosystem processes to the students who interact with them directly. Undergraduate research at these institutions can generate meaningful scientific contributions while providing rich training experiences for students entering the workforce or pursuing graduate study.

This poster presents a scaffolded undergraduate research program anchored in evolutionary genomics of tamarisk beetles (*Diorhabda* spp.), scaling from large-enrollment courses to independent research at Colorado Mesa University. Each tier feeds into and reinforces the others, creating a sustainable framework. The modular structure is portable: individual tiers can be adopted independently, and the framework transfers readily to other biocontrol or invasive insect systems.

Tier 1 engages BIOL301 Genetics Lab students in Sanger sequencing mitochondrial CO1, contributing to a growing lineage distribution database with discoveries including parasitoid fly associations overlapping species distributions. Institutions with access to tamarisk-invaded riparian corridors or other biocontrol release sites are well-positioned to contribute specimens and sequence data, expanding geographic coverage. Tier 2 builds on this through structured eukaryotic gene annotation in collaboration with the Genomics Education Partnership, with students investigating the evolution of diapause-related genes in the four tamarisk beetle species, *D. carinulata*, *D. sublineata*, *D. elongata*, and *D. carinata*. Tier 3, piloting Fall 2026, will use Oxford Nanopore MinION sequencing to validate novel gene annotations and assess isoform diversity under experimental conditions. Tier 4 comprises independent research projects, including an investigation of evolutionary processes underlying rapid latitudinal range expansion.

The framework is broadly applicable to any biocontrol or invasive insect undergoing evolution, making it a realistic entry point for PUIs across regions and taxa. For example, yellow starthistle root crown weevils (*Ceratapion basicorne*) and eriophyid mites on whitetop and Russin olive (*Aceria* spp.) serve as additional systems. Interinstitutional collaborations, including student site visits, are integral to sustaining and expanding the program's research capacity.

414T Efficient Repetition in Molecular Evolution: Integrating Frequent Practice Across Discussion Sections, Online Homework, and Problem-Based Review Sessions

Hayley A Goss[1], Jackie E Shay[2] [1]Marine Science Institute, University of California, Santa Barbara, [2]Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara

Molecular evolution is increasingly common in upper-division undergraduate curricula, yet remains underrepresented in the education literature, resulting in limited guidance on effective instructional strategies (Ziadie & Andrews, 2018; Forsythe & Hsu, 2023). While repetition and practice improve learning outcomes across disciplines, the perceived effectiveness of repetition and students' preferred repetition strategies have not yet been investigated in the context of learning molecular evolution. In this study, we investigated the role of repeated practice in an upper-division molecular evolution course using pre- and post-course surveys, mid-quarter reviews, topic-specific formative assessment results, and weekly student reflections. Students engaged with course content through lectures, question-based knowledge checks, applied practice problems, online auto-graded homework, and question-based review sessions. On average, students were exposed to each topic ~ 4-6 times prior to formative assessments.

Student-reported comfort increased significantly across all 10 learning topics between pre- and post-course surveys. Although students identified coursework requiring multi-step application and interpretation, such as Tajima's D and dN/dS interpretations, as most challenging, they performed well on these topics in formative assessments. Students credited their learning to applied practice problems, online homework with feedback, and application-based review sessions. Students identified applied practice problem sets as the most helpful resource. They routinely requested additional problem sets and reported a belief that they could overcome challenges with the material through additional practice. Despite the course incorporating extensive repetition-based assignments, students did not report a greater perceived workload than in other courses. Overall, our findings indicate that repetition through targeted applied practice sets, online homework, and question-based review sessions may support students' learning in upper-division molecular evolution courses.

415W Community coalescence reveals strong selection and coexistence within species in complex microbial communities

Sophie Jean Walton[1], Qing Xu[2], Hannah R Gellert[3], Richa Sharma[3], Chih-Fu Yeh[3], Jonas Cremer[3], Katherine S Xue[4], Dmitri A Petrov[3], Benjamin H Good[2] [1]Biophysics, Stanford University, [2]Applied Physics, Stanford University, [3]Biology, Stanford University, [4]University of California, Irvine

Large microbial communities, such as the human gut microbiome, harbor extensive genetic variation within each of their component species, with conspecific strains in different local communities varying at thousands of genomic loci. However, the nature of selection on this intra-species variation remains poorly understood in species-rich community settings. To address this, we performed pairwise community coalescence experiments of *in vitro* gut microbial communities derived from different human donors, revealing the emergent fitness differences between conspecific strains as they competed within larger communities. Specifically, we used shotgun metagenomic sequencing to track the frequency of conspecific strains from each host for over 50 generations across many species in parallel, including prevalent human commensals such as *Bacteroides uniformis* and *Parabacteroides distasonis*. We found that most pairs of strains experienced strong and context-dependent selection, even when their parent communities were originally selected in the same nutrient environment. However, these fitness differences typically attenuated over time due to biotic interactions within the community, leading to extended coexistence within many species, and competitive exclusion in others. These results support the view that conspecific strains can fulfill distinct ecological roles when competing within a diverse community, even when their genomic diversity exhibits the hallmarks of a single biological species.

416W Uniform bacterial genetic diversity along the gut

Michael Wasney[1], Leah Briscoe[2], Richard Wolff[3], Hans Ghezzi[4], Carolina Tropini[4], Nandita Garud[5] [1]Human Genetics, UCLA, [2]Bayer, [3]UCSD, [4]UBC, [5]Department of Ecology & Evolution, UCLA

Environmental gradients throughout the digestive tract shape spatial variation in the composition and abundance of bacterial species along the gut. However, much less is known about how genetic diversity *within* bacterial species is distributed along the gut. Understanding this distribution is important because bacterial genetic variants confer traits that influence both microbiome function and host physiology, including local inflammation and nutrient metabolism. Thus, to understand how the microbiome functions at a mechanistic level, it is essential to understand how genetic diversity is organized along the gut. In this study, we profiled genetic diversity of approximately 30 common gut commensal bacteria in five regions along the gut lumen in germ-free mice colonized with the same healthy human stool sample. Although species composition varied considerably along the gut, genetic diversity within species was substantially more uniform. Driving this uniformity were similar strain frequencies along the gut, implying that multiple, genetically divergent strains of the same species can coexist within a host without spatially segregating. Additionally, the approximately 60 unique evolutionary adaptations arising within mice tended to sweep throughout the gut, showing little gut region specificity. We then analyzed metagenomic samples collected along the guts of conventional mice and healthy humans and found similar dynamics with their natural microbiomes, suggesting that uniform bacterial genetic diversity may be common to multiple host species. Together, our findings demonstrate that uniform spatial distribution of genetic diversity along the gastrointestinal tract is a robust feature of mammalian gut ecosystems.

417W Using strain-resolved metagenomics to track microbiome transmission in a wild primate

Reena Debray[1], Caroline Dickson[2], Shasta E. Webb[3], Pamela Ferretti[4], Jack Gilbert[5], Susan Alberts[6], Ran Blekhman[4], Elizabeth Archie[3], Jenny Tung[7] [1]Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, [2]Biological Sciences, University of Notre Dame, [3]University of Notre Dame, [4]University of Chicago, [5]University of California, San Diego, [6]Duke University, [7]Max Planck Institute for Evolutionary Anthropology

Animal-microbiome associations are ubiquitous, yet the mechanisms by which they arise, specialize, and persist through evolutionary time remain relatively unknown. Animal social behaviors have been proposed as a key mechanism maintaining the evolution of microbial specialization by enabling reliable transmission between individuals. To test this hypothesis, we integrated strain-resolved metagenomics with detailed behavioral and demographic data from an intensively studied population of wild baboons in Kenya. We generated shotgun metagenomic sequences from longitudinally collected fecal samples and tracked bacterial strains as they moved among hosts over time. We identified social transmission by using fine-grained grooming interactions to resolve strain acquisitions that followed the social network. Compared to species with transmission patterns following shared environments, socially transmitted species persisted for longer durations within baboons, but were less likely to be found in other primate species, suggesting a history of specialization and coevolution with their hosts. Further, the genomes of socially transmitted species were depleted for genes involved in motility, sporulation, and stress resistance – all traits thought to promote survival outside of hosts. Taken together, our results provide the strongest evidence to date that social structures shape the transmission and the selective landscape of the microbiome.

418W How does environmental difference, distance, and dimensionality shape Pareto fronts?

Elisa Visher[1], Gavin Sherlock[2], Dmitri Petrov[3] [1]Biology / Genetics, Stanford University, [2]Genetics, Stanford University, [3]Biology, Stanford University

How is it that trade-offs between environments are so idiosyncratically measured and seem so escapable, but are so necessary for describing any number of ecological and evolutionary dynamics? Here, we explore whether trade-offs can be better understood as probabilistic systems by extending concepts such as joint distributions of fitness effects (jDFEs) and Pareto fronts to clarify when trade-offs should arise and how they govern evolution in heterogeneous environments.

We experimentally a evolve barcoded *Saccharomyces cerevisiae* library in 1) single media environments spanning three levels of difference (SC-Ura 1.5% fructose; SC-Ura 1.5% galactose; SC-Ura 1.5% glucose + 0.5M NaCl) from a focal condition (SC-Ura + 1.5% glucose) and (2) spatially heterogeneous environments composed of two to four combinations of these media. We found that the rates of evolution in the four individual environments differed and were correlated with environmental difference from the focal condition; populations in more different environments evolved faster, consistent with greater distance from their fitness peaks.

Using these single-environment rates, we calculated expected rates of evolution in heterogeneous environments. We found that Increased environmental dimensionality led to significantly lower adaptation rates than predicted from the average of component environments, indicating a cost of generalism. We are now measuring the fitness of adaptive genotypes across all media to test whether the shape of the jDFE depends on environmental difference and/or their distance from their fitness peak. We ask whether more different environments exhibit steeper Pareto fronts, and whether populations closer to their fitness peak find more costly adaptations (accelerating costs pleiotropy). Finally, we evaluate whether evolutionary dynamics in heterogeneous environments can be predicted from jDFEs, with generalists favored under shallow trade-offs and specialists favored under steeper constraints.

419W Microbial single-cell RNA sequencing data reveal opposing transcriptional relationships at population and single-cell scales

Leandra Brettner, Kerry Geiler-Samerotte Biodesign Center for Mechanisms of Evolution, Arizona State University

Population-level measurements have long shaped our understanding of microbial physiology, underpinning widely accepted principles linking ribosome abundance, growth-associated transcription, and cellular growth rate. These relationships are often interpreted as reflecting fundamental regulatory strategies of individual cells. However, microbial populations are inherently heterogeneous, and the extent to which population-derived principles accurately describe single-cell behavior is coming into question. Here, we compared population-level and single-cell transcriptomes in *Saccharomyces cerevisiae* and *Bacillus subtilis* to examine how growth and stress regulation are organized across biological scales. At the population level, we observe the expected positive coupling between growth-associated transcription, ribosomal RNA abundance, and growth rate. In contrast, single-cell analyses reveal a markedly different picture. Ribosome levels vary widely among individual cells within the same population, population growth rate explains little single-cell variance in rRNA abundance, and growth-associated gene expression is consistently anticorrelated with rRNA within samples. Similar discrepancies between population and single-cell relationships are observed for stress-associated programs and transposable element expression. Single-cell transcriptomes further reveal persistent physiological substructure within exponentially growing cultures, violating the classical assumption of balanced growth: uniform biosynthetic allocation. Together, these results demonstrate that population-level scaling relationships can mask opposing and heterogeneous single-cell behaviors, limiting their interpretability for microbial physiology and evolution. Explicitly accounting for cellular heterogeneity is therefore essential for linking genomic regulation to microbial growth, adaptation, and evolutionary dynamics.

420W Capture, characterization, and experimental evolution of barcoded mobile genetic elements

Milo Johnson[1], Darian Doakes[2], Thien Vu[2], Adam Deutschbauer[3], Britt Koskella[2] [1]Integrative Biology, UC Berkeley, [2]UC Berkeley, [3]Lawrence Berkeley National Lab

"Great fleas have little fleas upon their backs to bite 'em,And little fleas have lesser fleas, and so ad infinitum" - Augustus De Morgan

In this study, we seek to expand our basic understanding of some of the smallest fleas we know: mobile genetic elements (MGEs). Leveraging a growing collection of over 150 randomly barcoded transposon (RB-TnSeq) libraries created in a wide variety of bacterial strains in Adam Deutschbauer's lab over the last 15 years, we screen for MGEs that carry the transposon and thus are traceable outside of host cells. RB-TnSeq libraries are traditionally used as collections of knockouts that can help functionally profile genomes across different selective environments, but they can also be used as collections of potentially transferable DNA, which can be isolated through conjugation with model recipients. We use bulk barcode sequencing of transconjugants to identify active mobile elements and determine how knockouts of each of their genes affect their ability to mobilize. Because each MGE is tagged with an antibiotic resistance cassette and associated with a set of unique barcodes and gene knockouts, we can pool together elements and use bulk assays to measure host range and the relative efficiency of transfer or integration of each element in each host, as well as the genetic determinants of host range.

Over the last 5 months, we have screened 80 RB-TnSeq libraries for conjugative elements capable of transfer into Escherichia coli, with 20 more libraries planned for early in 2026. Together, these 100 libraries span 3 phyla and 21 families and contain 58 complete conjugation systems. We have already confirmed transfer of 16 unique conjugative plasmids, integrative and conjugative elements, and, in one case, a temperate phage. Apart from one model conjugative element included as a positive control, none of these elements have been studied before.

This spring, we will experimentally evolve isolated elements by forcing them to be transferred back-and-forth between two E. coli strains. We will use this system to test classical experimental evolution questions in mobile elements: How much parallelism do we observe? Is coexistence common? Do less fit elements evolve faster? We will also test how selection depends on the balance of horizontal and vertical transmission by varying the frequency of conjugation and measuring changes in both conjugative efficiency and the fitness effect of the element in the host.

421W Investigating the evolution of fungus *Fusarium oxysporum* (Fo47) under different abiotic stressors

Madison Newman, Li-Jun Ma Biochemistry and Molecular Biology, University of Massachusetts, Amherst

Fungi may live many different lifestyles, as a pathogen, parasite, endophyte/symbiont, and decomposer; and they interact intimately with host plants. Climate change may cause a negative impact on plant-fungal relationships, especially for mutualistic interactions. In the *Fusarium oxysporum* species complex (FOSC), there are many endophytic strains which are beneficial to host plants, but how these strains will evolve under climate change-induced conditions is underexplored. My research goal for this project is to investigate how an endophytic strain of *F. oxysporum*, namely Fo47, mutates under different abiotic stress conditions: pH, salinity, and temperature induced stress; and the impact it has on its interaction with the host plants. For this, I will perform an evolutionary experiment where I grow fungal spores on different media under these different abiotic stress conditions. Thereafter, I will sequence the evolved strains and compare them to the ancestral reference strain to determine how the *F. oxysporum* endophyte, Fo47, evolved under these abiotic pressures. By understanding how endophytes evolve under these conditions we can learn how climate change will impact plant-fungal interactions, and how we might mitigate some of its undesirable effects.

422T Mapping the landscape of viral-immune coevolutionary dynamics

Tatiana Ruiz-Bedoya[1], Michael Lässig[2], Michael M Desai[1] [1]Harvard University, [2]University of Cologne

Viruses and the immune system have coevolved for millennia, and these evolutionary dynamics dictate which variants will emerge, spread, and escape our vaccines. A major challenge in predicting such evolutionary outcomes is that the genotype space viruses explore is too large to be sampled comprehensively, and the rules constraining how they navigate it differ from those in traditional evolutionary models. In Influenza, over a billion mutations can arise in one host during a single viral generation, and even single substitutions can substantially drive immune escape or zoonotic emergence. Efforts deployed annually to evaluate protection by the current vaccine can only explore small parts of these genotype-phenotype maps and often fail to predict which viral variants will escape immunity. One way to understand these dynamics is to follow how evolution has sampled this massive sequence space. We can trace the origin, rise, and fall of mutations throughout the virus's evolutionary history, building the missing link between historical patterns and the *evolutionary potential* of circulating strains. To do this, we aim to sample the phylogenetic history of Influenza, introduce mutations of interest, synthesize thousands of these custom DNA variants, and test them for phenotypes relevant to viral fitness. We can then build high-resolution genotype-to-phenotype maps that represent paths taken and yet-to-be-taken by the virus and build predictive models of future evolution for the major Influenza antigenic protein hemagglutinin (HA) in both seasonal and avian flu. This task has been unfeasible until now because synthesizing full-length custom-designed genes at scale is prohibitively costly, and higher-throughput mutagenesis tends to add random rather than custom mutations. To address this, we have built a new DNA synthesis and assembly pipeline that computationally designs and experimentally assembles libraries of thousands of HA variants, with 98% sequence accuracy by optimizing homology regions and using selective amplification. We have designed and assembled HA libraries of more than 5,000 sequences across avian and seasonal flu. With this, we aim to understand 1) patterns of vaccine/immune escape of seasonal flu, and 2) mutational patterns leading to zoonosis in avian flu. Using recently developed sequencing-based immunological and receptor-entry assays in safe non-replicative (BSL-2) systems, we aim to quantify binding, neutralization, and host-cell entry across these libraries and build genotype-to-phenotype maps that recover historical mutational trajectories and highlight plausible future routes of Influenza evolution to inform the design of longer-lived vaccines.

423T Pangenomic analysis of *Streptococcus pyogenes* reveals accessory genome diversification and the evolution of cluster-specific virulence

Evan Ho, Omar Cornejo Ecology and Evolutionary Biology, University of California, Santa Cruz

Streptococcus pyogenes infections contribute to a wide variety of diseases ranging from mild to life-threatening, yet determinants of disease severity remain poorly understood. We conducted a pangenomic analysis of 1017 publicly available *S. pyogenes* isolates encompassing 5995 genes (1364 core, 4631 accessory). Phylogenetic analysis and genetic clustering revealed substantial genomic diversity and recombination patterns within the species. We identified 250 accessory genes under positive selection, suggesting ongoing diversification. Functional enrichment analysis showed virulence-associated pathways were unevenly distributed across genetic clusters. These findings challenge the view of *S. pyogenes* as a monolithic pathogen, revealing instead that accessory genome variation drives the emergence of specialized groups with distinct pathogenic profiles and varying disease capacities.

424T Comparative Gut Viromes of Wild Lowland (*Lagothrix lagotricha tchudii)* and Yellow-tailed Woolly Monkeys (*Lagothrix flavicauda*)

Reese Hotten-Somers[1], Sofia M Weaver[2], Elvis Charpentier Uraco[3], Melissa Zarate[1], Evelyn Pain[2,4], Fanny Cornejo[4], Miriam Portocarrero Lapiz[5], Sam Shanee[6,7], Christopher Schmitt[1,2] [1]Anthropology, Boston University, [2]Biology, Boston University, [3]Equipo Primatológico del Perú, [4]Yunkawasi, [5]Asociación de Conservación Oso Dorado Hierba Buena / Allpayacu, [6]Neotropical Primate Conservation, [7]Asociación Neotropical Primate Conservation Perú

Virome metagenomic analysis can not only serve as a tool to understand the enteric host viruses but it can also provide critical insight into individual health across a range of ecological conditions. Peruvian woolly monkeys (*Lagothrix lagotricha tschudii*) and the yellow-tailed woolly monkey (*L. flavicauda*) are endangered, closely-related platyrrhine primates endemic to Perú, with the latter exclusively inhabiting colder Andean cloud forests (1800 – 2900 masl) and the former living in low-to-moderate elevation tropical forests (450 – 2600 masl). Research into primate viromes, particularly for platyrrhines, is severely limited and the endemic viruses of woolly monkeys have never been studied. We collected fecal samples from 73 *L. flavicauda* in Región Amazonas and 27 *L. lagotricha tschudii* individuals in Regiones Pasco (high elevation) and Ucayali (low elevation), Perú. We extracted viral RNA *in situ* at the Wildlife Conservation Lab at Los Amigos Biological Station, sequenced viral RNA using the Oxford Nanopore platform, and used the CZID pipeline for taxonomic assignment to assess virome compositions. We found similar alpha diversity in the viromes of both taxa as assessed by taxon richness (2.471 ± 1.586 vs 2.582 ± 1.512) and Shannon index (0.669 ± 0.573 vs 0.652 ± 0.495). However, there were significant differences in beta diversity and differential enrichment in key viral taxa, including increased relative abundance of *Salmonella phage 7t3* and *Hachidavirus* in *L. lagotricha tschudii*, and *Burkholderia phage BcepB1A* and *Flyfo siphovirus Tbat2_3* in *L. flavicauda.* PERMANOVA results suggest that the viromes of the two species are not significantly different (pseudo-F = 1.80, p = 0.096), although there is significantly increased heterogenous variation among individual *L. lagotricha tschudii* virome compositions compared to those among *L. flavicauda* (centroid 0.357 vs 0.238; pseudo-F = 3.49, p = 0.029), perhaps reflecting their respective ecological variation. Surprisingly, *Plasmodium* was detected in 27 individuals among both taxa suggesting woolly monkeys as a potential malarial reservoir, even though most sampling was done outside known malarial ranges. Our results suggest that fecal virome composition may respond to differing ecologies and may also be used to identify potential pathogens. These findings are critical for understanding zoonotic disease dynamics and informing conservation efforts for these endangered primates.

425T Adaptations spreading across human gut microbiomes arise from complex multisite adaptive architectures

Peter J Laurin[1], Nandita Garud[1,2] [1]Ecology and Evolutionary Biology, University of California, Los Angeles, [2]Department of Human Genetics, University of California, Los Angeles

Adaptation is pervasive in human gut microbiomes, giving rise to fundamental traits like the ability to digest foods and metabolize drugs. Given large mutational input per gut microbiome, adaptation to selective pressures should be rapid and repeatable, resulting in independent origins of adaptations in separate hosts. Despite this expectation, recent work has shown that broadly-beneficial adaptations can spread globally across human gut microbiomes via migration and subsequent recombination onto novel strain backgrounds rather than mutating *de novo* within each host. In this work, we test whether adaptations that have spread across hosts have multiple origins arising from rapid mutation. We find, surprisingly, that many adaptations have only one or a few origins. Using a stochastic model of an across-host selective sweep, we show that adaptations attaining detectable frequencies require adaptive mutation rates orders-of-magnitude lower than the mutation rate at a single base pair, and, consequently, that adaptations spreading across hosts bear structural or epistatic variants more complex than a single site mutation. In summary, we demonstrate that recombination across human gut microbiomes permits the spread of widely-beneficial adaptations with complex genetic architectures that otherwise would require a long waiting time to generate within an individual host.

426T Evolutionary Interactions Between the Microbiome and Host Life History Traits

Nourhan Mahmoud[1], Valeria Chavarin[1], Parvin Shahrestani[2] [1]Natural Science and Mathematics, California State University, Fullerton, [2]California State University, Fullerton

The microbiome plays a critical role in shaping host physiological and metabolic processes, yet its contribution to host evolutionary dynamics remains insufficiently explored. In *Drosophila melanogaster*, hosts exhibit genetic influence over microbial composition, while variation in microbial communities has been shown to alter key life history traits such as development rate and lifespan. Despite this reciprocal relationship, the temporal relationship between host trait evolution and microbiome change remains unclear. We used laboratory selection for fast development to examine associated changes to relative abundances of acetic acid and lactic acid bacteria within the fly microbiome over 30 generations. In parallel, we assessed how supplementation with acetic acid bacteria affects the rate of evolution when selection for fast development is applied. We found that selection for rapid development initially favors host trait evolution prior to the evolution of enhanced microbiome control. While work is in progress for the supplementation study, we expect that constant abundance of acetic acid bacteria will constrain the response to selection for fast development while accelerating host genetic regulation of microbial composition. Together, this work contributes to a broader understanding of the microbiome's role in life history evolution and aging.

427T Proportion of spontaneous beneficial mutations in yeast can be surprisingly high in some strains and environments

Joseph Matheson, Sergey Kryazhimskiy Ecology, Behavior and Evolution, University of California San Diego

One of the central suppositions of evolutionary theory is that the vast majority of spontaneous mutations are deleterious or neutral, and only a small fraction of them are beneficial. However, a recent experimental study proposed that such canonical left-skewed distribution of fitness effects (DFE) of mutations may be characteristic only for well adapted genotypes, whereas global epistasis among mutations may cause the DFEs of poorly adapted genotypes to be right-skewed with a large proportion of beneficial mutations. To test this hypothesis, we conducted a mutation accumulation (MA) experiment starting with 20 "background" strains of yeast *Saccharomyces cerevisiae* that had various initial growth rates across a set of laboratory conditions. We propagated ~20 MA lines per strain (~400 MA lines total) for 61 generations with minimal action of natural selection, which allowed each line to accumulate an estimated three to nine spontaneous mutations. We then measured how the growth rates of these lines changed during this experiment. We found that in the permissive nutritionally rich environment (YPD at 30°C), initially fast-growing strains on average lost fitness, as expected. Surprisingly, initially slow-growing strains on average gained fitness in YPD, even in the near complete absence of selection. However, the same MA lines exhibit a different pattern with respect to growth rate in another, more stressful, environment (SC at pH 7.0 at 30°C). In the latter condition, approximately equal fractions of MA lines gained and lost fitness irrespective of their genetic background. These results suggest that the statistical distribution of the effects of spontaneous mutations is more complex and variable than previously thought and can be substantially right-skewed at least in some genetic and environmental contexts.

428T Dissecting the evolution of gut-microbiome assembly in *Caenorhabditis* nematodes

Austin C. Link, Gavin C Woodruff School of Biological Sciences, University of Oklahoma

The nematode *Caenorhabditis elegans* has been a prominent facet of model organism biology for half a century now. Recently, many groups have investigated the ecological niche of *Caenorhabditis elegans*, the microbes it is associated with in nature, and the genetic basis of host-microbe interactions, utilizing the power of model organismal biology to address ecological questions. Interestingly, the sister species of *Caenorhabditis elegans* was recently discovered: *Caenorhabditis inopinata*. While *Caenorhabditis elegans* is a self-fertile nematode that lives in rotting plant detritus all over the world, having many isolates harboring a vast amount of genetic variation, *Caenorhabditis inopinata* lives within fresh figs, an environment wholly ecologically divergent from its sibling species and other members of *Caenorhabditis*. Furthermore, *Caenorhabditis inopinata* demonstrates remarkable morphological divergence from *Caenorhabditis elegans*, being twice as long, and has an opposite reproductive mode, being obligately sexual as opposed to self-fertile. Finally, their life histories are also divergent, with *Caenorhabditis inopinata* maturing much slower and being less fecund than *Caenorhabditis elegans*.

The fig in which *Caenorhabditis inopinata* lives, *Ficus septica*, has a distribution that stretches from East Asia to Australia, but *Caenorhabditis inopinata* has only been sampled for in Taiwan and Japan. This means that, unlike studies using *Caenorhabditis elegans*, current studies using *Caenorhabditis inopinata* utilize only a fraction of the potential genetic variation within the species. Consequently, when studying key divergent phenotypes between the two nematodes, not having isolates from the full biogeographical range of the species is a bottleneck that limits discovery. Having *Caenorhabditis inopinata* isolates from the entire range of its biogeographical distribution would enable the investigation of the genetic factors that regulate its microbiome by doing quantitative genetic studies. Comparing these underlying genetic factors to those that regulate host-microbe interactions within *Caenorhabditis elegans* would illustrate the divergence or conservation of gut-microbiome assembly phenotypes in this study system. We already have approximately forty distinct *Caenorhabditis inopinata* strains, but we want to aim for more. Thus, our mission is to sample *Ficus septica* across its entire distribution, starting in Australia, isolate enough *Caenorhabditis inopinata* to more accurately represent the variation within the species, characterize this variation, and use curated *Caenorhabditis*-associated microbial communities to investigate host-microbe evolution within the *Caenorhabditis* genus. This endeavor complements work we have already done to illustrate divergent life history responses to *Caenorhabditis*-associated bacterial isolates in these two species.

429T Population genomics of the VNI lineage of *Cryptococcus neoformans* from North Carolina

Erick Figueroa-Ildefonso[1], Kayla| T. Wilhoit[1], Marhiah C. Montoya[1], Claudia Zirión-Martinez[2], John R. Perfect[1], Paul M. Magwene[2] [1]Duke University, [2]Biology, Duke University

Cryptococcus neoformans is an opportunistic fungal pathogen causing ~112 000 deaths annually, it can be found in environmental niches where develops traits relevant for its pathogenicity. Among the *C. neoformans* lineages, VNI causes most of the human infections globally and is structured in three major sublineages. However, most of the studies on *C. neoformans* with global collections were conducted mainly with clinical isolates, with a significant underrepresentation of environmental isolates and from regions like North America. This lack of diversity limits our understating of population genetic variation locally and between environmental and clinical sub-populations. Given that most virulence genes are conserved across pathogenic and non-pathogenic species, a better understanding of the local population genetic structure and diversity of *C. neoformans* is relevant to explore how population genetic variation impacts pathogenicity traits. In that regard, we conducted the population genomic characterization of a sample of more than 300 isolates collected in North Carolina across 30 years from clinical and environmental sources. The screening for structural variants suggests the presence of chromosome duplications only in clinical isolates, supporting this as a mechanism of adaptation to within-host conditions. The maximum likelihood phylogeny shows the isolates from North Carolina clustering with the VNIa and VNIb sub-lineages and that are highly clonal. The VNIa sub-lineages are structured in five sub-clades, including a new sub-clade composed of mainly environmental isolates. Population genetics analyses shows evidence of recombination within the VNI lineage, suggesting sexual reproduction. These findings highlight the high diversity of sub-clades in NC isolates, despite being restricted to a single geographic region. The presence of the undescribed VNIa sub-clade highlights the importance of the environmental isolates to understand the genetic diversity of *C. neoformans* in nature. The population genetic structure and clonal characteristics described in this study pose a challenge to consider when conducting a genomic analysis aimed at discovering the genetic factors underlying differences in virulence traits in the subpopulations of *C. neoformans*.

430W MADS Box Genes and the Genetics of Floral Development in *Amborella trichopoda* Baill.

Sanam Parajuli[1], Madhav P. Nepal[2], Bibek Adhikari[2] [1]South Dakota State University, [2]Biology and Microbiology, South Dakota State University

The ABCDE model is a well-known general model of floral development in angiosperms with perfect flowers, with some modifications in different plant taxa. The Fading Borders Model was proposed to better explain floral patterning in basal angiosperms that typically possess spirally arranged floral organs. The MADS-Box gene family is central to these models and has greatly expanded in higher plants which is associated with increasing complexity in floral structures. *Amborella trichopoda* is a basal angiosperm with simpler floral features, and the genetic and functional roles of MADS-Box genes in floral development remain poorly understood in the species. The major objectives of this study were to perform a genome-wide identification and characterization of MADS-Box genes in *A. trichopoda*, and to analyze their expression in floral buds and mature flowers. We identified 42 members of the MADS-Box gene family in *A. trichopoda* with a Hidden Markov Model (HMM)-based genome-wide survey. Among them, 27 were classified into Type II or MIKC group. Based on our classification and orthology analysis, a direct ortholog *APETALA1* (*AP1*), an A-class floral MADS-Box gene was absent in *A. trichopoda*. Gene expression analysis indicated that MIKC-type genes were differentially expressed between male and female flowers with B-function orthologs: *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) in the species having differential expression between the two sexes, and E-function orthologs being upregulated in female flowers. Based on these findings, we propose a modification in the Fading Borders Model in *A. trichopoda* with a modified A-function, B- and E-function orthologs' expression being sex-specific, and C- and D-function genes having roles similar to that in the classical ABCDE model. These results provide new insights into the genetics underlying floral patterning in the basal angiosperm.

431W Characterizing and mitigating confounding by unreplicated evolutionary events in phylogenetic regression

Qifan Wang[1], Joshua G. Schraiber[2], Michael D. Edge[2], Matt Pennell[1] [1]Computational Biology, Cornell University, [2]Quantitative and Computational Biology, University of Southern California

When testing for the direct effect of one attribute on another using interspecific data, researchers need to account for the confounding influence of other attributes that have phylogenetic structure. A suite of approaches have been developed for this purpose, most notably phylogenetic regressions such as phylogenetic generalized least squares (PGLS). However, such methods can be misled by large-effect evolutionary events. We suggest one solution by borrowing from the statistical genetics literature and digging into the past of phylogenetic comparative methods. In genome-wide association studies (GWAS), confounding by population structure – an analogous problem – is often addressed by using the linear mixed model (LMM), where random effects are parameterized by a relatedness matrix. As in phylogenetics, GWAS studies are also susceptible to large-effect demographic events. A common practice is to use models with both random effects and fixed effects for individuals' coordinates on the eigenvectors of the same relatedness matrix. Schraiber et al. (2024) showed that PGLS and the LMM for GWAS are special cases of the same general model, suggesting that, in addition to including the phylogenetic structure as a random effect, one also include eigenvectors of the variance-covariance matrix as fixed effects in phylogenetic regressions. Building on the results of Schraiber et al. (2024), we use mathematical analysis and simulations to show under which scenarios this hybrid strategy is effective. We develop a novel method to visualize how different branches of the phylogeny contribute to the eigenvectors of the variance-covariance matrix. We also illustrate how quantile-quantile plots can help assess whether the phylogenetic structure has been effectively controlled. Applying the mixed approach to the co-evolution of gene expression across a cichlid fish phylogeny, we show that including the leading eigenvectors likely reduces the false positive rate. To facilitate the use of our approaches, we present a new R package, EIGER. We argue that these approaches can both help address the statistical problems arising from large, historical events and more generally, provide richer insights into the nature of phylogenetic confounding – and perhaps confounding in the statistical genetics setting as well.

432W Biophysical constraints on mRNA decay rates shape macroevolutionary divergence in steady-state abundance

Madhumitha Krishnaswamy[1], Catherine Felce[2], Alexander L Cope[3], Joshua Schraiber[4], Lior Pachter[5], Matt Pennell[1] [1]Computational Biology, Cornell University, [2]Division of Physics, Mathematics and Astronomy, California Institute of Technology, [3]Department of Biological Sciences, and Evolutionary Studies Initiative, Vanderbilt University, [4]Quantitative and Computational Biology, University of Southern California, [5]Division of Biology and Biological Engineering, and Department of Computing and Mathematical Sciences, California Institute of Technology

Evolutionary changes to gene expression are understood to be a major driver of phenotypic divergence between species. Studies which fit evolutionary models to multi-species 'omic' datasets indicate that steady-state mRNA expression levels show patterns consistent with evolutionary constraints, likely as a consequence of stabilizing selection. However, as all previous work has used bulk RNA measurements, it has been impossible to determine which of the many cellular processes that contribute to steady-state abundances underlie the divergence between species. Here we develop a novel paradigm for addressing this open problem, and use evolutionary systems drift as a general null model of macroevolutionary change. Using multi-species single-cell expression data and biophysical models, we estimate mRNA transcriptional burst sizes, splicing rates and decay rates across multiple species. We then derive phylogenetic models that describe the divergence of these rates under alternative evolutionary scenarios and fit these to the comparative data. In line with the expectations of systems drift, we test whether there are compensatory changes between transcriptional bursting and mRNA decay to maintain steady-state mRNA levels near their evolutionary optimum. We find evidence for biophysical constraints on the rates of mRNA decay, such that macroevolutionary divergence in expression is primarily a consequence of variation in transcriptional bursting.

433W Rapid evolution of synteny associated with multiple origins of dioecy and XY sex determination

Mark Hibbins[1], Casssandre Pyne[2], Mykhailo Sukmaniuk[2], Emily Glasgow[2], Meng Yuan[2], Yunchen Gong[2], Spencer Barrett[2], Stephen I Wright[2] [1]Ecology and Evolutionary Biology, University of Rochester, [2]University of Toronto

Chromosomal rearrangements can be major drivers of evolutionary innovation, shaping processes including local adaptation, speciation, and sex chromosome evolution. Multispecies synteny datasets are rich in information on genomic rearrangements, but statistical approaches that enable insights to be obtained from this information are still in their infancy. Here, we present a novel framework for the application of phylogenetic comparative methods to multispecies synteny datasets. We apply this approach to *Rumex*, a clade of flowering plants that exhibits rapid karyotypic evolution, including multiple origins of XY sex determination from hermaphroditic ancestors. Leveraging new genome assemblies, we find evidence for accelerated chromosomal rearrangement associated with evolutionary transitions to dioecy, illustrating how explicit phylogenetic hypothesis testing can generate new insights into adaptive hypotheses for rearrangements.

434T Tree-based transmissibility measures in compartmental epidemic models

Edoardo Manoni[1], Peiyu Xu[2], Andrew Clark[1], Jaehee Kim[1] [1]Computational Biology, Cornell University, [2]Molecular Biology & Genetics, Cornell University

Estimating how transmissibility differs among pathogen strains is a recurring goal of genomic surveillance, informing risk stratification and control. Because transmission chains are rarely observed, empirical analyses in genetic epidemiology often rely on tree-based summaries of pathogen genomes as proxies for relative transmissibility across strains (e.g., *Mycobacterium tuberculosis* and SARS-CoV-2): (i) SNP/evolutionary distance threshold clustering, summarized by cluster sizes and clustering rates, and (ii) terminal branch lengths in time-scaled phylogenies. These summaries are frequently interpreted as monotonic proxies for transmission intensity, based on the heuristic that faster spread yields more recent common ancestry, higher clustering, and shorter terminal branches. However, they also depend on sampling design, epidemic phase, latency, and host contact heterogeneity, making their connection to transmissibility unclear.

We derive analytic moments of these measures as functions of the transmission rate. Under an idealized setting of homogeneous mixing with no population structure and a large-population limit, we couple the sampled pathogen genealogy to SIR and SEIR dynamics and obtain closed-form expectations and variances for threshold-based clustering and terminal-branch statistics under explicit sampling schemes; under a molecular clock, a SNP threshold corresponds to a cutoff on pairwise coalescent times. We then systematically evaluate when these moments break down by relaxing the idealized assumptions in stochastic forward simulations, varying population size, contact-network topology, SEIR latent period, and sampling fraction and intensity. Our work provides a baseline for interpreting commonly used tree-based transmissibility proxies, identifies settings where these proxies are informative versus potentially misleading, and informs study design for genome-based epidemic surveillance.

435T Why some frogs took their eggs and left the pond

Katherine Martinson[1], Rebecca L Young[2], Arbel Harpak[2] [1]Biology, University of Texas at Austin, [2]University of Texas at Austin

The transition from aquatic to terrestrial life represents one of the most central shifts in vertebrate evolution. Frogs make an excellent system for studying its genetic basis, specifically in regard to their reproductive diversity and plentiful publicly available genomic data. Here, we used a suit of phylogenetic regression and phylogenetic genotype-phenotype association tools to ask: What were the evolutionary and ecological drivers of the switch to terrestrial egg laying in frogs? Among ecological and behavioral correlates of terrestrial vs. aquatic egg laying, smaller female body size and larger egg diameter stood out as highly indicative of terrestrial egg laying, even after adjusting for phylogenetic relationships. We used these results to inform our prior expectations for biological pathways that may be involved in genetic adaptation. Using methods to study both coding and non-coding variation among frog species, we point to candidate loci that we hypothesize are involve in the shift to terrestrial egg laying.

436T Expected value of the Sackin index under a biased speciation model

Daniel Bauman[1], Noah Rosenberg[2] [1]Institute for Computational and Mathematical Engineering, Stanford University, [2]Biology, Stanford University

Models of random phylogenetic trees are important mathematical tools in the analysis of phylogenetic data. Tree balance is a property of phylogenetic trees that helps describe and compare tree shapes and that can vary substantially among phylogenetic models. Here, we investigate the Sackin index, one of the oldest and most widely used tree balance measures, under a variety of phylogenetic models beyond the standard Yule and Uniform models. A general biased speciation model is a biologically interpretable tree-generating model in which species branch into pairs of descendants according to an intrinsic branching rate, which is then split between the pair of descendants. The proportion of the branching rate of the parent that is assigned to each descendant, the split fraction, is a model parameter. Our main result is a formula for the expected value of the Sackin index under the general biased speciation model. Special cases of the general biased speciation model include a case that sets the split fraction to be a constant for each branching event, and two special cases inspired by the Aldous beta-splitting model: a random choice of split fraction is sampled independently at each branching event, and drawn from a symmetric beta distribution in the Blum-François beta-splitting model or an asymmetric beta distribution in the Sainudiin-Véber beta-splitting model. As a corollary of our main result, we find formulas for the expected value of the Sackin index under these three cases of the biased speciation model. We use the formulas to study how the balance of trees depends on the model parameters. Finally, because the Yule model is a special case of the Blum-François beta-splitting model, we find a new proof for the classical result describing the expected value of the Sackin index under the Yule model.