Keynotes
Local genetic adaptation in humans and other primates
Aida M Andres

Presented by self
University College London (UK)

Genetic adaptation is a critical evolutionary process for populations facing environmental change. Change may be due to movements to novel habitats, such as when early humans migrated to colonise every corner of the world. Or it may be due to changes in the environment, such as when the natural habitat of African primates is altered by deforestation, anthropogenic pressure and climate change. Environmental change can have great consequences even when it’s geographically localised. Local genetic adaptation generates population differentiation and can markedly increase the diversity of a species, and locally adapted populations can have reduced fitness in different areas of the species’ range – likely including areas that may naively seem more favourable to the species. Recent genomic advances mean that these processes can now be investigated not only in well-studied, accessible species, but also in elusive, endangered ones that cannot be sampled directly. I will discuss my group’s work investigating the relevance of local genetic adaptation in two closely related species that differ markedly in their demographic history, ecology and geographic range: humans and chimpanzees. Our work suggests that local genetic adaptation has significantly contributed to population differentiation in both species, highlights major selective pressures in each species, and points to consequences of such evolutionary processes in conservation and human health.
Genetic diversity in the Anthropocene and the role of molecular studies for the future of biodiversity
Alicia Mastretta-Yanes
Presented by self
CONACyT - UNAM (México)

Genetic diversity is the basis of evolution and thus of all biological diversity. The evolutionary potential, and thus the resilience, of species and ecosystems depend on it. Human sustenance and nutrition also benefit from genetic diversity: it allowed the invention of agriculture, and today it continues to be pivotal in adapting crops to new environmental conditions and cultural preferences. Since its creation in 1992, the Convention on Biological Diversity (CBD), an international legal instrument for the conservation of biological diversity, its sustainable use, and benefit-sharing, has recognized the importance of genetic diversity. However, targets and indicators to measure it were lacking for decades, and genetic diversity was neglected in the CBD relative to species and ecosystem diversity, focusing mostly on domesticated species and their wild relatives. But not anymore. In December 2022, 196 countries signed the Kunming-Montreal Global Biodiversity Framework (GBF) of the CBD, an agreement to halt biodiversity loss by 2030. Among the commitments, an unprecedented milestone was set: conserving genetic diversity and monitoring and reporting its status for all species, not just those of socio-economic and cultural value. Molecular studies on the patterns and processes that impact the evolution of life followed a similar trend, first being focused on species of medical or agronomic interest, then later broadening their scope to the whole tree of life as genomic data became more widely affordable. The intersection of the new commitments to conserve genetic diversity and the unparalleled amount of molecular data that exist today offers a golden opportunity. This is the right time to think of new methods to answer old questions and to ask meaningful questions to build a better future.
PhyloG2P: the new science of connecting genomes to phenotypes via phylogenies

Scott V. Edwards

Presented by self
Harvard University (USA)

‘PhyloG2P’ is the latest in a series of approaches, including PhyloGWAS, for linking genomic variation to phenotypes using phylogenetic trees. PhyloG2P methods capitalize on different genomic signatures – presence/absence of genes, loss of function of genes, insertions and deletions, and shifting evolutionary rates – to find the genomic drivers of phenotypic change on clades that are not amenable to classical genetic approaches. In this talk I will present details of the PhyloAcc (‘Phylo-A-see-see’) family of Bayesian approaches to PhyloG2P and their application to diverse phenotypic traits in birds and mammals. PhyloAcc methods focus on associating changes in evolutionary rates of noncoding regions, such as conserved non-exonic elements (CNEEs), to changes in a binary or continuous trait. The basic PhyloAcc model finds CNEEs undergoing acceleration in concert with a change in a binary trait, such as convergent losses of flight in birds. We have used this approach on genome-wide CNEEs to find conserved putative enhancers that have undergone acceleration convergently in flightless birds (Paleognaths). We have further cataloged regions of open chromatin and applied massively parallel reporter assays (MPRAs) to demonstrate changes in function of enhancers undergoing acceleration. We have also recently produced PhyloAcc-GT, which incorporates gene tree heterogeneity into the model, as well as PhyloAcc-C and Halcyon, which associate accelerations in the genome with changes in a continuous phenotypic trait, such as maximum longevity in mammals. PhyloG2P methods provide powerful insights into genome-phenome connections in otherwise intractable lineages and help determine the relative contributions of protein and regulatory evolution to phenotypic diversification across the Tree of Life.
Adaptation and maladaptation in plant genomes.

Stephen Wright

Presented by self
University of Toronto (Canada)

What factors facilitate and constrain adaptation, and how does this influence plant genome evolution and diversity? In this talk, I will discuss research in the lab focused on two cases of extraordinary evolutionary parallelism acting on different timescales: the rapid evolution of herbicide-resistant weeds and the formation and degeneration of sex chromosomes. Using a combination of population and comparative genomics, these studies highlight the remarkable multifaceted adaptive potential of agricultural weeds, and the importance of recombination landscapes in driving the evolutionary trajectory and maladaptation of sex chromosomes.
Graduate Student Excellence Awards
Investigating the ecological suicide ("ecocide") theory in Rapa Nui with ancient DNA data
Bárbara Sousa da Mota

Bárbara Sousa da Mota, J. Víctor Moreno-Mayar, Tom Higham, Signe Klemm, Moana Gorman Edmunds, Francisco Torres Hochstetter, Martin Friess, Jesper Stenderup, Miren Iraeta-Orbegozo, Véronique Laborde, Evelyne Heyer, Morten E. Allentoft, Hannes Schroeder, Olivier Delanoeau, Anna-Sapfo Malaspinas

Curtin University (Australia), Musée de l'Homme (France), Muséum national d'Histoire naturelle (France), Rapa Nui Museum (previous affiliation) (Chile), Rapanui archaeologist (Chile), Regeneron Genetics Center (USA), Swiss Institute of Bioinformatics (Switzerland), University of Copenhagen (Denmark), University of Lausanne (Switzerland), University of Vienna (Austria)

Rapa Nui (Easter Island), in East Polynesia, is one of the most remote places in the world. After the peopling of the island a thousand years ago, it underwent profound environmental changes that have been associated with human presence, including the extinction of endemic birds and, more infamously, the complete disappearance of trees around the 1600s. It has been hypothesized that deforestation triggered famine, warfare and cannibalism, eradicating most of the Rapanui at the time - a so-called ecocide. However, there is a growing body of evidence that does not support that the Rapanui underwent a population collapse in the 1600s. Here, we addressed this question by analyzing 15 newly sequenced ancient Rapanui genomes (0.4x-26x) dated to the early 1800s. Using pseudohaploid and imputed data, we examined whether there was increased relatedness and consanguinity among the individuals by estimating kinship, inbreeding and runs of homozygosity (ROH). We found no close relatives up to 3rd degree and found very low inbreeding levels. Although effective population size was low (maximum ~2000 individuals), we show that the population grew steadily after the peopling of the island. Furthermore, effective population size and ROH estimates for simulated genomes under strong population-collapse scenarios did not match the observed data. In other words, the ancient Rapanui genetic diversity is inconsistent with a population collapse taking place in the 1600s. This is the first contribution from genetics to debunk the most famous example of ecological suicide.
A lethal mitonuclear incompatibility in complex I of natural hybrids

Benjamin M Moran

Benjamin M Moran, Molly Schumer
Stanford University (USA)

The evolution of reproductive barriers is the first step in the formation of new species and can help us understand the diversification of life on Earth. These reproductive barriers often take the form of "hybrid incompatibilities," where genes derived from two different species have diverged such that they no longer interact properly. Theory predicts that hybrid incompatibilities may be more likely to arise at rapidly evolving genes and in multi-protein complexes, both of which implicate the mitochondria as potential hotspots for incompatibilities. However, there has been sparse empirical data to evaluate these predictions, especially outside experimentally manipulated contexts. Here, we describe a mitonuclear incompatibility involving three genes within respiratory Complex I in naturally hybridizing swordtail fishes. Through a combination of morphometrics, histology, and respirometry, we develop an integrative view of the phenotypes which underly severe selection against this incompatibility in the wild. Ancestry incompatible with mtDNA at either of the two nuclear genes creates a distinct lethal phenotype of either stalled embryonic development or postnatal organ dysfunction, but both genes modify the other’s effects. Finally, we document a history of accelerated evolution of the genes involved, and provide evidence for introgression of the incompatible genes into a congener. This work thus provides the first glimpse into the genetic architecture, physiological impacts, and evolutionary origin of a multi-gene incompatibility impacting naturally hybridizing species. Our results are consistent with the idea that strong pairwise interactions may overshadow more complex mitonuclear interactions, which will be revealed with increasing experimental power.
Evolutionary shift of a tipping point forestalls collapse in a microbial community.
Christopher Blake

Christopher Blake, Michael J McDonald
Monash University (Australia)

Climate change is driving global ecosystems rapidly towards tipping points, critical thresholds where small environmental changes can lead to drastic ecological shifts. While progress has been made towards identifying these tipping points, one crucial question remains unanswered: Can evolution modify a tipping point and delay a potential ecosystem collapse? Here, we investigate the power of evolutionary processes to alter tipping points and potentially safeguard communities from collapse. We evolved a two-species microbial system over 4000 generations to map the changes in ecological stability resulting from coevolution. After that, we adapt our two-species community to an acute environmental stress and reassess the tipping point. Our findings reveal that through adaptation, communities can not only delay their tipping points but also alter their overall stability landscape, expanding the range of environmental conditions under which they can flourish. By developing a mathematical model, we identify how specific traits, such as growth rate, carrying capacity, interspecific interactions, and resistance to environmental change, can enhance ecological resilience. In this study, we demonstrate that the adaptation of key species within an ecosystem can significantly impact its tipping point, potentially postponing or preventing collapse. These findings open doors for novel strategies like microbiome engineering and directed evolution, which could be employed to bolster the resilience of ecosystems facing the challenges of climate change.
The provenance of Proto-North Dravidian ancestry in the Indus valley
Jaison Jeevan Sequeira

Jaison Jeevan Sequeira, George van Driem, Ranajit Das, Mohammed S Mustak

Department of Applied Zoology, Institut für Sprachwissenschaft, Mangalore University (India), Universität Bern (Switzerland), Yenepoya (deemed-to-be) University (India), Yenepoya Research Centre

Research has shown that the present-day population on the Indian subcontinent derives its ancestry from at least three components: Iranian plateau farmer-related, Pontic-Caspian steppe pastoralist-related and Andamanese hunter-gatherer related. However, with more sequences of ancient and modern genomes and fine structure analyses, we can expect more complex ancestries. In this study, we focus on the North-Dravidian linguistic groups to propose a fourth putative source which may have branched out from the basal Middle Eastern component that gave rise to the Iranian plateau farmer-related ancestry. The Elamo-Dravidian theory and the linguistic phylogeny of the Dravidian family tree provide a number of possible chronological fits for the genetic findings presented here but also leave a number of open questions. Our findings show a correlation between the linguistic and genetic lineages in the North Dravidian language speakers when they are modeled together. We suggest that this source, which we call 'Proto-North Dravidian' ancestry, emerged around the time of the Indus Valley Civilisation. This ancestry is distinct from all other sources described so far, and its plausible origin around 4,500 years ago on the Iranian plateau bordering the Indus valley strengthens the idea of a Greater Dravidian heartland before the arrival of Indo-European languages on the Indian subcontinent. Admixture analysis shows that Proto-North Dravidian ancestry is carried by most of the present day non-tribal populations. Our research highlights the importance of population-specific fine structure studies to avoid oversimplification of ancestral reconstruction.
Beyond Continental Groups: Unveiling Dynamic Human Genetic Communities with a Novel Network Analysis Pipeline

María José Palma Martínez

María J Palma Martínez, Yuridia Posadas García, Brenda E López Ángeles, Claudia Quiroz López, Anna C F Lewis, Kevin A Bird, Tina Lasisi, Arslan Zaidi, Mashaal Sohail

Brigham and Women’s Hospital, Centro de Ciencias Genómicas, Escuela Nacional de Antropología e Historia (Mexico), Institute of Health Informatics, UNAM (Mexico), Univ. of California (USA), Univ. of Michigan (USA), University of Minnesota (USA), and Harvard Medical School (USA).

Traditional clustering and visualization approaches in human genetics often operate under statistical and conceptual frameworks that assume inherent, discrete groupings. These methods can inadvertently simplify the multifaceted relationships among individuals, functioning to unveil assumed underlying differences between static groups. Recently there has been a growing embrace of more sophisticated and dynamic methodologies that better reflect the interconnected and fluid nature of human populations. Here we introduce a network-based computational pipeline and visualization tool grounded in relational thinking as a departure from typological models of analysis and interpretation that strive to categorize individuals into a predefined number of sets. Our pipeline offers the flexibility of using different metrics, such as genetic relationship matrices (GRM), principal components, identical-by-descent (IBD) segments to construct networks. Through the application of community detection (Louvain), we infer communities at multiple resolutions. These aspects facilitate more dynamic data-driven, assumption-free analyses and visualizations of population structure that can identify patterns at different temporal and spatial resolutions suited to the particular research questions. We applied our pipeline to a dataset merged from the 1000 Genomes and Human Genome Diversity Project. We also infer communities for trait-specific variation to capture co-occuring patterns relevant for specific traits. Our analysis not only reveals the limitations of traditional groupings but also captures the complexities introduced by recent and distant demographic events and evolutionary processes. To enable broader engagement with these intricate genetic landscapes, we provide a user-friendly web application (https://sohail-lab.shinyapps.io/GG-NC/) for interactive visualization.
MAYEX is an ancient long non-coding RNA recruited for X chromosome dosage compensation in lizards

Mariela Tenorio

Mariela Tenorio, Diego Cortez, Selene Fernandez-Valverde, Katarzyna Oktaba, Samantha Cruz-Ruiz, Sinai López, Fania Santiago, Joanna Serwatowska, Jose Corona-Gomez, Sergio Encarnacion, Magdalena Hernández, Robert Ossiboff, Fausto Méndez de la Cruz, Mario Zurita, Cynthia Flores - Aguirre, Diego Arenas-Moreno
BABS, CCG, CDPM, CINVESTAV, IB, IBT, IPN (Mexico), LANGEBIO, UF (USA), UNAM (Mexico), UNSW (Australia)

Long non-coding RNAs (lncRNAs) are essential regulatory elements of sex chromosomes that act to equalize gene expression levels between males and females. XIST, RSX, and roX2 regulate X chromosomes in placentals, marsupials, and Drosophila, respectively. Given that the green anole (Anolis carolinensis) shows complete dosage compensation of its X chromosome, we tested whether a lncRNA was involved. We found an ancient lncRNA, MAYEX, transcribed in multiple tissues of both males and females that gained male-specific expression and sequence conservation more than 89 million years ago, likely after the origin of the XY chromosomes in pleurodonts.

Using RNA-seq, ChIP-seq, Hi-C, ChAR-seq, CHART, MS, and CRISPRi data, we found that, specifically in males, MAYEX binds to histone 4, an acetyltransferase, and other transcription activators, to modify the epigenetic landscape of the X chromosome and then significantly increase the expression levels of the X. The MAYEX system operates in all tissues during development and in adults. MAYEX is the first lncRNA in reptiles linked to a dosage compensation mechanism that balances the expression of full sex chromosomes. The similarities between mammals, the fruit fly, and now lizards, indicate that lncRNAs have been concurrently recruited in distant groups during evolution to establish chromosome-wide cis-regulatory mechanisms that control the expression levels of entire sex chromosomes and restore balanced expression ratios between males and females, constituting a remarkable instance of convergent regulatory molecular evolution.
Lactase persistence (LP), the ability to digest milk into adulthood, is a striking example of natural selection in humans attributed to mutations upstream of the LCT gene. Within the past 5,000 years, LP-associated alleles have rapidly increased in frequency and are present in nearly 80% of Europeans. Yet there is limited understanding of LP in non-Europeans. India is the world’s largest producer of milk, from both cattle and water buffalo, and it is consumed across the subcontinent. Puzzlingly, a previous study found that LP-associated alleles identified in Europeans are on average seen in ~10% of Indians. We assembled a genome-wide dataset of 7,568 present-day South Asians and 1,223 ancient DNA samples (7500 BCE – 1650 CE) from South and Central Asia. For a subset of contemporary Indians, we had access to diet questionnaires showing 30–70% of Indians routinely consume milk. Across India, dairy consumption is correlated to LP-associated allele frequency, with highest rates in the north, decreasing south and east. Turning to ancient DNA, we find LP-associated alleles were first seen during the historical period in South Asia. Notably, like in Europeans, LP-associated allele frequency is associated with steppe pastoralist-related gene flow in India that occurred ~1900–1500 BCE. This indicates strong selection pressure to retain the LP-associated haplotype, which is at over 65% frequency in some Indian pastoralist groups. Our phenotypic data combined with contemporary and ancient genomic data aid in understanding the origins and evolution of LP outside Europe.
Human sex ratio at birth has virtually no detectable genetic variation: Was Fisher wrong about the evolution of sex ratio?
Siliang Song

Siliang Son, jianzhi Zhang
University of Michigan (USA)

The human sex ratio (fraction of males) at birth is close to 0.5 at the population level, an observation commonly explained by Fisher’s principle that, when both sexes require equal parental investment, natural selection favors the production of the less common sex until the sex ratio reaches 0.5. However, past human studies yielded conflicting results regarding the existence of sex ratio-influencing mutations—a prerequisite to Fisher’s principle, questioning whether the nearly even sex ratio is instead dictated by the random X/Y chromosome segregation in male meiosis. Here we show that, because a person’s offspring sex ratio has an enormous measurement error, a gigantic sample is required to detect sex ratio-influencing genetic variants. Conducting a genome-wide association study that is more powerful than previous studies, we detect a rare allele at rs144724107 upstream of ADAMTS14, a constituent of the ADAMTS proteinase gene family with roles in fertilization, that is associated with a 18% sex ratio reduction. Given the difficulty in precisely measuring offspring sex ratio, it is unsurprising that its estimated heritability is effectively zero. In fact, the estimated heritability would be virtually zero even if offspring sex ratio is as genetically variable as the highly heritable human standing height. These analyses, along with simulations of human sex ratio evolution respectively under purifying and positive selection, demonstrate the compatibility of a nearly zero estimated heritability with Fisher’s principle, suggesting the potential presence of small-effect and/or rare, large-effect genetic variants influencing the offspring sex ratio in humans.
Rapid coevolution preserves the epigenetic establishment of telomere protection

Sung-Ya Lin

Sung-Ya Lin, Hannah Futeran, Mia T. Levine
University of Pennsylvania (USA)

During sperm maturation, paternal chromosomes undergo extreme compaction. Immediately following fertilization, maternally deposited proteins reverse this paternal chromosome compaction and reestablish canonical protein packaging. This process is conserved from flies to humans, yet many of these maternally deposited proteins evolve adaptively. The causes and consequences of this adaptive evolution remain mysterious. In Drosophila, one such adaptively evolving protein, HipHop, protects chromosome ends from lethal fusions. Maternal HipHop is loaded de novo onto zygotic paternal telomeres, where it recruits other essential subunits of the telomere protection complex. We hypothesized that the de novo establishment of paternal telomere protection requires recurrent innovation of HipHop. To test this, we swapped hiphop from Drosophila yakuba (hiphop[yak]) into D. melanogaster. We found that most embryos of hiphop[yak]/+ females failed to hatch, suggesting that HipHop[yak] is toxic. These embryos arrest at the first mitosis and show delayed paternal chromosome condensation, chromatin bridging, and mitotic catastrophe — phenotypes associated with paternal telomeres that fail to assemble the end-protection complex. Importantly, one other member of this complex both physiologically interacts with HipHop and evolves adaptively. We hypothesized that HipHop[yak] interferes with paternal telomere recruitment of this D. melanogaster protein, called HOAP. Consistent with this hypothesis, swapping in HOAP[yak], the conspecific interaction partner of HipHop[yak], rescues embryonic lethality. Furthermore, interspecific chimeric HipHop proteins revealed that evolution of HipHop’s HOAP-interaction domain alone is necessary and sufficient for end-protection. Our study demonstrates that essential protein coevolution shapes the epigenetic establishment of vital chromosomal landmarks on zygotic paternal DNA.
S1 - Structural phylogenetics: investigating deep evolutionary history using protein structure.
Modeling archaic hominin protein structures reveals evidence of reduced efficacy of selection

Ava Xu

Ava Xu, David Rinker, Bian Li, Christopher Moth, Gregory Sliwoski, Jens Meiler, John Capra
AB Hancock Jr. Memorial Laboratory for Cancer Research (USA), San Francisco (USA), University of California, Vanderbilt University (USA)

Protein-coding genetic variants distinguishing anatomically modern humans and their archaic hominin relatives, the Neanderthals and Denisovans, are of great interest due to their potential contribution to lineage-specific traits. However, the functional effects of most lineage-specific variants are unknown. We combine high-coverage archaic genome sequences and powerful machine learning methods to evaluate the structural and functional effects of lineage-specific variants. We focus on 1050 fixed or very high-frequency lineage-specific derived missense variants that are likely strongly enriched for positive selection. To account for potential human-specific biases, we consider a diverse range of functional metrics including biophysical (ddG), evolutionary (within and between species sequence constraint), and deep learning (AlphaMissense, ESM1b) models. Variants derived in the archaic lineage are significantly more disruptive to protein structure and function compared to human-derived variants. This result holds for every functional metric used and is robust to the thresholds used to define lineage-specificity. Moreover, it is robust to filtering sites of potential ancient DNA damage. We also compare the effects caused by lineage-specific variants to a matched set of 409 annotated pathogenic missense variants from the ClinVar database. As expected, both human- and archaic-derived missense variants have smaller effects on proteins than human disease-associated missense variants. Our results find a substantial role for genetic drift within archaic populations and support previous studies concluding that the effective population size and efficacy of selection were lower in archaic hominins than modern humans.
Life has been evolving on this planet for four billion years. Evolution allows for exploration of the vast sequence space comprised by polymers of the twenty standard amino acids – a space that exceeds the number of molecules in the universe. From within this space, biology has discovered unique solutions to the challenges of growing and persisting in the ever-shifting diversity of Earth’s environments. Undoubtedly, biogeochemically critical microbial enzymes are central to this dynamic, long-term interaction between life and the environment. In this talk, I will focus on the evolutionary sequence and structural reconstruction of key enzymes that evolved early in the history of life and persisted through planetary extremes, such as the Great Oxidation Event. These examples include the RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) and nitrogenases. Particularly, I will discuss our laboratory’s approach combining ancestral sequence reconstruction, structural inferences and synthetic biology to reveal the lost evolutionary trajectories of critical enzymes across deep time.
Animal viruses pose a threat when they evolve traits that enable spillover into humans. The first step necessary for many viruses to cross species boundaries is to bind with high affinity to novel host receptors, allowing entry to initiate infection. However, the molecular mechanisms by which viruses evolve to recognize novel host receptors are not well-characterized. An extreme example of variation in viral receptor usage was recently described in a group of coronaviruses called merbecoviruses—a large subgenus that includes the epidemic human Middle East respiratory syndrome coronavirus (MERS-CoV) and its diverse relatives from mammals. Though MERS-CoV and closest relatives from bats and camels utilize the DPP4 protein as their entry receptor, more distant bat merbecoviruses were recently shown to utilize ACE2 protein receptor; the genetic mechanisms by which these differences in receptor usage evolved have not been characterized. To isolate the evolutionary interval when merbecovirus receptor usage changed, we perform phylogenetic reconstructions of ancestral merbecovirus receptor-binding domains and characterize receptor-binding specificities via a yeast-display-based deep mutational scanning platform. From these data, we identify the evolutionary coincidental sequence substitutions that enable this change in receptor usage. We describe a combinatorial mutational scanning experiment for subsequent genetic dissection of the causal historical substitutions, including the role of epistasis or promiscuous intermediates in this historical evolutionary transition. This work identifies the mechanisms of a dramatic change in protein function while informing efforts to predict or prevent future spillover from this viral family of future pandemic potential.
How did Nature discover the coding rules needed to read blueprints templated in the sequences of their own genes? The reflexive translation of symbols in one chemical language to another defined genetics. Yet, the co-linearity of codons and amino acids is so commonplace an idea that few even ask this question. That readout is done by two sets of proteins, called aminoacyl-tRNA synthetases (AARS). AARS must enforce the rules first used to assemble them. We will find the root of translation only when we can describe and experimentally validate the earliest AARS*tRNA “cognate pairs” and the “structural codes” they used to recognize both amino acid and RNA substrates. Phylogenetics and AI now help identify new AARS “urzymes” from novel sources. We’ve shown that urzymes actually prefer to acylate T?Cminihelices, rather than full-length tRNAs. Catalysis by urzymes does not need active-site amino acids that were not present at early times. Studies of RNA specificity confirm details of the code used by urzymes to recognize cognate RNA minihelices. We have adduced substantial evidence that both strands of a single ancestral gene coded for Class I and II AARS on opposite strands. Translated bidirectional gene sequences are inside out, and so fold into different 3D structures. That leads to contrasting amino acid and RNA substrate binding modes. These experimental data therefore argue that these initial substrate specificities are rooted in the base pairing between their coding sequences. This work provides phylogenetics a platform for discovering the origin of genetics.
Polyphyletic insertions in the bacterial DNA-directed RNA polymerase

Claudia Alvarez Carreno

Claudia Alvarez Carreno, Angela Huynh, Christine Orengo, Loren Dean Williams
Georgia Institute of Technology (USA), NASA Center for the Origin of Life (USA), University College London (United Kingdom)

Transcription is the Central Dogma process in which the RNA polymerase (RNAP) transcribes DNA into RNA. Bacterial RNAP contains five subunits called alpha1, alpha2, beta, beta’, and omega. All these subunits have archaeal orthologs. Though RNAP-beta and RNAP-beta’ are universally distributed, their multi-domain architectures vary significantly between archaia and bacteria, and among bacterial species. Here, we use sequences and structures to retrace the deep evolution of RNAP-beta and RNAP-beta’ in bacteria. We present the results of a comparative structural analysis of experimentally determined structures and AlphaFold predictions to investigate the deep divergence of bacterial RNAP-beta and RNAP-beta’. Our combined analysis suggests that phylogenetic distributions of bacterial lineage-specific insertional domains follow the Tree of Life. We observe that bacterial lineage-specific domains in RNAP-beta belong to a group of domains that we call BEAN (Broadly Embedded ANnex) and that in RNAP-beta’, bacterial lineage-specific domains are HAmmerhead/BArrel-Sandwich hybrid (HABAS) domains. We discuss mechanisms that might account for the presence of homologous insertional domains in non-equivalent locations of the same protein.
Foldtree. Empirical benchmarking of structural phylogenetics methods

David Moi, Charles Bernard, Christophe Dessimoz
Institut Pasteur (France), UNIL DBC (Switzerland)

With the emergence of accurate AI generated structural models it has become possible to generate phylogenetic trees of protein families incorporating this structural information. It is still unclear which is the best evolutionary model to represent the phylogenetic signal of the final folded protein as we transition away from methods that represent this object as a series of independently evolving positions towards a more holistic view incorporating tertiary structure and interactions between residues. In this work we present an empirical benchmarking of thousands of protein families using a battery of structural phylogenetics approaches currently available. Currently our experiments show that the use of structural alphabets to align residues improves alignment quality and subsequent tree topology inference. However, constraining this structural character alignment to statistical models typically used for maximum likelihood phylogenetics does not appear to improve tree topology quality in our benchmarking experiments over the use of simpler models. The methods we find to be most apt at recapitulating species tree signals from our benchmarking experiments are showcased using difficult phylogenies sometimes involving horizontal transfers and long evolutionary histories stretching back billions of years. We show structural phylogenetics applied to the structural superfamilies of Fuseins (containing Hap2, Fsx1 and viral class II fusogens), the RRNPPA peptidic quorum sensing family and other illustrative examples. We will also present preliminary work regarding the expansion of structural alphabets tuned specifically for phylogenetic analysis.
Structural phylogenetics of the jelly roll fold sheds light on viral evolution

Desiree Beate Langer

Desiree B. Langer, Ashar J. Malik, Peter F. Stadler, Peter W. Hildebrand, Caroline Puente-Lelievre, Nicholas Matzke, Jane R. Allison, Anthony M. Poole

- (New Zealand), Baker Heart and Diabetes Institute (Australia), Bioinformatics Group, Computational Biology and Clinical Informatics, Institute of Computer Science, Institute of Medical Physics and Biophysics, School of Biological Sciences, School of Chemistry and Molecular Biosciences, University of Auckland (New Zealand), University of Leipzig (Germany), University of Queensland (Australia)

Comparing virus coat architectures has revealed the presence of a common jelly-roll fold in viruses infecting all three domains of life, spanning both RNA and DNA viruses. While structural similarities are evident, protein sequence conservation is low, making traditional sequence-based tools impractical for building viral evolutionary trees. Our lab developed a novel method for constructing evolutionary trees based on protein structure, which can aid in reconstructing deep relationships even when sequence similarity is limited. Our method makes use of molecular dynamics to generate statistical support values for phylogenies built from protein structure. These results will be compared with a method that overcomes the dependence on distance-based methods and uses a structural alphabet (3Di), which is used for the recently developed program Foldseek for deep homology search. Our phylogenies reconstruct expected viral relationships, indicating that evolutionary signal can be extracted from the comparison of protein structures. Our results also shed light on the evolution of jelly roll architectures themselves, revealing that double jelly roll (DJR) folds have emerged multiple times from the single jelly roll (SJR) fold via ancient gene duplication.
Unicore Enables Ultra-fast and Accurate Phylogenetic Reconstruction with Structural Core Genes

Dongwook Kim

Dongwook Kim, Sukhwan Park, Martin Steinegger
Republic of), Seoul National University (Korea

Breakthroughs in AI-driven protein structure prediction are transforming biology. Efforts [1,2] to integrate predicted protein structures into phylogenetic analyses have shown promise, however, have been constrained by the availability of accurate models or computational demands. The ProstT5 [3] protein language model and Foldseek’s 3Di structural alphabet [4] are circumventing the need for slow structure prediction with over three orders of magnitude faster prediction of structural sequences from amino acid input, and are posed to revolutionize phylogenetic reconstruction.

Here, we present Unicore, an ultra-fast and accurate method for structure-based phylogenetic inference that leverages predicted 3Di sequences, focusing on the identification of highly conserved structural core genes. We demonstrate this approach to reliably reconstruct phylogenetic relationships with linear run-time scaling over the size of the target proteomes, while being congruent with conventional approaches using orthologs or core genes. Unicore is universally applicable to any given set of taxa, even spanning superkingdoms and overcoming limitations of previous methods requiring orthologs of fixed taxonomic scope. As an ultra-fast solution for structural phylogenetic reconstruction without taxonomic limitations, we expect Unicore to pave the way to a new era of structural phylogeny.

Exploring protein evolution through natural and simulated sequences using experimental and modeled 3D structures

Hector Romero

Hector Romero, Mauricio Langleib, Martin Graña
Institut Pasteur (Uruguay), Universidad de la República (Uruguay)

Evolving new molecular functions and fine tuning existing ones is the core of evolutionary processes at the molecular level. In turn, in many cases this is the foundation of changes in higher order structures such as complexes, networks (metabolic, signalization, etc.), both at cellular and organismic levels. The path followed by molecular machines through this process has fascinated biologists for decades. Coupling present sequencing capabilities with the power of deep learning methods for protein structure prediction, we explore the fate of theoretical mutations and substitutions within predicted structures and compare it with natural observations using both experimental (when available) and modeled data. First, we took a forward in time approach, in which we simulate evolutionary processes at the sequence level using well known markov models of molecular evolution both at the DNA level (e.g. HKY, GTR, etc.) and directly at the protein level (e.g. Dayhoff, JTT, LG, etc.). Sampling through different evolutionary times and modeling 3D structure in each moment performing 100 different trajectories. And then compare the foldability and difference with the original structure. Then, we used actual sequence data of homologs to model (and when available used experimental 3D structures) to obtain ‘natural’ variants of the potential trajectories.Measures like TMscore, RMSD, stability of the protein, etc. are taken into account in order to evaluate the viability of the variants, and study their distribution in the different models.Finally we compare the actual sequences and their variability with the simulated ones.
A fundamental problem in bioinformatics is determining the degree of evolutionary relatedness among Twilight Zone homologs that by definition have diverged to a point where they share less than 30% identity and yet retain similar structures and/or functions. Many researchers have leveraged experimentally determined structures in the quest to classify Twilight Zone proteins. Such endeavors are difficult and require great investments of time and money. Using both simulated and real-world data we characterize molecular weight-hydrophobicity physicochemical dynamic time warping (MWHP PCDTW or MWHP DTW for short) as a method requiring only the amino acid sequence to quantify similarity of related proteins in the Twilight Zone. This is a step forward in determination of the degree of relatedness among Twilight Zone proteins and most notably allows for the discrimination of random similarity and true homology in the 0–20% identity range. This method was previously presented expeditiously as a result of the Covid-19 pandemic because it was able to functionally cluster ACE2-binding betacoronavirus receptor binding domain’s (RBD) a task that has been allusive using standard techniques. Here we show that one reason that MWHP DTW is superior to standard alignment techniques with regard to comparisons within the Twilight Zone is because standard alignment produces incorrect alignment when comparisons are made within the Twilight Zone. Further, we present an extended definition of the Twilight Zone that incorporates the dynamic relationship between structural, physicochemical and sequence-based metrics.
Elucidating the origin and molecular evolution of the cholesterol biosynthesis pathway.

Janeth Alejandra García

Janeth A. García, Antonio Lazcano, Ricardo Hernández-Morales
Departamento de Biología Evolutiva, Facultad de Ciencias, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México. (Mexico)

Cholesterol is an essential component of animal cell membranes. The biosynthesis of this sterol has been reported in other eukaryotic organisms, such as fungi and plants. It is believed that the appearance of cholesterol in cell membranes is correlated with the evolution of eukaryotic organisms and the increase in atmospheric oxygen that occurred 2.32 billion years ago, molecular oxygen is already required for the completion of cholesterol biosynthesis, but recent research has found that the halophilic bacterium Enhygromyxa salina may be able to synthesize cholesterol de novo, raising questions about the antiquity of this pathway. Therefore, this project will attempt to address the origin and molecular evolution of the cholesterol biosynthetic pathway by comparing the primary and tertiary structure of the enzymes involved in the biosynthesis. Preliminary results of our study suggest that the ability to synthesize sterols may be more widespread than previously thought, opening new perspectives on the evolution and distribution of this pathway across domains of life.
How does evolution assemble protein complexes made of interdependent parts? We use the Toll-like receptor 4 (TLR4) complex as a model for this question. TLR4 recognizes bacterial lipopolysaccharides (LPS) and activates inflammation. Here, we explored the evolution of the CD14, an important support protein that transports LPS to the TLR4 complex. Prior to our work, it was thought that CD14 evolved in the ancestor of tetrapods, ~70 million years after the TLR4 complex formed. Using phylogenetic and syntenic analyses, we found that CD14 is present in ray-finned fishes. Using cell-based assays, we validated that CD14 from two extant fishes were indeed functional, suggesting CD14 evolved concurrently with the rest of the TLR4 complex. We next sought to understand the evolution of CD14 function. CD14 is a paralog of TLR2 that evolved by serial truncation after duplication. Using ancestral sequence reconstruction and structural modeling, we found that CD14 evolved its LPS delivery function when a truncation exposed a hydrophobic surface, creating a functional site de novo. Ancestral TLR4 and CD14 complexes induce inflammation in experimental assays, demonstrating that the complex was active in the bony vertebrate ancestor. The TLR4 complex was then modified along descendent lineages. In particular, the model vertebrate Danio rerio lost CD14 but gained CD14-independent TLR4 activity. This work improves our understanding of the evolution of the TLR4 complex, reveals that functional sites can evolve by exposure of a hydrophobic surface, and helps map the results of studies in Danio rerio to human biology.
Evolution and functional role prediction of the CYP6DE and CYP6DJ subfamilies in Dendroctonus (Curculionidae:Scolytinae) bark beetles

Juan Manuel Quijano Barraza

Juan Manuel Quijano Barraza, Gerardo Zúñiga, Claudia Cano Ramírez, María Fernanda López, Gema L Ramírez Salinas, Moises Becerril
Instituto Politécnico Nacional (Mexico)

Dendroctonus beetles are natural components of coniferous forests that spend most of their lives under the bark, where they are exposed to toxic terpenes present in the oleoresin. Cytochrome P450 (CYP) enzymes are involved in the detoxification of these compounds. It has been demonstrated that CYP6DE and CYP6DJ subfamilies hydroxylate monoterpenes, whose derivatives can act as pheromone synergists be pheromones themselves. Given the functional diversity of CYPs, we investigated whether these cytochromes have retained their function throughout the evolution of these insects. We performed a Bayesian phylogenetic analysis to determine subgroups of cytochromes in these subfamilies. Subgroups were mapped and reconciled with the Dendroctonus phylogeny. Molecular docking analyses were performed with the cytochromes of each subgroup and enantiomers of α-pinene and β-pinene, (−)-3-carene, β-myrcene and R-(+)-limonene. In addition, functional divergence analysis was performed to identify changes in catalytic site and/or protein folding. The mapping and reconciliation analysis showed different patterns for each subgroup. Functional predictions indicated that the cytochromes analyzed can hydroxylate all monoterpenes with preferential affinities to different monoterpenes. The CYP6DE subfamily has experimented type I and II divergence, whereas the CYP6DJ subfamily has evolved under strong functional constraints. Results suggest cytochromes of the CYP6DE subfamily evolve to reinforce their detoxifying capacity hydroxylating mainly α- and β-pinene to (+) and (−)-trans-verbenol, compound used as a pheromone by several Dendroctonus species; whereas the CYP6DJ subfamily appear to retain its original function related to the detoxification of these compounds.
Secondary structure conservation sheds light on the forces that drive the evolution of condensate proteins
Lisa E Kursel

Lisa E Kursel, Kaan Goktepe, Ofer Rog
University of Utah (USA)

Condensates are a newly-appreciated way to organize the cell. Like membrane-bound organelles, condensates enable spatial and temporal regulation. However, condensates are regulated by distinctive molecular interactions. Rather than stable, tight interfaces, condensates assemble through weaker, labile interactions. This mode of assembly could allow for sequence variability, limiting the applicability of current knowledge on protein evolution. Even when carrying out essential functions, condensate members tend to lack conserved amino acid sequence and identification of orthologs relying on per-site identity has been challenging. We addressed this paradox by analyzing the evolution of proteins from two condensates, centrosomes (essential for cell division) and the synaptonemal complex (SC, essential for reproduction), in Caenorhabditis. We found that centrosome and SC proteins have an unusual evolutionary signature; they are significantly more diverged than other proteins, but retain conserved secondary structures including coiled-coil domains. We found similar evolutionary signatures in Drosophila and Eutherian mammals despite no homology across clades. This suggested we could identify condensate proteins by leveraging their diverged primary sequence but conserved secondary structure. We demonstrated this by cloning a novel SC protein, PpaSYP1, in a distantly related nematode, Pristionchus pacificus. PpaSYP1 has no homology to any known SC protein. We made a knockout and tagged version of PpaSYP1 and confirmed that it localizes to the SC and is required for reproduction. Overall, we observe patterns of functional and structural conservation that do not depend on primary sequence conservation, shedding light on the forces that drive the evolution of condensate members.
Corrin biosynthesis in the last universal common ancestor
Luca David Modjewski

Luca D. Modjewski, Val Karavaeva, Natalia Mrnjavac, Filipa L. Sousa, William F. Martin
Department of Functional and Evolutionary Ecology, Heinrich Heine University Düsseldorf (Germany), Institute of Molecular Evolution, University of Vienna (Austria), Vienna Doctoral School of Ecology and Evolution

Corrins are cobalt-containing tetrpyrroles that serve as cofactors for numerous methyl transfer and radical-dependent reactions. They are essential in the most ancient CO2-fixing biochemical route as cofactors for the corrinoid iron-sulfur protein, CoFeS, which catalyzes a cobalt-to-nickel methyl transfer in the acetyl-CoA pathway. CoFeS occurs in H2-dependent archaeal methanogens, perhaps the most ancient microbial lineage known, dating to 3.5-3.95 billion years ago, based on carbon isotope data, and in the acetyl-CoA pathway of H2-dependent bacterial acetogens. Was the corrin synthesis pathway transferred from the archaeal to the bacterial lineage during evolution, or was it present in the last universal common ancestor, and in either case, how did the primordial acetyl-CoA pathway function prior to the enzymatic synthesis of nature’s most complex cofactor? Here we analyze 26 enzymes for corrin biosynthesis in acetogens and methanogens, showing that phylogenies trace the pathway to their common ancestor. Prebiotically synthesized corrins are unlikely functional forerunners of CoFeS function, yet in laboratory simulations of serpentinizing hydrothermal systems, Fe-, Ni- and Co-containing solid state surfaces catalyze methyl transfer reactions of the acetyl-CoA pathway, presenting natural analogue of corrin function. Corrin biosynthesis was a massive evolutionary innovation, demanding the existence of a steep selective barrier. The data suggest that insolubility of its solid-state predecessor catalyzing essential metal-to-metal methyl transfers in a primordial hybrid acetyl-CoA pathway (part abiotic, part enzymatic) was that hurdle in early biochemical evolution.
Supercharged Protein Analysis in the era of AI

Martin Steinegger

Presented by self
Seoul National University (Korea)

Protein analysis has witnessed a revolution through machine-learning methods. At the forefront are highly accurate structure prediction methods such as AlphaFold2 and ESMFold. These have generated an avalanche of publicly available protein structures. The AlphaFold database and ESMatlas contain over 214 and 620 million predicted structures, respectively, covering nearly every protein sequence in our largest protein reference databases. This unprecedented access to structural information is not just critical for structural biology but impacts most fields of biology. In this talk, I will discuss how this data is revolutionizing genomic and proteomic annotations and introduce fast and sensitive methods to search and cluster this data to extract new biological insights.
Biosurveillance of zoonotic coronaviruses through an alignment-free physicochemical clustering approach
Rajeev K Azad

Jamie D Dixson, Rajeev K Azad
University of North Texas (USA)

To infer functional relatedness among betacoronaviruses, we developed an alignment-free remote homology detection technique, which was leveraged to gain new insights into coronavirus evolution, specifically in the context of the COVID-19 pandemic caused by novel coronavirus SARS-CoV-2. Novel methods for predicting the capacity for coronaviruses in general to infect human cells are needed. Our proposed method utilizes physicochemical properties of amino acids to develop a fully dynamic waveform representation of proteins that encodes both the amino acid content and the context of amino acids. These waveforms are then subjected to dynamic time warping and distance evaluation to develop a distance metric that is relatively less sensitive to variation in sequence length or primary amino acid composition. Using this method, we show that in contrast to alignment-based maximum likelihood and neighbor-joining phylogenetic analyses, all bat betacoronavirus spike protein receptor binding domains (RBDs) known to bind to the ACE2 receptor are found within a single physicochemical cluster. Further, other RBDs within that cluster are from pangolin coronaviruses, two of which have already been shown to bind to ACE2 while the others are suspected, yet unverified ACE2 binding domains. This finding is important because both SARS-CoV and SARS-CoV-2 use the host ACE2 receptor for cell entry. Surveillance for coronaviruses belonging to this cluster could potentially guide efforts to stifle or curtail potential and/or early zoonotic outbreaks with their associated deaths and financial devastation.
Characterizing the origin, evolution, and function of the synaptic protein PSD95 (DLG-4 gene) in single-celled relatives

Riya Nilkant

Riya Nilkant, Lisa Y Mesrop, Samuel Lobo, Anissa Morrison, Onur Sakarya, Soojin V Yi, Kenneth S Kosik

University of California Santa Barbara (USA)

The origin and evolution of complex synaptic machinery are facilitated by new protein-protein binding partnerships. Evolutionary changes in protein sequence and binding properties of PDZ domains can evolve new interactions that serve as primers for assembling complex protein modules with novel synaptic functions. Here we focus on the evolutionary origin and functional divergence of the tri-PDZ domain architecture of PSD-95, a key scaffolding protein in post-synaptic cells. We use phylogenetics, ancestral reconstruction, and protein structure predictions to determine the ancestral PDZ domain and the evolutionary changes that occurred in the binding interactions resulting in new binding partnerships. By comparing PDZ domains in unicellular relatives and Metazoans, we found the n-terminally positioned PDZ is ancestral to the first two PDZ domains and originated along the ancestral lineage leading to the Holozoan clade. Although the interaction of conserved GLGF residues determines the binding specificity of PDZ domains, we found that unicellular PDZ domains still bind in silico to endogenous synaptic-like ligands even in their absence. We observed the hydrophobicity through a molecular dynamics simulation to be similar across these unicellular PDZ domains and Metazoans despite the lack of residue conservation. We further found conservation of hydrogen bond interactions at positions 0 and -2 of PSD-95 ligands across humans, the unicellular relative C.owczarzaki, and certain predicted ancestors. We posit that the hydrogen bond interactions at specific ligand positions, and not conserved residue identities, are likely evolutionary drivers of new ligand partners for PSD-95.
The increasing availability of genomic sequences is driving forward our understanding of the diverse life forms on Earth. However, the ability to generalize organism-specific knowledge is limited by how well we relate genomes to each other. Genomes can be compared in terms of orthologous genes — genes of different species that derive from a single gene in the last common ancestor. Currently, the majority of orthology prediction software is based on the amino acid sequence, the most abundant information about proteins. However, the sequence signal of distantly related genes is weak, distributed in saturated positions, and confounded by evolutionary forces. In this work, I show how the more conserved protein 3D-structure can improve orthology prediction. I devised a method that combines sequence k-mers with k-mers generated from a local structural alphabet. The enriched sets of k-mers can be used to generate a reference classification of proteins into Hierarchical Orthologous Groups (HOGs) — a coarse-grained representation of protein families. The structure-informed reference HOGs, here named AlphaHOGs, can be exploited to infer orthology in a set of proteomes, by using the recently developed software FastOMA. As a test case, we reconstruct the ancestral genome of the first multicellular animal with an unprecedented resolution, paving the way to higher-level analyses such as the ancestral gene content and protein interaction network, potentially shedding light on the current uncertainties about the origin of the animal lineage.
Quasi-Ordered Recognition Dynamics in Sperm-Egg Interactions in Abalone

William T Higgins

William T Higgins, Damien B Wilburn
The Ohio State University (USA)

Protein evolution does not occur at a fixed rate over time, but rather as a dynamic process influenced by physiological and genetic factors. Some protein sites are subject to higher degrees of evolutionary conservation than others, such as catalytic residues or residues involved in forming protein-protein interactions. In the marine mollusk abalone, sperm lysin and egg VERL are rapidly coevolving proteins that mediate species-specific interactions during fertilization. The VERL binding interface on lysin consists of many rapidly evolving residues that are loosely packed in a “quasi-ordered” state that is primarily stabilized by the coordination an aromatic triad: W3, W62, Y133. This hydrophobic packing is observed in red abalone and closely related taxa, but Y133 is a derived characteristic compared to ancestral N133. This evolutionary transition coincides with a Q63A mutation adjacent to W62. The precise interactions and dynamics between these residues are not well understood. We employed molecular dynamics simulations using hydrogen mass repartitioning to evaluate the effects of A63Q and Y133N mutations in red abalone lysin to gain further insights about the protein dynamics on longer timescales. Preliminary analyses support that epistatic constraints have preserved this spatial coordination via these residues through alternative molecular interactions. Complementary solution NMR experiments to evaluate these trends are ongoing. Few quasi-ordered proteins have been biophysically characterized, and patterns inferred from lysin structural evolution to maintain species-specific fertilization may provide useful insights into other protein systems with complex molecular dynamics.
S2 - Novel approaches to study plant domestication: disentangling the complex evolutionary history of crops
Exploring the patterns of microbiota recruitment and inheritance on the chocolate tree, *Theobroma cacao* L

Alejandro Caro-Quintero

Presented by self

Tandem Max Plack Leader on Holobiont Research (UNAL), Universidad Nacional de Colombia Biology Department

The diversity and function of the microbiota associated with perennial tropical plants, such as *T. cacao*, remains a relatively unexplored area of research. It is becoming clear that these interactions have shaped the ecology and evolution of the host, and the role of microbiota in the development, growth, and health of hosts is becoming increasingly clear. Investigating whether this microbiota is affected by agriculture and domestication is crucial in understanding the complex relationship between plants and microbes. The microbiota might be recruited by roots from the surrounding environment or inherited from the maternal plant carried by the seeds. In the case of *T. cacao*, recruitment and inheritance processes likely influence the microbiome’s constitution. Here, I present some of our recent findings on exploring these mechanisms. First, under the CacaoBio program, we undertook an expedition to study the microbiome of *T. cacao* and its close relatives in their indigenous environments across the Amazon and Choco regions. Our study revealed critical drivers behind the microbial community structures across various tree tissues. A significant finding was the distinct microbial recruitment strategy observed in the rhizospheres of *T. cacao* plants from geographically isolated areas, setting them apart from other *Theobroma* species. Second, regarding seed-borne endophytes in cacao, we investigated the composition of the seed-borne endophytic microbial community associated with *T. cacao*. We explored the composition and diversity of seed-borne endophytes in cacao pods of commercial genotypes, recently liberated genotypes, and landraces to evaluate microbial vertical transmission and establishment in various tissues during plant development. We observed a higher abundance of *Pseudomonas* and *Pantoea* genera in the landraces and recently liberated genotypes. In contrast, the commercial genotypes presented many bacteria species but in low abundance. We isolated some of these seed-borne endophytes to evaluate their potential as growth promoters. We found that *Bacillus*, *Pantoea*, and *Pseudomonas* isolates presented high production of indole acetic acid and ACC deaminase activity. Our results suggest that cacao domestication could lead to the loss of essential bacteria for seedling establishment and development. Furthermore, the biotechnological potential of the cacao microbiota to enhance productivity and quality is being explored. Along this line, we have demonstrated improved seedling establishment and grafting success for regional cacao genotypes by harnessing seed-borne plant growth-promoting bacteria.
Meta-analysis of wild and domesticated crop phenotypic spaces

Arthur Wojcik, Stella Huynh, Ewen Burban, Karine Alix, Pierre Gérard, Amandine Cornille, Harry Belcram, Agnès Rousselet, Martine Le Guilloux, Rebecca Stevens, Marie Brault, Sylvain Glémin, Yves Vigouroux, Domenica Manicacci, Catherine Dogimont, Maud Tenaillon

EcoBio (France), GAFL (France), GQE - Le Moulon (France), IRD (France)

Plant domestication has led to a divergence between domestic (D) and wild (W) forms, at phenotypic and genomic levels. While the phenotypic domestication syndrome has been extensively documented, comparative analysis of phenotypic spaces between species have yet to be undertaken. The underlying pattern of phenotypical evolution have also to be linked to biological traits. Here, we explore the phenotypic spaces of 14 crop-wild pairs of diploid plant species with various mating systems and life-history traits, allowing comparative analyses. We measured traits on 20 W and 20 D populations in a randomized-block design. We performed variant calling on 40 W/D short-read sequenced genomes. We observed a marked disjunction of W/D phenotypic spaces in all systems. However, the strength of the phenotypic domestication syndrome depended neither on the genomic divergence nor the time to domestication. Second, the relative breadth of W/D phenotypic spaces followed a variety of trend - from smaller D than W in African rice to larger one in grape - that did not relate to mating system or life-history traits. The D/W genomic diversity ratio was <1 in all systems except for perennials, as chosen species did not suffer from domestication bottlenecks. It also does not correlate to the ratio of domestic to wild phenotypic space. Altogether, our results show uncorrelated patterns between genomic and phenotypic domestication syndromes. However, analysis of traits variance-covariance matrices reveals an uncoupling between W/D matrices as time to domestication increases, suggesting that phenotypic evolution has impacted genetic correlations.
Population genetic analysis reveals the domestication and dispersal history of common buckwheat

Jeffrey Fawcett

Jeffrey A Fawcett, Euki Yazaki, Yang Liu, Harriet V Hunt, Ryoma Takeshma, Mariko Ueno, Takanori Ohsako, Yumei Dong, Meifang Li, Hideki Hirakawa, Tatsuya Ota, Chengyun Li, Martin K Jones, Yasuo Yasui
Kazusa DNA Research Institute (Japan), Kyoto Prefectural University (Japan), Kyoto University (Japan), NARO (Japan), RIKEN iTHEMS (Japan), SOKENDAI (Japan), University of Cambridge (United Kingdom), University of Cambridge (United Kingdom), Yunnan Agricultural University (China)

Common buckwheat (Fagopyrum esculentum) is a rare example of a crop domesticated in southwest China, outside of major domestication centres, and spread widely across Eurasia likely over a thousand years ago. Here, we first sequenced and constructed a chromosome-scale reference genome of a selfing accession of common buckwheat. We then resequenced a few hundred accessions including ~50 of the wild progenitor species and performed various population genetic analyses. We found that a wild population from southeast Tibet is most closely related to all cultivated accessions and observed genetic differentiation between cultivated accessions from different geographic regions. We also identified a number of selective sweeps involved in the domestication and dispersal processes, including one strong selective sweep shared by all cultivated accessions which likely originated from standing variation within the wild population from southeast Tibet. Our results suggest that all cultivated common buckwheat accessions originated from a wild population in southeast Tibet and allow us to discuss the domestication and dispersal history of common buckwheat.
Fruit color variation in date palms (Phoenix dactylifera L.) and their wild relatives is determined by an ancient trans-specific polymorphism and parallel degradation of anthocyanin pathway genes

Jonathan M. Flowers

New York University Abu Dhabi (UAE)

Trait variation in crop species is often controlled by complex genetic and developmental networks and reconstructing their evolutionary origins is often complicated. Pigmentation traits have offered some of the most insightful examples into the origins of trait diversity and continue to yield unprecedented insight into genetic basis of evolutionary parallelism and convergence in crops and other organisms. Fruit color in cultivated date palms (Phoenix dactylifera L.) varies from red to yellow within species and similarly differs among its crop wild relatives. Here we describe the genetic basis and evolutionary origin of this trait across the Phoenix genus. We report an ancient origin of a previously reported causal LTR-retrotransposon insertion in an R2R3-MYB activator of the anthocyanin pathway. We estimate the age of the insertion and find it constitutes an ancient trans-specific polymorphism shared by divergent species. Fruit color variation between species appears to be caused both by this ancient insertion and parallel loss-of-function mutations in anthocyanin pathway genes. Our results suggest that both ancestral polymorphism and parallel evolution at minimally pleiotropic loci explain fruit color diversity in this genus of palms.
Exploring the fructan syndrome in Agave tequilana—a non-typical crop plant

June Simpson

Presented by self
Cinvestav, Genetic Engineering Department, Unidad Irapuato (Mexico)

Most plants store carbohydrates as starch or sucrose, however 15% of angiosperms store carbohydrates in the form of fructans. The fructan syndrome refers to species which carry out fructan metabolism by employing specific enzymes for the synthesis and degradation of these polymers. The enzymes responsible for fructan metabolism are closely related to and probably evolved from vacuolar and cell wall invertases. In addition to their function as storage carbohydrates, fructans are also important in abiotic stress tolerance mechanisms due to their capacity to aid in the protection of cell membranes under conditions of drought or frost. Roles for fructans in maintaining the osmotic balance of cells and as signaling molecules under biotic stress have also been proposed. Chicory, onion and cereal crops such as wheat, oats and barley are fructan syndrome species although the types of fructan polymers and their presence in different organs and at different developmental stages varies widely. Specific genotypes have probably been selected for the capacity to withstand cold stress at higher latitudes or in some cases to survive overwintering, traits which are related to fructan metabolism. Our model of study is Agave tequilana which has the capacity to synthesize and store the large quantities of fructans that provide the raw material for the tequila industry and probably play a role in the adaptation of these species to extreme arid conditions. In comparison to other agave species used to produce alcoholic beverages, A. tequilana is the most efficient in accumulating high levels of fructans within a shorter life cycle leading to the commercial and exclusive propagation of this genotype for tequila production. We are currently studying the genes and related enzymes involved in fructan metabolism in A. tequilana to gain a deeper understanding of the roles of fructan metabolism in adaptation to stress which will be relevant for non-fructan syndrome species in the current scenario of climate change.
Population genetics of year-long bean (Phaseolus dumosus Macfady, Fabaceae) in Southern Mexico and Guatemala
Katya Garduño Obrajero

Katya Garduño, Azalea Guerra García
Cinvestav Unidad Irapuato (Mexico), UNAM (Mexico)

One of the crucial processes that allowed humans to transition from nomadic to sedentary lifestyles was the development of agriculture. This process began simultaneously around 11,000 years ago during the Neolithic period in different parts of the world and Mesoamerica is one of the centers of domestication. Among the plant species domesticated in this region, are five bean species, which are an important component of the diet in Mexico and are considered one of the seven basic and strategic crops for the country. The year–long bean (Phaseolus dumosus) is a widespread crop in the Southeast regions of Mexico and Central America. Its wild populations only occur in the highlands of Guatemala, where domestication presumably started. In this study, we aim to understand the population structure, diversity, and genetic differentiation of 8 populations of P. dumosus (six cultivated and two wild populations) by using genomic information from 43 individuals. After filtering, 19,321 SNPs were kept and 4 genetic groups were identified: three cultivated genetic clusters (Veracruz, Chiapas and Guatemala), and one wild cluster (Guatemala highlands). The highest genetic diversity was found in the cultivated population from Guatemala, meanwhile, the lowest was detected in the cultivars from the Trans-Mexican Volcanic Belt. The genetic differentiation was unexpectedly high when comparing the cultivars from Guatemala to the rest of the populations, including the wild one. This constitutes the first study at a population level that assesses the genetic diversity of the orphan crop bean species.
Wheat domestication and adaptation to new environments: the role of wild emmer.
Laura Botigué

Laura Botigué, Alice Iob, Nathan Wales, Ferran Antolín, Jana Helmbold-Doye, Marta Ortigas
Centre for Research in Agricultural Genomics (Spain), Egyptian Museum of Berlin (Germany), German Institute of Archaeology (Germany), NA (Spain), University of York (United Kingdom)

Wild emmer wheat (Triticum turgidum subsp. dicccoides) stands among the earliest plant species to give rise to domesticated forms in the Fertile Crescent, approximately 11,000 years ago. Today, one of its descendants, bread wheat, is a staple crop of immense socio-economic importance, contributing to approximately 20% of the world’s caloric intake. The domestication of wheat has long captivated interdisciplinary scholars, and the prevailing theory suggests a protracted and geographically diffused process spanning millennia. In this study, we delve into the population structure of wild emmer populations and their impact on the domestication process and subsequent dispersal of the first domestics. Our investigation involves quantifying, for the first time, the genomic proportion of ancestry from wild emmer populations in two distinct germplasms, encompassing landraces from Europe, Africa, and Asia. We also provide estimates of the time since admixture of these two wild populations, aligning with the domestication and dispersal patterns observed, particularly towards Africa. Furthermore, we explore the potential adaptive role of wild emmer from the Southern Levant in domestication and dispersal, revealing an enrichment of genes associated with resistance to biotic stress and drought. Finally, we analyze the exome of 2,000 year-old wheat specimens from Egypt to trace the presence of these putative adaptive haplotypes. Our findings deepen our understanding of the origins of domestic wheat and underscore the significance of modern domestic landraces of emmer wheat in elucidating the genetic basis of resilience in cultivated crops.
Composition of microbiota of floral nectar in wild and domesticated squash (Cucurbita)
Luis Alberto Villanueva-Espino

Luis Alberto Villanueva-Espino, Irais Avila Eulogio, Violeta Patiño Conde, Rafael Lira Saade, Eric Fuchs Castillo, Mauricio Quesada Avendaño
Campus Morelia (Mexico), Escuela Nacional de Estudios Superiores, Facultad de Estudios Superiores Iztacala (Mexico), Instituto de investigaciones en ecosistemas y sustentabilidad (Mexico), UNAM, Universidad de Costa Rica (Costa Rica)

Floral nectar is the most common floral reward in plant-pollinator interactions. Due to its high energy and nutritional content, it is also a favorable habitat for the growth of many microorganisms, mainly yeasts and bacteria. These microbes can significantly modify the chemical composition of nectar, affecting pollinator behavior and plant reproductive success. The effects of this tripartite relationship (plant-pollinator-microbiota) in the Cucurbita genus are addressed to understand domestication’s impact on this triangle. Using a metabarcoding approach, we determine the composition of bacteria and fungi in the floral nectar of six species of plants of the genus Cucurbita, three wild and three domesticated. We compared the differences between species, the floral sexes, and the effect of pollinator visitation on the diversity of the microbiota of these squash species. In general, we observed a higher richness of bacteria than fungi in the floral nectar microbiota of the studied species of Cucurbita. For bacteria, alpha diversity shows differences between species but not in floral sexes, visited vs. not visited flowers, and wild and domesticated species. For fungi, however, this parameter does not show any differences. The beta diversity for both groups indicates differences between species, floral sexes, and visited vs. not visited flowers, but not by wild and domesticated species. This study shows that domestication does not affect the microbial community in floral nectar. However, it provides evidence of how the microbiota varies based on other factors, including species, floral sex, and pollinator visitation.
Multipurpose species domestication in Mayan homegardens
Miriam Monserrat Ferrer

Miriam M Ferrer, Hector Estrada-Medina, Juan J Jiménez-Osorno, Patricia Irene Montañez-Escalante, Rocío Ruenes-Morales
Universidad Autónoma de Yucatán (Mexico)

During domestication, plants are expected to show a reduction in genetic diversity and in the microbiota associated with them in the agroecosystems in which they are managed, compared to that of forest populations. However, initial evaluations of genetic variability in Brosimum alicastrum (ox in Maya), Cordia dodecandra (k'oopte' in Maya), Selenicereus undatus (woob in Maya), and Spondias purpurea (abal in Maya) in homegardens in Yucatan suggest that the species have similar genetic diversity to sympatric forest populations due to gene flow among wild and managed individuals. We have also found that the most abundant bacterial and fungal endophytes of leaves in C. dodecandra are similar in individuals from homegardens and forests, even when they are grown in distant regions. These results suggest that native species have in homegardens an agroecosystem that functions as a hologenomic reservoir. On the other hand, we observed different cytotypes in B. alicastrum individuals from homegarden and forest and a ploidy series in S. purpurea (2n, 3n and 4n) varieties growing in homegardens. Also, in soils and rhizosphere associated with C. dodecandra, bacterial and fungal communities have a higher turnover, which may be associated with specific management practices of each agroecosystem. We believe that omics studies can help us to understand how management by the Mayan people promotes agrobiodiversity by maintaining the microbiome and increasing genomic variability of species with biocultural importance.
Human and ant-domesticated plants harbour different disease resistance gene repertoires compared to their wild relatives

Noah Bourne

Noah Guy Bourne, Guillaume Chomicki
Durham University (United Kingdom), University of Sheffield (United Kingdom)

Plant domestication is a pivotal innovation driving the success of human civilizations; however, the domestication process often includes unintended phenotypic and or genomic consequences for the plants involved. These include susceptibility to diseases in which their wild relatives are resistant. Convergently ant-domesticated epiphytes are susceptible to disease upon ant-exclusion compared to their non-farmed (wild analogues) relatives. The genomic basis of this apparent discrepancy is linked to domesticated plants having differing resistance gene (R-gene) repertoires to their respective wild relatives. Although due to the lack of consistent annotations in previous studies it is unclear whether domestication repeatedly leads to reduced R-gene repertoires and thus unequal disease resistance capability across the phylogeny of domesticated plants. Pan-genome approaches highlight that many R-genes in plants exist in the auxiliary portion of the genome with their diversity being maintained in a population by strong balancing selection which is often negated during domestication. This provides an evolutionary explanation to why reduced R-gene repertoires persist in crop plants. Using high quality genomes of crop plants and their wild relatives, including ant-domesticated Squamellaria and their non-farmed relatives, we use state of the art deep learning approaches to provide consistent annotations of R-gene repertoires. Among the 15 crop/wild relative pairs covering 10 plant families we see significant differences in the repertoire of specific R-gene types. Additionally, the inclusion of ant-domesticated Squamellaria provides an exceptionally unique insight into the evolution of domesticated plants over a larger timescales.
The post-domestication history of rice

Ornob Alam

Ornob Alam, Michael Purugganan, Rafal Gutaker
Kew (United Kingdom), New York University (USA), Royal Botanic Gardens

The dispersal of rice (Oryza sativa) following domestication influenced massive social and cultural changes across South, East, and Southeast (SE) Asia. While recent resequencing of traditional rice landraces from across Asia have revealed possible routes of dispersal across Southeast Asia, the early history and spread of rice across and out of China remain largely unresolved. This is mainly because of sparse sampling of traditional landraces of japonica rice — the ancestral rice domesticate — from China as these have largely been replaced by modern improved or hybrid varieties and indica rice. We addressed this dearth of sampling by collecting historical specimens of domesticated rice from herbaria in France, the UK, and the US, originally sampled from China and other parts of Asia by botanists in the 19th and early 20th century. Here, we performed whole-genome resequencing of 96 herbarium rice specimens and analyzed them alongside a previously published set of 367 japonica rice landraces. We used ancestral recombination graph-based approaches to reconstruct the demography and routes of dispersal of japonica rice following domestication in China. We also utilized ancestral recombination graphs to investigate the evolutionary history of adaptive and functional variants — including those involved in flowering time, immunity, and taste— in populations of japonica rice, revealing widespread local adaptation as rice spread to different regions, climates, and latitudes.
Domestication often involved radical morphological changes as plants adapted to novel anthropogenic environments. Understanding the origin of genetic variants underlying these changes has long been of interest. Despite the dramatic morphological differences between maize and its wild relative teosinte, research on a handful of well-described genes has found that maize domestication relied on standing genetic variation in ancient teosinte. But, researchers had concluded that one key trait and the genetic variation that causes it – exposed kernels and a mutation in tga1 – could exist only in cultivated populations. If true, the causative mutation in tga1 must have appeared de novo in cultivated populations. However, the puzzling absence of tga1 in many genome-wide scans for selection, typically well-powered for strong selection on de novo mutations, prompted us to consider that the causative mutation in tga1 actually predates domestication. Leveraging a large dataset of maize and teosinte genomes, we first identified the maize allele of tga1 in modern teosinte, suggesting that the purportedly deleterious allele can persist in the wild as standing genetic variation. Second, we conducted population genetic simulations and found that haplotype diversity patterns and distribution of mutational ages near tga1 can be best explained by selection on standing variation. Finally, we estimated the mutational age of the causative allele in tga1, finding that it predates even conservatively early estimates for the beginning of maize domestication. We conclude that, in contrast to several decades of thought, maize domestication drew upon standing genetic variation in tga1.
Tracing Domestication History in Potato: Admixture and Ancestry Inference in a Polyploid Species

Sergio Tusso

Sergio Tusso, Korbinian Schneeberger
LMU (Germany)

Understanding the history of crop domestication is fundamental to agricultural advancement and biodiversity conservation. In this project, we delve into the complex and unclear domestication history of potato in the Andean region of South America. The domestication of potato likely involved admixture from multiple species spanning the geographic range from Mexico to Argentina, during initial domestication 10000 years ago and subsequent plant breeding programs in more recent times. However, given the tetraploid nature of potato, traditional methods face significant challenges in the genomic inference of ancestral admixture. To address these challenges, we present a novel approach utilizing an extended dataset comprising resequencing data from 840 samples, including globally distributed cultivars and 84 wild and landrace species. Leveraging linkage information, we identify ancestral groups within cultivar samples, revealing clear introgressed blocks from multiple species and variation along the genome, indicative of a complex evolutionary history. Surprisingly, despite the evident admixture from multiple species, we also observe markedly reduced haplotype diversity, which has important implications for potato agricultural research and biodiversity conservation efforts. By demonstrating the application of linkage disequilibrium for ancestry inference, our study not only provides novel insights into the intricate history of potato, but also presents a potential alternative to for admixture inference of other polyploid species.
A uniform bacterial community inhabits the rhizosphere across cultivars of the common bean (Phaseolus vulgaris).

Víctor González

Victor González, Griselda López-Romo, Rosa I Santamaría, Patricia Bustos, Francisco Echeverría, Luis Roberto Reveles-Torres
Centro de Ciencias Genómicas (Mexico), Centro de Ciencias Genómicas-UNAM (Mexico), INIFAP (Mexico)

The success of common bean (Phaseolus vulgaris) crops relies, among other factors, on the microbial species inhabiting the rhizosphere. Although still inadequately characterized, beyond the established role of nitrogen-fixing bacteria, current metagenomic techniques allow us to catalog a large number of microorganisms hosted in the rhizosphere. Here, we studied the bacterial communities in the rhizosphere of nine common bean cultivars used in agriculture. Bulk soil and rhizosphere samples from P. vulgaris cultivars were collected in situ from plots with and without a cultivation history. In both plots, bacterial diversity in the bulk soil exceeded that in the rhizosphere. Diversity and taxonomic composition analyses confirmed the dominance of Proteobacteria in the rhizosphere. Furthermore, a uniformly distributed bacterial community encompassing various cultivars was present in the rhizosphere. The results of comparing bulk soil-rhizosphere metagenomes of various cultivated plants, including maize, wheat, tomato, cucumber, Arabidopsis, and common bean cultivars, showed a distinct rhizosphere effect for the latter. Specifically, common bean cultivars exhibited a different pattern in the presence of bacterial genera known to promote plant growth, such as Rhizobium, compared to the other plants studied. These findings suggest that the rhizosphere effect of common bean cultivars may be unique and deserve further investigation. Metagenome-assembled genomes (MAGs) reconstructed from metagenomes confirmed the presence of diverse biosynthetic gene clusters that are involved in defense, signaling, and plant growth promotion. The results underscore the significance of the prior soil microbiome for the subsequent cultivars in facilitating bacterial colonization of the rhizosphere.
Origin and diversity of the wild cottons (Gossypium hirsutum) of Mound Key, Florida

Weixuan Ning

Weixuan Ning, Karen M Rogers, Chuanyu Hsu, Zenaida V Magbanua, Olga Pechanova, Mark A Arick II, Ehsan Kayal, Guanjing Hu, Daniel G Peterson, Joshua A Udall, Corrinne E Grover, Jonathan F Wendel
Agricultural Genomics Institute at Shenzhen (China), Chinese Academy of Agricultural Sciences (Chad), Florida Department of Environmental Protection (USA), Iowa State University (USA), Mississippi State University (USA), USDA/Agricultural Research Service (USA)

Deciphering genetic diversity within wild forms of modern crops is essential for understanding domestication and the possibilities of wild germplasm utilization. Gossypium hirsutum is a predominant source of natural plant fibers and the most widely cultivated cotton species. Wild forms of G. hirsutum are challenging to distinguish from feral derivatives, and truly wild populations are uncommon. Here we characterize a population from Mound Key Archaeological State Park, Florida using genome-wide SNPs extracted from 25 individuals over three sites. Our results reveal that this population is genetically dissimilar from other known wild, landrace, and domesticated cottons, and likely represents a pocket of previously unrecognized wild genetic diversity. Furthermore, our results suggest this population may be truly wild, refuting the notion that it could be a naturalized derivative of cultivars used by the Calusa Native Americans, the former inhabitants of Mound Key before European colonization. The unexpected level of divergence between the Mound Key population and other wild cottons suggests that the species may harbor other relictual and genetically distinct populations that are geographically scattered in suitable habitats throughout the Caribbean. Our work thus has broader conservation genetic implications and suggests that further exploration of natural diversity in this species is warranted.
Deciphering Sorghum Domestication: The Pivotal Role of Condensed Tannins in Co-evolution Among Plants, Humans, and Birds

Xianran Li

USDA-ARS WHGQ (USA)

Among major cereal crops, sorghum is the only one with a high proportion of cultivars containing condensed tannins, potent secondary metabolites inducing bitterness perception, in seeds. This presentation explores a case of co-evolution among plants (domesticate), humans (domesticator), and birds (biotic environmental conditions) to unravel this evolutionary mystery. An examination of the geographic distribution of 11,557 sorghum varieties revealed that tannin sorghum are predominantly found in East and South Africa, while non-tannin sorghum predominates in West Africa. Analysis of functional polymorphisms of two key genes suggested that smallholders in different African regions independently domesticated non-tannin sorghum. Further investigation unveiled a parallel geographic distribution between sorghum and the pest Quelea quelea in Africa. Q. quelea, the world’s most abundant wild birds, mainly inhabits East and South Africa, but not West Africa. Smallholders in East and South Africa selected tannins to combat severe herbivore threats. To address how African smallholders could embrace "bad taste" tannin sorghum as a staple food, a parallel geographic distribution between bitter taste receptor (TAS2Rs) alleles in humans and sorghum was uncovered. Ancestral non-taste alleles are predominantly observed in East and South Africa, while derived taste alleles sensitive to tannins mainly originate from West Africa. Integration of these coevolutionary relationships suggested that in East and South Africa, smallholders (perceiving less bitterness due to tolerant TAS2Rs) selected tannins to combat agricultural pests, while in West Africa with mild threats from Q. quelea, smallholders (carrying sensitive TAS2Rs) selected non-tannin sorghum for better food.
Unraveling the population structure and origins of pigmented rice

Yoon Kyung Lee

Yoon Kyung Lee, Nese Sreenivasulu, Rafal M. Gutaker
International Rice Research Institute (Philippines), Kew (United Kingdom), Royal Botanic Gardens

Unpolished rice (Oryza sativa L.) provides nutritional benefits such as dietary fiber, essential vitamins, and minerals. The coloration in the rice pericarp, caused by the accumulation of anthocyanins and proanthocyanidins, not only adds to the nutritional value, but also has cultural significance. Aside from the rice domestication gene Rc, the mutation which led to the emergence of white rice, the subsequent origins of pigmented rices with various range of hues from red to purple, their evolutionary history and distribution is underexplored. To shed light on this, we conducted a comprehensive population structure analysis using genomic data from a diverse panel of 360 pigmented and unpigmented rice varieties. The study determines whether pigmented rice varieties share a common ancestry or if they have evolved independently in different geographic regions of Asia. To achieve this, Principal Component Analysis projection was employed, over the panel composed of three representative subgroups, including indica, tropical japonica, and subtropical japonica varieties. Further studies on the genetic diversity among different subpopulations may indicate adaptation to different environments. Our findings are expected to provide insights into the intricate relationship among genetic diversity, environmental factors, and human selection pressures. These insights will inform future breeding strategies aimed at achieving desirable nutritional values.
Understanding the genetic basis of eating quality traits in rice (Oryza sativa L.) is critical for breeding programs aimed at improving consumer satisfaction. In this study, we investigated the genetic diversity and selection signatures associated with eating quality traits in a diverse panel of 284 rice accessions from South Korea, Japan, China, Taiwan, and USA, comprising 171 improved varieties and 113 landraces. Population structure analysis revealed that most accessions from this panel belonged to the temperate japonica subgroup. Further subdivision within this population revealed three distinct clusters: landraces, improved cultivars, and recently bred improved cultivars. Phenotypic evaluation was conducted to characterize the eating quality trait by measuring the glossiness of cooked rice using the Toyo taste meter. Subsequently, a selection scan for genetic determinants associated with eating quality traits was performed using XP-CLR, focusing on 25 accessions exhibiting high eating quality and 25 with low eating quality. The study enhances our understanding of the genetic architecture of grain quality traits in japonica rice. It also highlights the complexities and nuances encapsulated within varietal diversity. The study identifies genomic regions with molecular signatures for selection associated with superior eating quality, which can inform targeted breeding strategies to enhance these attributes.
Unraveling the evolutionary history of indica rice post domestication
Yu-Chen Lin
Yu-Chen Lin, Michael Purugganan
New York University (USA)

Asian rice comprises two primary subspecies: japonica and indica, and it is cultivated across a broad latitudinal range and diverse local ecosystems. Archaeological and population genomic evidence indicates that indica rice underwent domestication in the Ganges Valley approximately four millennia ago, potentially spreading to Southeast Asia and China through introgression with japonica rice. A previous study revealed that the dispersal patterns of indica rice remain puzzling, largely due to widespread gene flows and a weak population structure. In this study, we employed an extensive sampling of indica rice landraces to scrutinize the population structure and demographic history of indica rice post-domestication. Our preliminary findings highlight a robust population structure linked to geographical origins, with admixture analysis suggesting two admixture events. Inferring selection and allele frequency trajectories through genomic data of the geographic subpopulations promises to offer insights into the evolutionary history of indica rice.
S3 - Progress and challenges for understanding the molecular evolution of sex chromosomes across Eukarya
Evolution of sex-biased gene expression (SBGE) during transitions to separate sexes in the Silene genus

Aline Muyle

Djivan Prentout, Aline Muyle, Niklaus Zemp, Adil el Filali, Bastien Boussau, Pascal Touzet, Alex Widmer, Jos Käfer, Gabriel Marais
34 EPHE, 4485-661 (Portugal), 69622 (France), CIBIO, CNRS, Campus de Vairão, Centre d’Ecologie Fonctionnelle et Evolutive (CEFE), Centro de Investigação em Biodiversidade e Recursos Genéticos, Columbia University, Department of Biological Sciences, ETH (Switzerland), ETH Zurich (Switzerland), F-59000 (France), Genetic Diversity Centre (GDC), IRD (France), ISEM, InBIO, Institute of Integrative Biology, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, Laboratório Associado, UMR 8198 – Evo-Eco-Paleo, USA (USA), Univ Montpellier, Univ. Lille, Universidade do Porto, University of Montpellier, Université Lyon 1, Université de Lyon

Sexual dimorphism is widespread among species with separate sexes and its extent is thought to be governed by the differential expression of thousands of genes between males and females (known as Sex-Biased Genes, hereafter SBGs). SBGs have been studied in numerous species, but rarely in a comparative way, which curtails our understanding of their evolution, especially during multiple independent transitions to separate sexes. We sequenced the transcriptomes of nine dioecious species, two gynodioecious species (separate females and hermaphrodites) and two hermaphrodite species from the Silene genus. Our dataset provides access to three independent transitions to dioecy (dating from less than 1 Myo to about 11 Myo). We demonstrated that male-biased expression emerges first during a transition to separate sexes, later followed by female-biased genes. Furthermore, we showed that, despite a mixture of selective regimes, positive selection significantly affects the evolution of some SBGs. Overall, this study provides new insights on the causes of SBG evolution during transitions to separate sexes.
Do Z chromosomes in Butterflies and Moths Show Evidence of Increased Rates of Adaptation?

Claudia C Weber

Tree of Life, Wellcome Sanger Institute (United Kingdom)

Chromosomes present in a single copy in one sex often have unusual patterns of sequence evolution, since the effects of recessive mutations are not masked. Therefore, substitution rates may be elevated compared to the autosomes. This observation has prompted speculation that tortricid moths, which are agricultural pests, are evolutionarily successful due to enhanced adaptation of newly Z-linked detoxifying enzymes (without directly examining the implicated proteins). However, both “fast” and “slow” Z evolution have been reported in lepidopterans, depending on the species and genes examined. Previous analyses were limited to small sets of species, and it remains unclear if positive selection on the Z chromosome is common across the taxon. Using hundreds of lepidopteran genomes from the Darwin Tree of Life project, we can revisit the evolution of the Z chromosome and ask: Is there evidence for accelerated evolution of Z-linked sequences across lepidopterans? Does the pattern differ for recently translocated sequences? Drawing on a catalogue of genomic rearrangements, we can also consider if newly Z-linked genes evolve faster than comparable sequences that remain autosomal. The dataset also provides an opportunity to examine if any observed acceleration is consistent with adaptation rather than drift. In addition to identifying fast-evolving proteins, pinpointing rapidly diverging sites and examining the nature of the changes will be informative. Past work on mammals reports that positively selected sites in xenobiotic-metabolising enzymes cluster in functionally important regions. Considering the structural contexts of substitutions might, therefore, provide clues about their functional consequences.
Despite significant progress made towards defining the relationships of placental mammals (Eutheria), some branches deep within the phylogeny remain controversial. This discordance is largely driven by several biological factors — including incomplete lineage sorting and gene flow — which make the elucidation of the true species tree challenging. Within Eukarya, meiotic recombination considerably influences the pattern and distribution of phylogenomic signals across the genome. Regions with historically low recombination rates are often enriched for the species tree. The majority of placental mammals possess a XX/XY sex chromosome system which are extensively differentiated in structure and sequence — with only the pseudoautosomal regions still recombining. Therefore, the male-specific portion of the Y (MSY) lacks recombination, has elevated rates of mutation, and is subject to background selection. Yet, the Y has rarely been used to address deep-level phylogenetic questions due to sequencing, assembly, and alignment difficulties across diverse clades. To expand our knowledge of mammalian Y-chromosome evolution and function, we present the first comprehensive phylogenetic analysis of placental mammal Y chromosomes across a diverse sampling of ordinal clades using a hierarchical alignment approach from whole genome sequence assemblies. We present a Y chromosome phylogeny for Eutheria and compare it to phylogenies constructed using the X and autosomes. Moreover, we discuss the utility of sex chromosomes in delineating controversial phylogenetic relationships and their molecular signatures of selection based on patterns of MSY gene loss across different lineages.
Primates exhibit a wide range of population sizes, mating systems, dispersal patterns, and demographic histories, each expected to change the relative genetic diversity of X chromosomes and autosomes (X/A ratio). Interrogating this relative diversity across 120 species from lemurs to great apes reveals that the typical X/A ratio is closer to 1/2 than the ¾ predicted in neutral randomly mating populations of constant size, with a mean of 0.62 across all species, with some exhibiting a ratio lower than 0.5. We remove the species-specific effects of demography and sex-biased admixture by comparing observed X/A ratios to the values expected from their effective population size trajectories (SMC++). Still, the X/A ratios remain depressed suggesting that linked selection on the X chromosome, in the form of selective sweeps, may play a stronger role than on autosomes. Such sweeps would affect species individually, producing larger differences in heterozygosity between species across genomic regions. Consistent with this prediction, we find that the coefficient of variation on pi is not only larger on X but that its relative size correlates with the X/A ratio. Lastly, observed mating patterns and testis size have a significant effect on the X/A ratio, showing that differences in reproductive success across genders strongly affect the relative diversity on chromosome X.
The Silene latifolia genome and its giant Y chromosome

Gabriel Marais

Presented by self
CIBIO (Portugal), CNRS (France), Laboratoire Biométrie et Biologie Evolutive (LBBE)

In some species, the Y is a tiny chromosome but the dioecious plant Silene latifolia has a giant ~550 Mb Y chromosome, which has remained unsequenced so far. Here we used a hybrid approach to obtain a high-quality male S. latifolia genome. Using mutants for sexual phenotype, we identified candidate sex-determining genes on the Y. Comparative analysis of the sex chromosomes with outgroups showed the Y is surprisingly rearranged and degenerated for a ~11 MY-old system. Recombination suppression between X and Y extended in a stepwise process, and triggered a massive accumulation of repeats on the Y, as well as in the non-recombinating pericentromeric region of the X, leading to giant sex chromosomes.
Cephalopod sex determination and its ancient evolutionary origin revealed by chromosome-level assembly of the California two-spot octopus

Gabrielle Coffing

Gabrielle Coffing, Silas Tittes, Scott Small, Jeremea Songco-Casey, Judit Pungor, Denise Piscopo, Adam Miller, Cris Niell, Andrew Kern
University of Oregon (USA)

The coleoid cephalopods comprise a remarkable branch of the tree of life, possessing a distinct set of novel traits. With the recent proliferation of genome sequencing technologies, we have the capability to explore the genomic foundations of these distinctive characteristics. In 2015, the California two-spot octopus (Octopus bimaculoides) became the first cephalopod to be sequenced, and since then, numerous additional reference genomes for various members of this group have been generated. Here, using PacBio long-read sequencing of genomic DNA and Iso-seq sequencing of full-length mRNA transcripts, we provide a novel chromosome-scale reference genome and annotation for O. bimaculoides. We use our assembly in combination with existing cephalopod genomes to create the first whole genome alignments from this group and characterize the landscape of sequence conservation in octopus. Notably, our assembly reveals evidence of a heterogametic chromosome in females, consistent with a ZZ/ZO male/female system—the first evidence of genetic sex determination in cephalopods. We compared this heterogametic chromosome to other cephalopods and our data support a single origin of this sex chromosome on the lineage leading to coleoid cephalopods. The Z chromosome is an evolutionary outlier with respect to divergence and repetitive element patterns. Our results suggest that the coleoid cephalopod Z chromosome originated between 455 and 248 million years ago, placing it among the oldest conserved animal sex chromosomes known.
The UV sex chromosomes of brown algae act as genomic cradles for new genes that evolve de novo
Josué Barrera-Redondo

Josué Barrera-Redondo, Jia Xuan Leong, Christos Papadopoulos, Cátia Igreja, Rémy Luthringer, Fabian B. Haas, M. Mar Albà, Susana M. Coelho
Hospital del Mar Medical Research Institute (Spain), Max Planck Institute for Biology (Germany)

Despite the vast diversity of sex-determination systems in nature, there are only three known types of sex-chromosome systems: the XY, the ZW, and the largely unexplored UV systems. Although these three systems share several core features, there are important differences between them that have broad evolutionary and genomic implications. Brown algae (Phaeophyceae) are complex multicellular eukaryotes that harbor an ancestral UV sex-determination system, but have also evolved a derived XY system, as well as several transitions to hermaphroditism. Thus, brown algae represent exceptional models for investigating the evolutionary dynamics of sex chromosomes. We show that the UV sex chromosomes in brown algae are at least 450 million years old. When we calculate the relative age of each gene in the brown algal genomes, we observe that the UV sex chromosomes are consistently enriched in younger genes compared to the autosomes, particularly in the pseudoautosomal regions. This pattern is accompanied by higher substitution rates, more repressive chromatin marks, and higher TE content. However, the brown algae that evolved an XY system or transitioned to a hermaphroditic lifestyle lost this pattern, suggesting that young gene enrichment is a feature of the UV system. Through a combination of synteny analyses, RNAseq, and riboseq data, we tested whether new genes are arising in the UV sex chromosomes through de novo gene birth processes (i.e., the birth of new protein-coding genes from noncoding DNA). We propose that the UV sex chromosomes act as genomic ‘cradles’, fostering the birth of new genes.
Evolutionary Insights into Ape Sex Chromosomes from Telomere-To-Telomere (T2T) Genome Assemblies: Palindromes and Multi-Copy Genes.

Karol Pal

Karol Pal, Robert Harris, Monika Cechova, Huiqing Zeng, Brandon D Picket, Prajna N Hebbar, Sergei K Pond, Kateryna D Makova, T2T Primates Consortium

Department of Biology, Department of Biomolecular Engineering, Department of Genome Sciences, Genomics Institute, National Human Genome Research Institute Home, National Institutes of Health (USA), Santa Cruz (USA), Temple University (USA), The Pennsylvania State University (USA), University of California, University of Washington (USA), gineering

The Telomere-To-Telomere (T2T) Primates Consortium is building complete genome assemblies for multiple primate species, leveraging methods recently developed for constructing gapless assemblies of the human genome. Here, we, for the first time, assembled T2T sequences of the X and Y chromosomes of five great apes and one lesser ape—all endangered species. These assemblies have enabled the first comprehensive analysis of palindromes and multi-copy genes evolution, the most challenging regions but also the hallmarks of ape sex chromosomes. Palindromes—long inverted repeats undergoing intrachromosomal recombination and gene conversion between their arms—have evolved on the Y to counteract the accumulation of mutations due to lack of recombination (outside of the pseudoautosomal regions), but they are also present on the X. We found Y palindromes to be 2–3 times longer and constitute a higher proportion of the chromosome as compared to X palindromes. Y palindromes were frequently species-specific, whereas X palindromes were largely shared among species. Interestingly, X palindromes had high gene density and harbored housekeeping genes. In contrast, Y chromosome palindromes carried multi-copy gene families with lineage-specific expansions (CDY in Pongo, and RBMY in Pan—both gene families involved in spermatogenesis). Besides confirming previously described multi-copy genes, we identified new ones, and searched for their autosomal homologs. We discuss evolution of these genes in light of our analyses of their selection and copy number variation. The potential role of these genes in determining mating patterns and levels of sperm competition should be evaluated in further functional studies.
Hidden sex-chromosomes in Middle-Earth - Revealing sex-chromosome diversity within the geckos of Aotearoa New Zealand

Ludovic Dutoit

Phoebe Webster, Sarah Walters, Alison Cree, Jo Monks, Tony Gamble, Mary Morgan-Richards, Samantha King, Tania King, Ludovic Dutoit
Marquette University (USA), Massey University (New Zealand), University of Otago (New Zealand), Wildlands (New Zealand)

Sex-determination systems are of critical importance to a range of fundamental evolutionary processes. The dynamics of speciation, local adaptation, and genetic conflict are all affected by sex chromosome systems. Yet, testing hypotheses empirically is challenging because of the typical lack of diversity in sex-determination systems over short evolutionary time scales. Within vertebrates, geckos exhibit an extraordinary diversity of sex-determination systems. Male and female heterogamety have evolved several times from an ancestral temperature sex-determination system that can still be found in a few species today. Aotearoa New Zealand is often considered a land of odd birds, but it is truly a land of wonderful lizards, inhabited by at least 48 gecko species in a monophyletic radiation. We will present work highlighting several sex-chromosome transitions within New Zealand geckos alone, suggesting that these species present an attractive model system to study extremely young sex-chromosome systems in vertebrates.
Alignment and Variant Calling in Reference Genomes to Improve Sex Chromosome Comparative Genomics
Melissa Wilson

Presented by self
Arizona State University (USA)

Mammalian sex chromosomes have a shared evolutionary history, as they were once a pair of homologous autosomes. Though X and Y are differentiated, they still share high levels of sequence similarity in some regions, like the pseudoautosomal regions (PARs) and the X-transposed region (XTR). Females typically have two X chromosomes and males typically have an X and a Y chromosome. Despite the importance genes on the X and Y chromosomes for comparative genomics and evolutionary history, there is limited understanding on how taking a standard approach for alignment and variant calling - one that is designed for the autosomes/diploid chromosomes - affects variant calling on the sex chromosomes. Here, we undertook a simulation study to assess the effects of standard autosomal versus sex chromosome complement-informed alignment, variant calling and variant filtering strategies on variants called on sex chromosomes. We find that aligning samples to a reference genome informed by the sex chromosome complement of the sample - either the entire Y hard masked when aligning genetic females (46, XX) or hard masking one copy of the PARs when aligning genetic males (46, XY) - increases the number of true positives called in the PARs, and additionally, in females only, the XTR. Unexpectedly, masking one copy of the XTR when aligning male samples results in a ten-fold higher rate of false positives in the XTR. We find that haploid calling on X and Y in male samples reduces the number of false positives compared to diploid calling but does not decrease the number of false negatives. We also find that using diploid-based filter thresholds on haploid chromosomes worsens variant calling for some filters. Finally, we find that - using a default alignment approach - joint genotyping the X chromosome between males and females slightly reduces the number of true positives called in females, while dramatically increasing true positives called in males, but this effect is mitigated when using a sex chromosome-complement informed alignment. By taking into account sequence similarities and ploidy differences on the sex chromosomes, we outline best practices for mapping, calling and filtering variants on the sex chromosomes. We recommend aligning samples to versions of the reference genome informed on the sex chromosome complement of the sample and to use biologically accurate ploidy parameters when calling variants and setting filtering thresholds. This approach results in a more accurate way of identifying variants on the sex chromosome and will help improve to account for the unique biology of chromosomes X and Y.
Comparisons between sex chromosomes and autosomes can provide fundamental insights into how the genome evolves and the forces that shape it. Specifically, differences in their transmission and ploidy drive their divergent evolutionary trajectories, with sex chromosomes being predicted to adapt and consequently evolve faster. However, in the well-studied model systems these effects are confounded: sex-linked genes are under less efficient selection due to their reduced effective population size, as well as more efficient selection because recessive mutations are exposed to selection in the heterogametic sex. Fungus gnats and gall midges are two fly families that present unique opportunities to disentangle these effects. They have XO sex determination, but males transmit only maternally inherited chromosomes. This phenomenon, known as paternal genome elimination (PGE), results in equal transmission of the X and autosomes, allowing the effect of hemizygous selection to be studied in isolation. We examine rates of divergence under PGE and find that, surprisingly, X chromosomes diverge more slowly than autosomes. Using population resequencing and gene expression data, we find that stronger purifying selection on the X drives its slower divergence. Our findings demonstrate the potential of systems with unusual inheritance for understanding fundamental evolutionary processes.
An evolutionary transcriptomics approach to understanding the origin of the HPG axis in sea lamprey, a basal vertebrate

Sara Victoria Good

Sara V Good, Avery Keen, Tamanna Yasmin
The University of Winnipeg (Canada), University of British Columbia (Canada), University of Manitoba (Canada)

Lampreys are jawless fish that belong to the oldest extant lineage of vertebrates and are the oldest lineage to harbour a pituitary. They have a complex life cycle that involves larval, juvenile and adult stages. Lampreys are considered models for the origin of the hypothalamus-pituitary-gonadal (HPG) axis. The objective of this research project is to understand the origin, mode of evolution, and expression of genes related to the HPG axis in gonadogenesis throughout development. First, we used phylogenetics and syntenic analyses to identify and classify putative orthologs of core vertebrate genes involved in the HPG axis in lampreys, based on whether the genes originated in i) the pre-vertebrate genome (amphioxus); ii) the post 1R genome; or iii) were cyclostome specific gene duplicates. Next, we employed an evolutionary transcriptomics analysis by further grouping genes into evolutionary strata based on their rate of evolution as estimated by the number of nonsynonymous to synonymous substitutions of genes relative to the outgroup hagfish. We then tested the hypothesis that slower evolving genes and/or those present in the pre-vertebrate ancestor, were more likely to be expressed during earlier developmental stages and exhibit less variation in gene expression over developmental stages, while lineage specific gene duplicates present within jawless fish or those exhibiting a higher rate of evolution exhibit higher variation in gene expression across sex and/or stage. The results of this research shed light on the core and peripheral genes involved in the early origin of the vertebrate-specific HPG axis.
Using phased genome assemblies to examine the evolution of heteromorphic sex chromosomes in Cannabaceae

Sarah Carey

Sarah Carey
HudsonAlpha Institute for Biotechnology (USA)

The Cannabaceae family has a deep history regarding their dioecious flowers and sex chromosomes. One century ago, the XY pair that controls the development of the sexes was identified in the hop Humulus lupulus var. lupulus, owing to the cytologically smaller Y chromosome relative to the X. Curiously, examinations across other species in Humulus and in the sister genera Cannabis uncovered varying cytological differences. In Cannabis, the Y chromosome is bigger than the X, while other varieties of Humulus remain homomorphic. Despite these early discoveries, we know little about the Cannabaceae sex chromosomes at the molecular level. This is largely due to the complexities of assembling XY pairs in genome references. Here we use a combination of Illumina, PacBio HiFi, and Dovetail Omni-C data to assemble fully-phased genomes for XY males of Cannabis and Humulus. We show that our assemblies match the known cytological differences in these XY pairs. We next use these assemblies to explore the timing of evolution, gene gain and loss, and structural complexity of these XY chromosomes to ultimately uncover genes that control the development of the economically valuable females of hemp and hop.
Shining a light on early stages of neo-sex chromosome evolution in Australian honeyeaters (Aves: Meliphagidae)

Sophia Catherine MacRae Orzechowski

Sophia Catherine MacRae Orzechowski, Doaa Doudin, Alex Drew, C. Jonathan Schmitt, Foyez Shams, Chris Wilson, Leo Joseph, Tariq Ezaz, Scott V Edwards

Australian National Wildlife Collection (Australia), Harvard University (USA), University of Canberra (Australia)

Mounting discoveries of avian neo-sex chromosomes are yielding rich opportunities to understand early stages of sex chromosome evolution in birds. Neo-sex chromosomes form when autosomes fuse to ancestral sex chromosomes, representing a type of sex chromosome turnover. We discovered and characterized the structure and phylogenetic distribution of neo-sex chromosomes in Australian honeyeaters by integrating long-read genomic, cytogenetic, and chromatin conformation data. We leveraged a phylogenetic framework to chart the progression of sex-linked recombination suppression across the newly added region. Finally, we generated transcriptomic data to investigate patterns of dosage compensation in the Blue-faced Honeyeater. Despite overall evidence for a single bout of recombination suppression 10-20 MYA involving 68% of the 55 Mb added region, we find little evidence of major gene loss or disruption of gene expression along the added W. However, we find extensive proliferation of transposable elements (TEs) on the added W, consistent with the refugium hypothesis for active TEs on sex-limited chromosomes. Our chromosome-level assembly reveals major rearrangements on the neo-W, including a 20 Mb inversion and a translocation of an ancestral W chunk that we hypothesize led to recombination arrest. Overall, our findings offer insight into the progression and consequences of recombination suppression on neo-sex chromosomes. Our 'chromosomics' approach provides us with fresh ammunition to get at the long-standing question of what happens when autosomes become sex-linked, and the extent to which each sex chromosome system has its own unique molecular journey.
Evolution of anisogamy in early diverging fungus Allomyces

Sujal Phadke

Sujal Phadke, Melissa Sayres
Arizona State University (USA), California State University (USA)

Why do many organisms have large eggs and tiny sperm? Addressing this fundamental question in evolutionary biology requires investigation into molecular pathways that regulate gamete dimorphism (anisogamy). It also requires investigation into selection pressures that may explain the prevalence of anisogamy. The fungus Allomyces belongs to the phylum Blastocladiomycota, which is an early diverging branch in fungi. Closely related species of Allomyces show anisogamy with dimorphism in gamete size, pigment, motility and pheromone production. Allomyces carries homologues of CatSper genes that are critical for mammalian sperm motility. Our preliminary RNA-seq analysis showed that male-biased samples of Allomyces overexpress CatSper homologs compared to the female-biased samples, which is reminiscent of the patterns of CatSper expression in mammals including humans. Furthermore, other groups have shown that synthetic analogue of Sirenin- the female pheromone of Allomyces, mimics the action of human hormone progesterone and can cause cross-species activation of human sperm motility through CatSper complexes in the membrane of human sperm. We hypothesized that genes responsible for gamete dimorphism and interaction are conserved between mammals and early diverging fungi including Allomyces. We identified homologs of additional gamete-specific proteins and investigated their patterns of expression in male-biased vs female-biased mutants in Allomyces. Our results suggest that the role of some of these proteins in gamete-interaction may have evolved at the root of Ophistokonts.
DNA methylation in female and male X chromosomes in human neurons and oligodendrocytes
Yong Hwee Eddie Loh

Robert Morgan, Yong Hwee Eddie Loh, Devika Singh, Isabel Mendizabal, Soojin V Yi
Arbor Biotechnologies (USA), Basque Foundation for Science (Spain), Biocruces Bizkaia Health Research Institute (Spain), Center for Cooperative Research in Biosciences (CIC bioGUNE) (Spain), Foundation Medicine, Georgia Institute of Technology (USA), Ikerbasque, Inc. (USA), University of California Santa Barbara (USA)

The presence of heterogametic sex chromosomes often leads to dosage imbalance between sexes. For example, if all human X-linked genes were expressed equally, females, with two X chromosomes, would have two times the RNA products of the X-linked genes compared to males. X chromosome inactivation (XCI) is considered to have evolved to silence one of the two female X chromosomes to balance the gene dosage, and one of the key mechanisms in this transcriptional silencing process is DNA methylation, whereby a methyl group (-CH3) is added to cytosine bases in DNA. Most DNA methylation studies to date often exclude the X chromosomes in their analyses. Moreover, many of these studies used array-based methods that examine only subsets of genomic positions, rather than the whole X chromosome. Consequently, our understanding of X chromosome DNA methylation lags behind that of autosomes. In this study, we investigated X chromosome DNA methylation using 89 neuronal and oligodendrocyte methylomes generated using whole genome bisulfite sequencing (WGBS), which provides an unbiased measurement of DNA methylation at single base resolution across the entire genome. This study furthers our understanding of gene dosage regulation by DNA methylation on the chromosomal level as well as on individual gene levels.
In the evolution of sex chromosomes, particular attention is given to the poorly understood phenomenon of "sex chromosome turnover." Sex chromosome turnover is defined as the phenomenon where the sex chromosome system changes from XY type to ZW type (or vice versa), or even when different evolutionarily originated sex chromosomes emerge within the same type (e.g., from XY to another XY). The Japanese frog (Glandirana rugosa) inhabiting Japan is a rare species that possesses both XY and ZW sex chromosomes within the same species, and the turnover of sex chromosomes from XY to ZW has been observed. Our research group aims to elucidate at the molecular level how the turnover of sex chromosomes occurs in the frogs. Previously, we determined the nuclear genome of the frog using short read sequencing (Katsura et al., LSA 2021). Furthermore, it has been suggested that sex chromosomes originating from at least three different chromosomes independently emerged within the population (Miura et al., Mol. Ecol. 2022). The frogs have 13 chromosomes, and in two populations (Tokai region and Hokuriku Tohoku region), the chromosome 7 has morphologically differentiated into ZW and XY sex chromosomes, as reported in previous studies. However, in other populations, sex chromosomes do not show morphological differentiation. In addition to presenting the results of sequence comparisons of morphologically differentiated XYZW sex chromosomes, I would like to introduce the findings from our analyses of populations and genome studies in the frogs.
S4 - Everything that is old becomes new: comparative genomics and museum specimens.
Viral and host metagenomics of 100-year-old insect museum specimens
Alexandra H Keene

Alexandra H Keene, Mark D Stengelein
Colorado State University (USA), Colorado State University (USA)

Recovery of virus sequences from old samples provides an opportunity to reconstruct historic virus-host interactions. Most studies of old viruses have used DNA or RNA from frozen or fixed samples. The millions of specimens in natural history museum represent a trove of old virus sequences, but it's not clear how well RNA survives in these samples. Here, we experimentally assessed the stability of RNA in dried insects over 72 weeks. Although RNA molecules from experimentally dried samples grew increasingly fragmented over time, RNA levels remained constant. RT-qPCR of host and virus RNA targets showed minimal differences between RNA in dried and frozen specimens. Next, we acquired D. melanogaster museum specimens from entomological collections across the United States to assess RNA survival in much older specimens. We recovered coding complete and partial sequences from known and novel viruses including several coding complete virus genomes from a fly collected in 1908. We found that the virome of D. melanogaster has remained relatively stable over the past century. Galbut virus, the most prevalent virus infection in contemporary D. melanogaster, was also the most common in historic samples. Finally, we investigated the genomic and physical features of surviving RNA and found that surviving RNA was chemically damaged, and apparently protected by RNA-RNA interactions or RNA-protein interactions. Overall, we concluded that RNA is stabler than typically thought. This highlights the utility of the millions of dried museum specimens to provide a clearer understanding of historic virus-host interactions and virus evolution.
Museomics improves and extends orang-utan mitochondrial phylogeny

Ana Agapito Vera

Ana Maria Agapito Vera, Viviani Mantovani, Samuel Wittwer, Irune Ruiz-Gartzia, Núria Albala Hermosilla, Esther Lizano, Logan Kistler, Tomàs Marques-Bonet, Michael Krützen

Smithsonian Institute (USA), Universitat Pompeu-Fabra (Spain), University of Zurich (Switzerland)

The demographic history of orangutans (genus: Pongo) has been shaped by ancient geological events, recent human-driven forest exploitation, and a unique dispersal system with extremely strong female philopatry. Because of the latter, phylogenetic analyses based on mitochondrial (mtDNA) genomes allow for a comprehensive reconstruction of past and present genetic diversity of this genus. Previous work revealed reciprocally monophyletic clusters in Borneo and Sumatra, aligning with the three recognized species. However, previous work was hampered by the lack of samples from areas where orangutans could not be sampled or are now extinct. In order to complement previous sampling efforts and potentially detect orangutan populations that might have been lost, we created a mitogenome phylogeny using 50 modern mitogenomes from previous studies and sequenced 61 mitogenomes from orangutan's museum specimens. Our analyses revealed a novel monophyletic subclade within Borneo, showing a geographic signal. Generally, divergence times were more recent than previous studies, but still within previous estimates. Our work highlights the utility of museum samples for obtaining a more complete picture of past genetic diversity, especially in organisms that for logistic or political reasons are hard to sample. The advance of ancient DNA techniques also allows to create whole-genome data from museum samples, allowing more complete insights into how orang-utan genetic diversity responded to anthropogenic perturbations while at the same time advancing our understanding of orangutan evolution in the Holocene.
Exploring mammalian evolution through comparative genomics with hundreds of species in the Zoonomia project

Elinor Karlsson

Presented by self
The Broad Institute (USA), UMass Chan Medical School

Zoonomia is one of the largest comparative genomics resources produced to date and a potentially powerful tool for therapeutic discovery. By comparing the genomes of 240 mammal species, we harness over 100 million years of evolution to explore genome function. Overall at least 11% of the human genome is constrained, and constrained bases are enriched for variants explaining common disease heritability more than any other functional annotation. We associate coding and regulatory variants with variation in hibernation and brain size across placental mammals, and reconstruct the phenotype of an ancient sled dog named Balto from his genome. Nearly half of highly constrained bases are unannotated in existing data resources, highlighting how much of the regulatory landscape of the genome has yet to be explored. We are now combining comparative genomics with cellular models to investigate the evolution of traits across species.
Ancient RNA expression profiles from the extinct woolly mammoth

Emilio Mármol Sánchez

The analysis of ancient DNA has recently gained considerable momentum, allowing the study of extinct and extant organisms that lived up to 2 million years ago. This has enabled the reconstruction of genomes and historical ancestry of multiple extinct species, as well as revealing the complex nature of the bygone ecosystems where they once thrived. However, current DNA sequencing techniques cannot alone provide information about tissue identity, gene expression dynamics or transcriptional regulation – all of which is encoded in the RNA fraction, and critical to fully understanding the biology of these now lost species. Here we report the oldest ancient RNA expression profiles ever recorded and dated to around 30,000 years-old from a permafrost-preserved woolly mammoth, representing the first instance of gene expression profiles characterized in an extinct species from the Late Pleistocene. We report tissue-specific gene expression patterns at both the coding and noncoding level, comparative analyses with mammoth ancient DNA data and genomic and transcriptomic traces supporting that a phenotypically identified female mammoth is actually a genetic male. Our findings provide evidence for the preservation of meaningful transcriptomes beyond preconceived temporal limits, doubling the age record registered so far. With our results, we anticipate a renewed interest in the study of ancient RNA molecules, pushing the field beyond DNA and fostering a new era of integrative paleogenomics and paleotranscriptomics studies.
The trait specific timing of accelerated genomic change in the human lineage.

Eucharist Kun

Eucharist Kun, Mashaal Sohail, Vagheesh Narasimhan
The University of Texas at Austin (USA), Universidad Nacional Autónoma de México (Mexico)

Humans exhibit distinct characteristics compared to our primate and ancient hominin ancestors including bipedal locomotion and enhanced neurocognitive ability, but the timing of accelerated changes in these traits is uncertain. To investigate if specific trait-associated variation show enrichment during particular periods of human evolution, we combine genome wide association study (GWAS) data from 70 traits, spanning multiple categories including AI-based image-derived morphological phenotypes of the brain, heart, and skeletal tissues with data from 12 different evolutionary regions obtained from comparative functional genomics, multi-species alignments from long read sequencing, and ancient DNA reflecting 4 different major evolutionary divergence points. These regions cover epigenetic differences in the brain between humans and rhesus macaques, various human accelerated regions (HARs) including regions from the Zoonomia Project, ancient selective sweeps, and Neanderthal introgressed alleles. Using two complementary approaches to examine enrichment between GWAS loci and genomic regions, we found that traits across respiratory, dermatological, reproductive, and metabolic categories were enriched through multiple timepoints, along with skeletal and brain traits, consistent with striking morphological changes between humans and other primates. We also find that different brain structures associated with various functions were enriched at different evolutionary depths, including the longitudinal fasciculus in human-gained epigenetic elements since macaques, the visual cortex in HARs, and the thalamus proper in Neanderthal introgressed alleles. Lastly, we show that more phenotypes are enriched in the periods of divergence between macaques and humans or chimps and humans, and less so during the divergence with Neanderthals.
Herbarium seed embryos as alternative sources of ancient DNA
Hong Phuong Le

Hong Phuong Le, Yoon Kyung Lee, Rafal Gutaker
KEW (United Kingdom), Royal Botanic Gardens

Natural history collections, which contain hundreds of millions of specimens classified in terms of time, space, and taxonomy, provide valuable resources for many fields of research. Since the first success of ancient DNA (aDNA) extraction in the 1980s, these repositories, including herbaria for plants, have been intensively used to study microevolutionary processes, especially genetic responses to anthropogenic activities over the past several centuries. Two crucial challenges of aDNA research comprise contamination by environmental DNA, and degradation. aDNA extracted from herbarium leaf specimens are short fragments, ranging from 50 to 100 bp due to a higher breakdown rate than that in bone assemblage. To maximise the amount of data retrieved and minimise destructive sampling, we have compared whole-genome sequencing libraries constructed for leaves and seed embryos in herbarium specimens. Additionally, damage patterns and the proportion of endogenous DNA in both samples have been evaluated. The results have shown that DNA isolated from embryos was less fragmented than that from leaves. Accordingly, large embryos in seeds are a very promising alternative source of genomic DNA from herbarium specimens.
Comparative Genomics of Glossophaga mutica (Phyllostomidae): Demographic History and Evolutionary Processes

Jesús Antonio Rocamontes Morales

Jesus Antonio Rocamontes Morales, Jorge Ortega, Enrique Ibarra Laclette, Gabriela Castellanos Morales
A.C. (INECOL) (Mexico), El Colegio de la Frontera Sur (ECOSUR), Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional (Mexico), Instituto de Ecología, Unidad Villahermosa (Mexico), Universidad Autónoma de Zacatecas (Mexico)

The discovery of cryptic species is common in phyllostomid bats. At times, traditional taxonomic tools such as morphology fail to reveal the existence of cryptic species within species complexes. In contrast, comparative genomics allows for the analysis of divergence and similarity patterns among genomes, and the comparison of the demographic history of a species from a single genome, to contribute to the resolution of taxonomic debates. For instance, within the genus Glossophaga, Glossophaga soricina was previously considered a species complex. Recently this species complex was separated into four species, one of which (Glossophaga mutica) is endemic to Mexico and Central America. Interestingly, G. mutica includes an insular population in the Islas Marías archipelago, near the coast of Nayarit in Mexico. Phylogenetic relationships between island and mainland populations are still unclear, as they exhibit differences in their morphological characteristics, suggesting that the insular population may have undergone a distinct evolutionary trajectory, which can be studied from a genomic perspective. In this project, we used a whole genome sequencing approach to compare the nuclear genomes of one G. mutica individual from the Islas Marías archipelago and one from the mainland to understand the differences in their evolutionary trajectories and demographic history to test the hypothesis of two distinct species in mainland and insular G. mutica. We present preliminary results of genomic assemblies and results of SNP discovery between the nuclear genomes of G. mutica from continent and island. Our results show 14,561,725 SNPs between G. mutica genomes.
Scattered about the mammalian phylogeny are species whose cells endure dramatic temperature variation that would harm human cells. Hibernators, such as little brown bat and 13-lined ground squirrel, drop to refrigerator-level body temperatures every winter, and desert species such as dromedary camel, undergo daily thermal cycling. Have cellular-resilience mechanisms in these species emerged independently under similar selective pressures, or do relevant genes and pathways overlap, suggesting shared evolutionary history? Powered by resources from our Zoonomia comparative genomics project and from San Diego Zoo’s Frozen Zoo®, the world’s only institution to bank cells of more than 1,000 taxa, we are using a classical common-garden framework to discover the evolutionary history of cellular robustness. We cultured primary dermal fibroblasts of 12 deeply diverged mammals, including both strict thermoregulators and heterotherms, and assessed responses across a range of temperatures. We identified several pathways universally impacted by temperature perturbation, including regulation of cholesterol biosynthesis, expression of heat shock proteins, and maintenance of the cell cycle. We found evidence for convergent shifts in gene expression in 1,054 genes in heat-flexible species under hot treatment (41°C) and 875 genes in hibernator species under cold treatment (32°C). Under hot treatment, dromedary camel transcriptome exhibited widespread post-transcriptional changes, including a large shift in isoform usage of HSF1, an early responder to heat shock, highlighting a potential role for alternative splicing. Our comparative framework holds great promise for unraveling the mechanisms of cellular robustness and discovering therapeutic avenues for diseases of compromised homeostasis.
From swampy ancestors to modern maize: tracing the allopolyploid origins of Zea
Michelle C Stitzer

Michelle C Stitzer, Taylor AuBuchon-Elder, Thuy La, Robert Bukowski, Qi Sun, M. Cinta Romay, Edward S Buckler, Elizabeth A Kellogg
Cornell University (USA), Donald Danforth Plant Science Center (USA), USDA-ARS (USA)

Early cytological, genetic, and genomic studies hypothesized maize originated from an allopolyploidy event, due to the presence of duplicated regions throughout the genome. Most lines of evidence suggested the parents of the allopolyploidy belonged to the highly productive Andropogoneae tribe of grasses, today containing ~1,200 species. Limited genomic sampling of these non-model, often polyploid, grasses has left considerable gaps in our understanding of the evolutionary history of the maize genome. Here, we use extensive herbarium and field collections to sample grasslands globally, where Andropogoneae species cover nearly 20% of terrestrial land. We generate whole genome short read sequence (~15x depth) of over 500 species, and phylogenetically targeted PacBio assemblies of 28 species. We reconstruct phylogenetic relationships for 9,500 conserved gene loci, and estimate divergence of transposable elements. Approximately half of loci place the maize lineage sister to subtribe Rhytachniniae, which is today found in wet environments in Africa, growing in river banks and seasonally swampy environments. The other subgenome donor is more obscure, with contributions from extant Ratzeburgiinae and Elionurus, each of which has a few species present today in the Americas. We interpret preferential retention and stronger purifying selection on gene copies inherited from the wet Rhytachniniae ancestor alongside biogeographic reconstructions that show the polyploid originated in Central America as evidence that the newly formed allopolyploid succeeded in a novel ecological niche. We conclude that considering the subgenome ancestry of maize can be helpful when interpreting modern maize diversity, especially for traits like drought and flooding tolerance.
Advancing social designs for biorepository and biodiversity data services in comparative genomics

Nico Franz

Presented by self
Biodiversity Institute, University of Kansas (USA)

This presentation focuses on issues of social design, sample/data governance, and credit sharing in comparative genomics. A relevant context is the aim to attain and publish reference genomes at the scale of all presently recognized eukaryote species, i.e., the Earth BioGenome Project. Inclusive and equitable social designs are relevant to the scaling of genomic sample acquisition, sample processing and storage, and metadata publishing. The perspective being shared is that of a biorepository and biodiversity data science culture that is connected to natural history collections at U.S. educational institutions (such as public universities), with a strong emphasis on recognizing the diverse research achievements made by students and early-career scientists. Natural history collections record and integrate sample data in accordance with the Darwin Core standard (https://doi.org/10.1371/journal.pone.0029715). Globally, nearly 3.0 billion Darwin Core occurrences representing 1.5 million recognized species are aggregated through the Global Biodiversity Information Facility (https://www.gbif.org/). This vast resource of biosample data, and the repositories through which the physical specimens are maintained, jointly constitute a research infrastructure whose existing and emergent designs will greatly impact effort and data ownership and governance in the genomic sciences. Three examples will be presented to illustrate this point. (1) The opportunity for systematic specialists working with samples - including students and early-career scientists - to have control over the taxonomic system(s) according to which samples are identified and aggregated in regional to global information networks (https://doi.org/10.1093/database/bax100). (2) The opportunity for individual collections, sample collectors and sample identifiers to be recognized for their itemized contributions to such environments (https://doi.org/10.3897/biss.4.59167). (3) The opportunity for filtered, provenance-aware propagation of annotations on sample data across multiple, decentralized data portals (https://doi.org/10.3389/fdata.2020.519133). Although these components are not guarantors of an equitable and inclusive culture for managing genomic samples and data, they are technical and social prerequisites for (e.g.) implementing CARE Principles in comparative genomics (https://doi.org/10.1038/s41559-023-02161-2). The presentation will showcase current and planned solutions to the three challenges, as implemented in the Symbiota software platform for publishing biocollections data (https://symbiota.org/).
Metagenomics for non-model species and museum collections pave way for studying changes in environmental microbiome.

Rafal Marek Gutaker

Rafal M Gutaker, Robin B Guevarra
Kew (United Kingdom), Royal Botanic Gardens

Natural history museum collections are repositories of time-stamped samples for plant and animal species. Most collections were made in the last couple of centuries and have documented some of the most radical anthropogenic environmental changes. Analysing such collections through the lens of metagenomics would facilitate research on the impact of agriculture, land-use change and widespread use of antibiotics on microbiomes. Two major bottlenecks are: difficulties in distinguishing between bona fide microbiome and post mortem colonizers, and the assumed need to remove host genome for non-model species. Using three examples: cultivated plants (rice), wild plants (ragweed) and insects (beetles), we show that the lack of host reference genome does not preclude informative metagenomic assignments, even in historical samples. Furthermore, as exemplified by plant herbarium, we show that it is possible to build collection-specific reference panel for post mortem contaminations, which in turn facilitates microbiome comparisons in time. Finally, we show one application of such approach and the discovery of previously unknown strains of nitrogen-fixing bacteria associated with rice cultivation over 100 years ago. In summary, we showcase that metagenomics on historical biological samples is possible. Given its huge potential for understanding environmental change, we encourage more research in this direction.
Integrating paleontological, morphological and molecular methods of telling evolutionary time - assessing the timescale of crocodilian evolutionary history.

Sidney Davies

Sidney Davies, Phil Donoghue, Davide Pisani
University of Bristol (United Kingdom)

In phylogenetics, the molecular clock represents the only viable means of converting the relative ages of species divergences into absolute times in order to provide a fully-dated phylogeny that is calibrated to geologic time. However, there are a diversity of methods available, both to account for rate heterogeneity across the tree and to incorporate information from the fossil record in order to provide reliable temporal constraints that are used to extrapolate the mutation rate and determine the time of divergence among species pairs. There is much debate over which methods are the most appropriate to employ in a variety of evolutionary scenarios, but despite this, very few studies have investigated the relative performance of different clock methods in a single, integrated study. We set out to achieve this goal using crocodilians as a model group to compare the ways in which various commonly-used methods used to incorporate fossil information into a molecular clock analysis behave in an empirical setting, such as node-dating, tip-calibration, total-evidence analysis and the fossilized birth-death process. This allows us to make direct comparisons on the behaviour of disparate clock methods in the context of a well-studied taxonomic group with a rich evolutionary history, and may provide insight on how we can expect various methods to perform in future empirical analyses.
The early evolutionary history of HIV remains relatively unknown due to a paucity of sequence data prior to the 1980s. Without ancestral HIV sequences, understanding of viral emergence is limited to results from phylogenetic models predicting past patterns from modern data. Accessing ancestral sequences from archival fixed tissues presents an opportunity to glimpse into the evolutionary past of the virus. The acquisition of viral sequences from archival tissue is, however, stymied by the low abundance of viral reads against the host background. To combat this methodological challenge, we compared four priming techniques during reverse transcription to optimize whole genome viral cDNA generation. We compared the use of random primers, gene-specific primers, and two combined approaches to optimize the coverage of viral targets. These approaches were tested on modern HIV-1 RNA spiked onto human RNA at various concentrations. The cDNA samples were then prepared for sequencing using single-stranded library preparation and sequenced. The mean coverage of the HIV-1 genome was similar across priming techniques. The copy number of specific amplicons was higher, however, in samples primed using gene-specific primers. These results demonstrate the potential utility of using a combined priming approach to generate near whole-genome data from viral species in archival samples. Understanding the emergence and spread of viruses affecting modern and historical populations can be improved by accessing viral sequences, but this can only be done if robust techniques specific to viral RNA in fixed tissues are developed, as we have done here.
A Backbone Tree of Formicidae —— The First Step to Reveal Ant Evolution in the Cretaceous Terrestrial Revolution and Post K-Pg Extinction Era

Wentao Tao, Davide Pisani, Philip Donoghue, Donato Antonio Grasso, Enrico Schifani
Department of Chemistry, Life Sciences, School of Biological Sciences (United Kingdom), School of Earth Sciences (United Kingdom), University of Bristol, University of Parma, and Environmental Sustainability (Italy)

With over 15,000 species, colonising nearly all terrestrial environments, ants are one of the most successful insects on the earth. However, despite genome-scale data have been used to explore their internal relationship, there are still some inconsistencies between different researches. Mysteries, such as the relationship of the subfamilies Leptanillinae and Martialinae with other living ants, as well as the root age and divergence time of internal clades, are still waiting to be revealed. Here, we build up the largest genome-scale dataset, containing 91 ant species and 10 outgroups. After going through a series of cleaning pipeline to remove hidden paralogs and long branch species, we build up trees under the best fitting mixture model, then calibrate them with 19 fossils following the best practice of fossil calibration so as to investigate the evolutionary history of ants. Our phylogenetic analysis supports the hypothesis that the leptanillomorph clade (Leptanillinae and Martialinae) is the sister group to all other living ants. While the novel molecular clock analyses suggest that crown ants originated in the early-Cretaceous and diversified afterward with a younger age compared to the previous research. Therefore, in the future, by adding more molecular and fossil data, combining with the latest tree-building and fossil-calibrating method, we can better know about (1) the diversification of crown ants in the late-Cretaceous under the background of the angiosperm’s explosion, and (2) how they occupied the niches previously belonging to stem group ants after the K-Pg event, finally achieved success today.
S5 - Human evolution in the genomic era.
Mutations in the human genome have deleterious, neutral or advantageous effects, which are quantified with a selection coefficient (s). The distribution of fitness effects (DFE) measures the probability of new mutations having a certain selection coefficient and, as such, is central to fundamental questions in population genetics such as how natural selection influences genetic diversity. The DFE has previously been inferred for a variety of species using the site frequency spectrum (SFS), measures of haplotype lengths as well as different groupings of summary statistics. A drawback of these approaches is that they focus on specific summary statistics of the genetic data which are not guaranteed to be sufficient to estimate the impact of natural selection. Notably, all patterns of genetic variation explained by summary statistics are encoded on the Ancestral Recombination Graph (ARG), which is a data structure that shows the structure of local genealogies across the genome. Natural selection changes coalescent patterns from local genealogies making the ARG a rich data source to infer the DFE. In this study we leverage information from the ARG to build a maximum likelihood approach to infer the DFE using information from the number of lineages with a derived allele at specific time points. We show through simulation that this approach can accurately infer a wide range of selection coefficients (0.1>2Ns<1000) and show an application of the method to infer the DFE of non-synonymous mutations of the Yoruba population from the 1,000 genomes project.
Simple, general tests for accelerated evolution and positive selection

Alexander L Starr

Alexander L Starr, Leslie Magtanong, Gabriella Cale, Alexander Hoefler, Hunter B Fraser
Stanford University (USA)

A large body of work has established that the genetic basis of human complex traits primarily lies in many non-coding variants with small phenotypic effects. However, widely used methods to detect ancient changes in the strength of natural selection are primarily applicable only to small genomic regions or protein-coding sequences, creating a mismatch between what is known about human genetics and the methods used to understand human evolution. To address this, we developed two simple, general methods to detect mono- and polygenic (1) accelerated evolution of and (2) positive selection on coding and non-coding sequence genome-wide. Using the first method, we identified hundreds of genes that have undergone accelerated evolution on coding and nearby non-coding sequences in the human lineage. We highlight the gene with the strongest evidence for cis-regulatory accelerated evolution, LMO4, which plays a key role in determining handedness and brain asymmetry. Although accelerated evolution can be the result of positive selection, it can also result from an increased local mutation rate or loss of evolutionary constraint. Therefore, we developed a new test for positive selection that is applicable genome-wide. Using this test, we find evidence for positive selection at non-coding sites near dozens of genes, including at the EVC/EVC2 locus previously linked to human-specific craniofacial morphology and many transcriptional regulators. Overall, we anticipate that our broadly applicable frameworks will improve our understanding of how natural selection has shaped the evolution of humans and other species.
Selective sweeps are significant evolutionary events where beneficial variants increase rapidly in frequency and reach fixation in the population. Sweeps affect diversity at surrounding variants via hitchhiking. Several methods exploit the excess linkage disequilibrium (LD) pattern caused by hitchhiking for sweep detection. In this project, the LD statistic Dz is utilized to detect selective sweeps in simulations and in 1000 Genomes Project data. Dz was introduced by Hill and Robertson as a mathematical stepping stone to compute $D^2$, which is the usual measure of covariance between allele frequencies. Dz computes the degree of LD between low frequency variants. It has recently emerged as a useful empirical measure of diversity. We show that the expectation of Dz is more informative about recent and ancient sweeps compared to other commonly used LD statistics, with adequate power for up to 100 thousand years ago in simulation with human-like parameters. Incorporating expected Dz in a scanning window in different populations from 1000 Genomes Project data recaptures sweep signals from widely studied sweeps. This project demonstrates Dz's potential for incorporation into the current line of sweep detection techniques, taking advantage of genetic patterns not considered by most current methods.
Investigating the relationship between runs of homozygosity and changes in human height over 35,000 years

Ana Victoria Leon Apodaca

Ana V. Leon-Apodaca, Zachary A. Szpiech
Pennsylvania State University (USA)

Demographic processes such as consanguinity, bottlenecks, and natural selection increase the likelihood of haplotypes being inherited identical-by-descent (IBD) from a recent common ancestor. When these identical haplotypes are inherited through both the maternal and paternal lineages, they manifest as runs of homozygosity (ROH). In humans, ROH is commonly associated with recessive Mendelian disorders, however, increased homozygosity has also been associated with complex traits such as height. Height is a highly heritable and polygenic trait influenced by both genetic and environmental factors. Research conducted on present-day human populations of European descent have found an association between genome-wide homozygosity and height, with more homozygosity resulting in lower height. Archaeological evidence and genomic data obtained from ancient human remains spanning the Upper Paleolithic (38,000 yBP) to the Iron Age (2,400 yBP) have shown there was a marked reduction in human height that occurred during the transition to agriculture that took place during the Neolithic, followed by an increase during the subsequent post-Neolithic periods of agricultural intensification. In this work, we investigate the relationship between historical demographic changes, ROH, and height; we assess whether changes in ROH patterns observed across ~35,000 years across ancient European populations are associated with human height, and whether genetic variants known to influence height are more often observed within ROH regions.
Estimating gene conversion tract length and rate from PacBio HiFi data

Anders Poulsen Charmouh

Anders Poulsen Charmouh, Peter Porsborg Sørud, Thomas Bataillon, Asger Hobolth, Lasse Thorup Hansen, Søren Besenbacher, Mikkel Heide Schierup

Aarhus University (Denmark), Aarhus University. (Denmark), Bioinformatics Research Centre, Department of Mathematics, Department of Molecular Medicine (MOMA)

A non-crossover event (NCO) occurs when the DNA double-strand break (DSB) is repaired using the homologous chromosome as template through the synthesis-dependent strand annealing pathway, or double Holliday junction pathway (Chen et al. 2007; McMahl et al. 2007). Some NCOs are observable as gene conversions which occur when one SNP is transferred unidirectionally from one sequence to another (Mansai et al. 2011; Lorenz and Mpaulo 2022). NCO is an understudied recombination mechanism despite most DSBs yielding NCO rather than a crossover (Cole et al. 2012; Halldorsson et al. 2019). Further, gene conversions have several effects on genomic diversity and hence many evolutionary implications as they can generate new combinations of alleles, cause de novo mutations due to the potentially mutagenic effect of recombination and even counteract the mutation load by reverting germline mutations through GC-biased gene conversion (Arbeithuber et al. 2015; Halldorsson et al. 2019). Here, we show that gene conversion events can be called directly through HiFi PacBio sequencing data, without using trio data or breeding experiment. We develop a new model of gene conversion rate and tract length and show that very few percent of all NCOs are observable as gene conversions. We estimate the rate of NCO for the genome of humans and other great apes. The results are relevant with respect to understanding NCOs and the extent to which they affect the genome.
On the genetic basis of tail loss evolution in humans and apes
Bo Xia

Presented by self
Broad Institute of MIT and Harvard, Harvard University, NYU (USA)

The loss of the tail is a major phenotypic change that have occurred during evolution leading to the lineage of humans and apes. Despite its long-discussed impacts on the locomotion style, the genetic mechanism of tail-loss in hominoids remains unknown. We present evidence that an Alu element inserted into the intron of TBXT gene of the hominoid ancestor might be a key genetic mechanism for such a phenotypic evolution. We demonstrate that this Alu element induced a hominoid-specific alternative splicing event of TBXT through a noncanonical way. Results from multiple mouse models further supported the notion that the hominoid-specific alternative splicing isoform is sufficient to induce a tail-loss phenotype. We further speculate that tail-loss evolution may have been associated with an adaptive cost of the potential for neural tube defects, which continue to affect human health today.
The human brain has evolved distinct traits compared to nonhuman primates, including differences in absolute size, relative size of prefrontal cortical regions, distribution of cell populations, and changes in gene regulation and expression. Literature indicates that these variations stem from a longer duration of neurodevelopment during late fetal and postnatal stages in humans compared to nonhuman primates. Early neurodevelopment is generally conserved across many vertebrate species, involving the transformation of a flat neural plate into a tube shape through extension, bending, and folding of neural tissue. This process has been observed through in-vivo morphological and molecular comparisons across zebrafish, frogs, chicks, mice, and humans. However, these stages differ in detail and give rise to the central and peripheral nervous systems and thus may be the origin of fine grain, human-specific traits in the brain. Advancements in in-vitro organoid models have expanded the scope of comparative studies on early ape brain development, which were previously limited by the availability of subjects. These aforementioned factors motivate the investigation of the role of early neurodevelopment in the evolution of the human brain through the generation of novel 3D organoids representing early neurodevelopment from stem cells of various primate species, as well as a comparison of early and late neurodevelopment using previously published RNA-seq data.
From Foraging to Farming: Tracing the Impact of Agricultural Adoption on Adaptation and Selection using whole-genome sequencing

Bridget Chak

Bridget Chak, Elise Lauterbur, Evan Wu, Maanasa Raghavan, David Enard, Luis Barreiro
University of Arizona (USA), University of Chicago (USA)

The Neolithic revolution marked a pivotal transition in human history; from altering infectious disease burdens to narrowing dietary composition and diversity. Despite its significance, we lack a comprehensive understanding of how the emergence of agriculture has shaped the genetic histories of and adaptation among different human populations. Leveraging cross-continental comparisons, we investigated the extent to which convergent evolution of phenotypes and biological functions related to immunity and metabolism are observed among hunter-gatherer and agriculturalist populations facing similar ecological environments or evolutionary pressures. Here, we collated a cross-continental dataset of newly generated and published high-coverage (~30x) whole-genome sequencing data from 10 hunter-gatherer, 2 pastoralist, and 21 agriculturalist populations from 6 geographic regions: Southeast Asia, South Asia, West Central Africa, East Central Africa, East Africa, and Southern Africa. This dataset includes a total of 974 individuals. We characterized a broad spectrum of genetic diversity and variation, including 62,908,806 SNPs and 15,319,830 indels. Using Flex-sweep, a machine learning-based method designed to detect diverse selective sweeps, we present evidence of population-private and -shared signals of positive selection among hunter-gatherer and agriculturalist populations, including significant adaptation in virus-interacting proteins. Overall, our study sheds light on how the process of agricultural adoption played out in varied global contexts with respect to the impact on human demographic and selection histories.
Ancient AMY1 gene duplications primed the amylase locus for adaptive evolution upon the onset of agriculture

Charikleia Karageorgiou

Charikleia Karageorgiou, Feyza Yılmaz, Kwondo Kim, Petar Pajic, Kendra Scheer, Christine Beck, Ann-Marie Torregrossa, Charles Lee, Omer Gokcumen

The Jackson Laboratory for Genomic Medicine (USA), The University of Connecticut Health Center (USA), University at Buffalo (USA), University of Connecticut (USA)

Starch digestion is a cornerstone of nutrition, with the amylase enzyme playing a crucial role in starch metabolism. The copy number of the human amylase gene has been linked to both metabolic diseases and the adaptation to agricultural diets. Prior studies have suggested recent origins of salivary amylase gene duplications. Here, we characterized 51 distinct amylase haplotypes across 98 individuals employing long-read DNA sequencing and optical mapping methods. Additionally, analysis of 31-mers linked to individual haplotypic blocks in the locus suggests that the first duplication of the amylase locus occurred more than 700,000 years ago, predating the human-Neanderthal split. Further, our analysis of ancient genomes revealed a shift towards higher amylase copy numbers in Neolithic farmers than in hunter-gatherer groups. The comparison of amylase haplotypes allowed us to characterize multiple nonallelic homologous recombination (NAHR) and microhomology-mediated break-induced replication (MMBIR) events, underlying different types of structural variants in the locus. Overall, our study reveals the mutational mechanisms and evolutionary forces that shape the remarkable structural complexity of the amylase locus.
Diverse patterns of transposable elements expressions across tissues exhibited by Rhesus macaque and possible regulation of gene expression of adjacent genes by tissue-specific TEs
Dae-Soo Kim
Kun-Hyang Park, Nakyong Kim, Dae-Soo Kim
Department of Core Facility Management Center (Korea, Korea Research Institute of Bioscience & Biotechnology (Korea, Republic of))

Transposable elements (TEs) constitute approximately half of primate genomes and play a pivotal role in gene regulation, potentially contributing to the evolution of phenotypic diversity. In this study, we examined TE expression patterns across 13 distinct tissue types in the rhesus macaque, a widely used non-human primate model, using publicly available RNA-seq and Cap analysis gene expression datasets. Our analysis revealed non-random distribution and differential enrichment of TEs across genomic regions, with intergenic TEs being predominantly silenced and highly expressed TEs notably concentrated in the thymus. We identified four distinct clusters of TEs exhibiting tissue-specific enrichment patterns, with enrichment levels varying by tissue type and TE age. Particularly noteworthy were the 365 short interspersed nuclear elements (SINEs) and 269 long interspersed nuclear elements (LINEs) highly enriched near transcription start sites (TSSs), suggesting potential regulatory roles. Furthermore, we uncovered 37 tissue-specific TEs highly expressed within coding DNA sequences (CDS) and 1226 upstream of genes, potentially driving tissue-specific gene overexpression. By examining pairwise comparisons and associations between TEs and genes, we elucidated the role of tissue-specific TEs in regulating nearby gene expression in primates. Our findings contribute to a deeper understanding of TEs as regulators of gene expression in a tissue-specific context, shedding light on the intricate mechanisms underlying phenotypic diversity in primates. This study enhances our comprehension of the complex interplay between TEs and gene regulation, offering valuable insights into primate genome evolution and functional genomics.
Indians have been underrepresented in human genomic studies. We generated ~2,700 high coverage genomes from India—including individuals from most geographic regions, speakers of all major languages, from tribal and caste groups—providing a comprehensive survey of genetic variation in India. We show that most Indians derive ancestry from ancient Iranian farmers, Eurasian Steppe pastoralists and South Asian hunter-gatherers. We uncover a common source of Iranian-related ancestry from early Neolithic cultures of Central Asia across diverse Indian groups, including Ancestral South Indians, Ancestral North Indians, Austroasiatic-related and East Asian-related individuals, suggesting that the Iranian-related gene flow occurred well before the Steppe pastoralist-related gene flow. Following these admixtures, India experienced a major demographic shift towards endogamy, resulting in extensive homozygosity and identity-by-descent sharing among individuals. At deep timescales, Indians derive ~1-3% ancestry from Neanderthals and Denisovans. By assembling the surviving fragments of archaic ancestry in Indians, we recover ~1.5Gb of the introgressing Neanderthal and ~0.6Gb of the introgressing Denisovan genomes, more than any previous archaic ancestry study. Indians have the largest variation in Neanderthal ancestry and the highest amount of population-specific Neanderthal segments among worldwide groups. Finally, we demonstrate that most of the genetic variation in Indians stems from a single major migration out of Africa that occurred around 50,000 years ago, with minimal contribution from earlier migration waves. Together, these analyses provide a detailed view of Indian evolutionary history and underscore the value of expanding genomic surveys to diverse groups outside Europe.
Understanding variation in chromatin contact patterns across human populations is critical for interpreting non-coding variants and their effects on gene expression and phenotypic variation. However, experimental determination of chromatin contacts at a population-scale is prohibitively expensive. To overcome this challenge, we developed a machine learning method to quantify the diversity 3D chromatin contacts at 2 kilobase resolution from genome sequence alone. We validated that it accurately generalized on contact data from diverse individuals. We then applied this approach to thousands of diverse modern humans and the inferred human-archaic hominin ancestral genome. Patterns of 3D contact divergence genome-wide are qualitatively similar to patterns of sequence divergence, but we find that 3D divergence in local 1-megabase genomic windows does not follow sequence divergence. In particular, we identify 392 windows with significantly greater 3D divergence than expected from sequence. Moreover, 26% of genomic windows have rare 3D contact variation observed in a small number of individuals. Using in silico mutagenesis, we find that most sequence changes do not result in changes to 3D chromatin contacts. However, in windows with substantial 3D divergence, just one variant can lead to divergent 3D chromatin contacts without high overall sequence divergence. In summary, inferring 3D chromatin contact maps across human populations reveals diverse contact patterns. We anticipate that these genetically diverse maps of 3D chromatin contact will provide a reference for future work on the function and evolution of 3D chromatin contact variation across human populations.
Why do we get sick? Unraveling Genetic Trade-offs between Fertility, Longevity, and Complex Diseases
Eva Brigos Barril

Modern humans exhibit lower fertility rates, extended longevity, and a higher incidence of complex diseases relative to other mammals. Life history theory suggests that varying levels of investment in reproductive function can impact the likelihood of developing diseases and longevity, suggesting intricate interactions between life-history traits and disease. However, despite a few identified mutations acting as genetic switches, we lack unambiguous and systematic genomic evidence to understand these trade-offs. We use genomic data to evaluate the shared pleiotropic landscape between fertility, longevity, and 61 complex diseases across 12 disease domains. We report a negative genetic correlation and pleiotropy between fertility and longevity. This trade-off is also evident when evaluating their shared genetic architecture with disease. We show that positive genetic correlations are common between fertility and disease. The demographic analysis using the UK Biobank supports the latter genetic observations, where individuals diagnosed with a disease tend to have more offspring than controls. In contrast, there is a predominant proportion of negative genetic correlations and pleiotropies with longevity. Specifically, we identified 297 pleiotropic SNPs shared between diseases and fertility and 569 with longevity. The analysis of the signature of selection shows that risk-allele frequencies are higher for pleiotropic variants that are linked to increased fertility. Also, these have undergone recent positive selection despite increasing disease risk. Overall, we highlight fertility's role in shaping current disease patterns and disentangle the complex network of interactions between fertility, longevity, and disease.
Exploring the Landscape of Gene Family Size and their Associated Pseudogenes in the Human Genome

Gabriela Procopio Leite

Gabriela Procopio, Beatriz Mello, Carlos Schrago
Federal University of Rio de Janeiro (Brazil)

Gene family dynamics play a crucial role in shaping the evolutionary history of species. Segmental duplication, a key driver of expansion, provides the raw material for the evolutionary process. Although often deemed as adaptative, these expansions should be carefully evaluated against the neutral duplication rate shaped by factors like recombination and adjacent mobile/repetitive elements composition. This implies certain genomic regions may be predisposed to duplication, independent of natural selection pressures. For instance, Nozawa et al. (2007) showed the number of pseudogene matches that of functional counterparts within the human olfactory receptor (OR) gene family. This equivalence suggests a similar probability of persistence and degeneration for novel copies in populations, casting doubt on the traditionally attributed selective advantage. Therefore, expansions may result from the interplay between duplication, purifying selection, and demography. Our study evaluates whether family sizes are positively correlated with the corresponding pseudogene number in the human genome, indicating their dynamics are mainly driven by fundamental molecular mechanisms associated with errors leading to segmental duplications. We investigate whether gene family expansions predominantly correlate with differential recombination rates and mobile elements composition. We found the number of pseudogenes is significantly positively correlated with paralog count. Moreover, this correlation is stronger for larger gene families. We map gene family size distribution and their associated pseudogenes across human chromosomes, juxtaposing this pattern against recombination rates and mobile element compositions. This comprehensive analysis offers an understanding of the processes shaping the genomic architecture of gene family size in humans.
Setting up an integrated pipeline of analyses to explore the genetic architecture of complex adaptive traits evolved by modern human populations

Giulia Ferraretti

Giulia Ferraretti, Paolo Abondio, Marta Alberti, Agnese Dezi, Phurba T. Sherpa, Paolo Cocco, Massimiliano Tiriricco, Marco di Marcello, Guido Alberto Gnecci-Ruscone, Luca Natali, Angela Corcelli, Giorgio Mrinelli, Davide Peluzzi, Stefania Sarno, Marco Sazzini
Dept. of Archaeogenetics, Dept. of Basic Medical Science, Dept. of Biological, Dept. of Cultural Heritage, Dept. of Emergency and Organ Transplantation, Dolakha, Explora Nunaat International, Geological and Environmental Sciences, Germany (Germany), Interdepartmental Centre Alma Mater Research Institute on Global Changes and Climate Change, Italian Institute of Human Paleontology, Italy (Italy), Laboratory of Molecular Anthropology & Centre for Genome Biology, Leipzig, Max Planck Institute for Evolutionary Anthropology, Montorio al Vomano, Mount Everest Summitters Club, Nepal (Nepal), Neuroscience and Sense Organs, Ravenna Campus, Rolwaling, Rome, Teramo, University of Bari Aldo Moro, University of Bologna

Biological adaptations mediated by complex/polygenic traits are increasingly supposed to represent most of the adaptive responses evolved by human populations to cope with a range of selective pressures. However, studies in such a research field still lack specifically defined methodologies able to formally test a realistic approximation of the polygenic model of adaptation, with many approaches relying on results from Genome-Wide Association Studies, which are biased by well-known conceptual limitations. Here, we combined three different methods to set up an integrated pipeline of analyses capable to detect functionally related loci putatively implicated in the modulation of complex/polygenic adaptive traits. Particularly, we tested the reliability of such an approach by investigating the genetic bases of high-altitude adaptations evolved by Himalayan populations, which provide a case-study characterized by a well-understood selective pressure. For this purpose, whole genome data from individuals of Tibetan ancestry were first scanned with a likelihood haplotype-based method that discriminates between strong and weak selective signals. Distributions of the latter were then used to identify networks of genes significantly enriched in such signatures and strictly related from a functional perspective. Finally, we shortlisted the obtained results by focusing on those putative adaptive genes confirmed also by a supervised-machine learning algorithm trained on simulated data build on demographic parameters inferred for Tibetan populations. Accordingly, we pinpointed angiogenesis-related pathways as processes pervasively shaped by the action of natural selection in Tibetans and potentially related to their improved tissue blood perfusion, as previously proposed.
Inference of selection acting on coding sequences in the human lineage.
Hossameldin Loay

Hossameldin Loay, Donate Weghorn
Center of Genomic regulation CRG (Spain), Universitat Pompeu Fabra UPF (Spain)

Signs of selection within DNA sequences are usually detected using molecular clock models. However, these methods depend on certain intrinsic assumptions. First, they assume the same mutation rate across different genomic compartments and for different mutation types. Second, they often assume infinite sites: identical by state mutations are assumed to be identical by descent (no parallel mutations), and a mutated site is not allowed to re-mutate to a different state (no multiple hits). Sequencing more extant individuals of the human population shows that the infinite sites assumptions are violated, where more individuals harbor identical by state mutations from independent mutation events. Under the infinite sites assumptions, substitution rates (fixation rates) are equal to the underlying mutation rates. A difference between the two values is dependent on the rate of acquiring new mutations: population size multiplied by mutation rate. Here, we show that we can leverage the mutation rate variability across different genomic regions as well as different mutation types to have an estimate for an effective population size that can be used to correct for the bias induced by parallel mutations. Moreover, we correct for multiple hits without assuming detailed balance often assumed by some models like the HKY model. We use the corrected mutation rate values and the estimated effective population size to calculate gene-specific selection coefficients. We are also able to retrieve this information for specific historic evolutionary episodes that are not accessible by traditional methods.
The genetic population history of the last Himalayan hunter-gatherers

Inez Derkx

Inez Derkx, Simone A Biagini, Sudarshan Subedi, Prajwal Rajbhandari, Elena Bosch, Jaume Bertranpetit, Lucio Vinicius, Andrea Migliano
Research Institute for Bioscience and Biotechnology (RiBB) (Nepal), Tribhuvan University (Nepal), Universitat Pompeu Fabra (Spain), University of Zurich (Switzerland)

The Himalayas, with their complex terrain and diverse populations, have played a pivotal role in shaping human migration and adaptation in Eurasia. While extensive research has focused on various Nepali populations like the Sherpas, the Raute of western Nepal, a nomadic hunter-gatherer group of roughly 150 individuals, remain relatively unexplored. With their foraging lifestyle under increasing pressure, it is crucial to investigate their genetic structure and potential population history and origins. This study aims to shed light on the Raute\'s genetic structure and their place within the broader Himalayan genetic and historical context. We analysed newly genotyped data from the Raute and individuals from neighbouring settlements to examine patterns of consanguinity, population history, and recent genetic connectivity. Our results not only augment the existing linguistic and cultural narratives of the Raute from a genetic standpoint but also highlight the methodological challenges in researching small, consanguineous groups. Through this research, we contribute to the broader comprehension of the Himalayan region\'s intricate human history and the existence of hunter-gatherer societies within it.
Unraveling the Genomic Influence of Archaic Hominins on 3D Genome Interactions, Immune Pathways, and Gene Expression in Modern Humans

Isabela Alvim

Isabela Alvim, Matthew Leavesley, François-Xavier Ricaut, Nicolas Brucato, Murray Cox, Irene Gallego Romero
, Centre de Recherche sur la Biodiversi et l'Environnement, Massey University (New Zealand), School of Humanities and Social Sciences, St. Vincent's Institute (Australia), Universite de Toulouse (France), University od Aucklans (New Zealand), University of Papua New Guinea (Papua New Guinea)

Modern humans have inherited as much as 8% of their genome from extinct hominins, with the highest frequencies of Neanderthal and Denisovan DNA today found in the Pacific region. This ‘archaic‘ DNA is enriched for immune pathway genes and regulatory elements, and clusters non-randomly across the genome in a pattern characteristic of the co-associating domains that form during 3D folding of DNA in the nucleus. Yet how archaic DNA interacts in 3D space and thus influences gene expression remains poorly understood. To address this knowledge gap, we employ a multi-faceted approach: (i) we identify archaic DNA regions in the genomes of 12 individuals from Papuan New Guinea; (ii) subsequently, we empirically determined how archaic DNA aligns in the three-dimensional space of living immune cells from these individuals through Hi-C experiments; (iii) we then explore how this 3D structure influences DNA accessibility and gene expression by integrating ATAC-seq and RNAseq data. Finally, (iv) we compare the results to data from the same experiments performed with 6 African samples, which have no archaic introgression, and 6 European samples, which show lower levels of archaic introgression coming mostly from Neanderthals. This study contributes to a deeper understanding of our genetic inheritance from archaic hominins by shedding light on the complex interplay between archaic DNA, three-dimensional genomic organization, and gene regulation in immune cells.
A large pool of novel translated open reading frames is neutrally evolving in the human genome

Jiwon Lee

Jiwon Lee, Aaron Wacholder, Maria Chikina, Martin Jinye Zhang, Anne-Ruxandra Carvunis
Carnegie Mellon University (USA), Center for Evolutionary Biology and Medicine (USA), Joint CMU-Pitt Ph.D. Program in Computational Biology (USA), University of Pittsburgh School of Medicine (USA)

Human annotated proteins are translated from sequences spanned by open reading frames (ORFs) that are largely conserved. In contrast, recent advances in ribosome profiling have uncovered a vast number of unannotated human noncanonical ORFs (nORFs) that are evolutionarily novel, existing mostly in the primate lineage. Because they are generally short (<110 amino acids) and taxonomically restricted, nORFs have been missed by current functional annotations. Thus, they are largely uncharacterized, and their prevalence in the human genome is an open question. In this work, we expand the known set of nORFs by integrating 1,067 human ribosome profiling experiments, and investigate their evolutionary roles and contribution to phenotypes. Our study utilizes SNP information from TOPMed whole genome sequencing (WGS) data (N=132,345), and GWAS summary association statistics across 59 diseases and complex traits (avg. N=351,446). We identify 66,000 nORFs, significantly expanding upon nORFs reported in recent studies. We show that these nORFs have a nonsynonymous-to-synonymous SNP ratio close to 1 (adjusted pN/pS=0.976), suggesting they are evolving close to neutrally. By meta-analyzing LD score regression (LDSC) results across 59 nonredundant GWAS phenotypes with genetic correlation <0.1, we further find that nORFs are deficient in heritability, but provide independent information relative to the 52 functional annotations in the baseline-LD model (LDSC ?=−6.45±1.15x10−7, p=1.91x10−8). Though there is a growing number of case studies reporting nORFs with biologically important function, our results suggest that the bulk of human nORFs are evolving neutrally and have low contributions to complex traits and disease.
Genetic adaptation of Asian human populations to their environment

Johanne Adam

Johanne Adam, Romain Laurent, Nina Marchi, Raphaëlle Chaix, Evelyne Heyer, Laure Ségurel

Biometry and Evolutionary Biology Laboratory Intitute (France), Muséum National d'Histoire Naturelle - CNRS - Université Paris Cité (France)

Positive selection drives the spread of advantageous mutations in populations, contributing to phenotypic diversity across species. While many studies uncovered examples of human local adaptation, they mostly focused on Western European and Eastern Asian populations. This has left little knowledge about human genetic adaptation in other areas of Asia, such as Northern, Central and South-East Asia. Yet, there is an important cultural and environmental variability across these populations, and in humans, cultural practices and climatic constraints are known to result in local adaptation events. To address this gap, we studied a large genomic dataset of 894 individuals from 27 diverse Asian populations. Our goal is to provide a comprehensive analysis of adaptive pressures across the entire Asian continent, further incorporating publicly available genomes from East Asia and South Asia. Using iHS statistics, we conducted genome-wide scans for selection and investigated the most prominent signals shared across the Asian continent or across sub-regions. We further specifically quantified the relative importance of climatic and dietary variability in driving local adaptation. Our analysis led to the identification of novel genomic regions under selection, associated with farming, herding, or hunting-gathering practices. Additionally, we explored the interaction between climate and subsistence strategies in influencing selective events. This approach allowed us to pinpoint specific regions that were potentially involved in diet adaptation across diverse climates. We assembled gene sets based on the functions inferred from adaptive events identified in our study and applied selection enrichment analyses to uncover new cases of polygenic selection.
Genomic histories of the Adivasi and Sinhalese populations of Sri Lanka

Jose A Urban Aragon

Jose A Urban Aragon, Esha Bandyopadhyay, Constanza de la Fuente Castro, David Witonsky, Matthias Steinrücken, Amali S Fernando, Anjana H Welikala, Joanne T Kotelawala, Kamani H Tennekoon, Ruwandi Ranasinghe, Niraj Rai, Maanasa Raghavan

Birbal Sahni Institute of Palaeosciences (India), University of Chicago (USA), University of Colombo (Sri Lanka)

Sri Lanka, yielding the oldest-yet evidence of Homo sapiens in South Asia (~36,000 years ago), is central to our understanding of the early peopling of South Asia. Historically, Sri Lanka has been a vital node in the Indian Ocean trade network. Extensive migrations have, consequently, shaped the rich socio-cultural tapestry of this island nation. To investigate if the population history of Sri Lanka mirrors this socio-cultural diversity, we analyzed whole-genome sequence data from two present-day Sri Lankan populations: Adivasi (N = 19) and Sinhalese (N = 35). Based on some anthropological records, Adivasi are believed to be descendants of early inhabitants of Sri Lanka. Sinhalese are the most prevalent contemporary group in Sri Lanka, with oral traditions and chronicles suggesting that the Indo-European-speaking Sinhalese migrated from India in the first millennium BCE. We evaluated the genome-wide affinities of the Sinhalese and Adivasi with global populations, including the Sri Lankan Tamils (STU, 1000G Project), finding that all three Sri Lankan populations fall on the modern South Asian genetic cline and share high levels of genetic drift each other and with South Indian populations. Moreover, while the three Sri Lankan populations were modeled with genetic sources used to model other South Asians, our analyses suggest that the Adivasi have maintained a small long-term effective population size, reflecting a unique demographic history compared to the other two Sri Lankan populations. Overall, our study provides novel genomic insights on the population structure of Sri Lanka.
Decoding the evolution and functions of Human Accelerated Regions with deep learning

Katherine Pollard

Presented by self

Gladstone Institute of Data Science and Biotechnology (USA)

Human accelerated regions (HARs) are sequences that have been highly conserved through millions of years of vertebrate evolution and then changed dramatically in the human genome since divergence from our common ancestor with chimpanzees. This evolutionary signature suggests that HARs play important roles and that their functions may have been lost or changed in our ancestors, making HARs exciting candidates for understanding the genetic basis for what makes us human. However, it has been challenging to determine what HARs do and why the evolutionary forces constraining HAR sequences in other species suddenly changed in our lineage. In this talk, I will described updated methods for identifying accelerated regions in any lineage using large multiple sequence alignments and machine learning approaches that have shed light on the evolutionary histories of HARs. These modeling approaches are generating new hypotheses about the fastest evolving regions in the human genome, which we are testing using high-throughput genomic tools for functional characterization of non-coding sequences. This prediction-first strategy exemplifies my vision for a proactive, rather than reactive, role for data science in biomedical research.
Global patterns of Holocene natural selection
Laura L Colbran

Laura L Colbran, Iain Mathieson
University of Pennsylvania (USA)

Ancient DNA (aDNA) provides direct access to the evolutionary history of individual variants and has been used extensively to detect and classify occurrences of natural selection. These efforts have largely focused on European populations. Populations in other parts of the world share many selective pressures with Europeans, but have also been exposed to independent environments. The extent to which the response to Holocene selection is shared across populations is unknown. New aDNA from different parts of the world provides an opportunity to study the common effects of shared selective pressures and to identify novel selected loci and traits. In this study, we tested for selection in populations in Europe, East Asia, Africa and Central/South America. We modelled admixture proportions in 6,567 individuals from 12 ancient populations, and 1,228 individuals from 14 modern populations and tested for significant deviations using a maximum likelihood approach. In addition to known loci such as LCT and ADH1B, we identified novel loci such as DDB2 in Europe, Asia and the Americas and PDK4 in the Americas. We also adapted this approach to test for polygenic selection on complex traits and found that balancing selection on waist-hip ratio and directional selection on skin pigmentation are found on every continent. However other signals of directional selection on viral immunity in Eurasia, white blood cell counts in Africa and bone mineral density in Europe are more region-specific. This study represents the first comprehensive aDNA-based comparison of selective pressures across much of the inhabited world.
Functional characterisation of the regulatory activity of archaic DNA in Island Southeast Asia

Maddy Comerford

Maddy Comerford, Davide Vespasiani, Navya Shukla, Matthew Leavelsley, Christopher Kinipi, François-Xavier Ricaut, Nicolas Brucato, Irene Gallego Romero

CNRS, Center for Genomics, Evolution and Medicine, Health Services, Human Genomics and Evolution, Melbourne Integrative Genomics, School of Biosciences, School of Humanities and Social Sciences, St. Vincent’s Institute of Medical Research (Australia), University of Melbourne (Australia), University of Papua New Guinea (Papua New Guinea), University of Tartu (Estonia), Université de Toulouse (France), Walter and Eliza Hall Institute of Medical Research (Australia)

Interbreeding between modern humans and archaic hominins has contributed to the genomes of present-day human populations. Introgressed archaic variants are not uniformly distributed throughout the genome, but are enriched in regulatory regions, particularly in enhancers active in immune cell types, suggesting they have the potential to contribute to gene regulation and organism-level traits. However, due to the lack of diversity in genomic databases, research on the impacts of archaic introgression has focused mainly on the Neanderthal variants present in European populations, to the detriment of our understanding of the Neanderthal and Denisovan variants present in the rest of the world. Using a Massively Parallel Reporter Assay (MPRA), we functionally characterise the regulatory activity of over 26,000 Neanderthal and Denisovan variants present in modern-day Papuan populations. To prioritise evolutionary and phenotypically relevant variants, we focused primarily on variants located within accessible, active chromatin in immune cell types and segregating at allele frequencies > 0.15, which together suggest they might have been targeted by positive selection. By comparing the regulatory activity of these archaic variants and their non-archaic counterparts in three cell lines of different genetic background, we assess the contribution of archaic introgression in shaping modern human genetic diversity and demonstrate the value of MPRA as tools to study how gene regulatory changes have shaped phenotypes across evolution. Overall, this project will give insight into the regulatory activity of archaic variants in Island Southeast Asia and contribute to ongoing efforts to increase human diversity in genomics studies.
Rice is a staple crop in East Asia, and its domestication created one of the largest dietary and environmental changes in East Asia in the last 10,000 years. However, except for a few examples ((e.g. ADH1B Arg47His), our knowledge about human adaptation during rice domestication is limited. We aimed to reveal detailed human demographic history and adaptation in East Asia during the last 10,000 years using whole genome sequence data of ~7000 individuals of East Asian ancestry. We found substantial effective migration in East Asia mirroring major geographic barriers, using FEEMS. A population size increase after rice domestication was observed in some populations through IBDNe analyses. We also identified variants that may have undergone selection in East Asia during different time periods meaningful to rice domestication, using the methods Relate, ASMC and SDS. The variants selected during the last ~10,000 years and ~5,000 years were enriched in genes associated with alcohol-related traits and plasma polyunsaturated fatty acid (PUFA) levels, respectively. PUFA levels are known to lower cardiovascular disease (CVD) risk, and higher rice consumption is known to be associated with CVD risk factors such as Type 2 diabetes (T2D). We conducted a PheWAS in All of Us Research Program that demonstrated lower risk of CVD in East Asians compared to Europeans, despite their higher risk of T2D. We identify new relationships between genetic adaptations and modern disease risk in East Asians that may be influenced by rice domestication.
Modeling gene regulatory mechanisms contributing to the evolution of the human cerebral cortex
Marybeth Baumgartner

Marybeth Baumgartner, Yu Ji, Mark A Noble, Atreyo Pal, James P Noonan
Yale School of Medicine (USA)

The vast expansion of the human cerebral cortex distinguishes us from our primate relatives, and this cortical expansion is the foundation of uniquely human higher-order cognition. Numerous developmental innovations, such as increased proliferation of cortical progenitor cells, contributed to this cortical growth. Ultimately, these developmental innovations arose from genetic changes in the human lineage, which altered the molecular and cellular programs underpinning development. Efforts to illuminate these human-specific genetic changes have identified thousands of Human Accelerated Regions (HARs), highly conserved regulatory elements that exhibit a high rate of human lineage-specific sequence change, and Human Gain Enhancers (HGEs), regulatory elements with increased enhancer activity in the fetal human brain compared to that of non-human primates and mice. To understand how HARs and HGEs impact the events of corticogenesis, it is essential to model the effect of these elements on cortical development in vivo. While a growing body of evidence from cross-species regulatory genomics datasets and in vitro assays have implicated a subset of HARs/HGEs in human cortical development, this shortlist of candidate elements still numbers in the hundreds. Here, we present approaches to (i) identify HAR and HGE candidates with high predicted effect sizes, based on the increasing number of high-quality primate genomes, chromatin contact data, and functional genomics datasets available; and (ii) broaden the scope of in vivo model design beyond single HARs/HGEs, to accurately replicate the human regulatory landscape and thereby maximize the likelihood of recapitulating uniquely human molecular and developmental phenotypes.
Selective dynamics of interruptions at short tandem repeats

Michael E Goldberg, Harriet Dashnow, Kelley Harris, Aaron R Quinlan
Fred Hutchinson Cancer Research Center (USA), University of Utah (USA), University of Washington (USA)

Short tandem repeats (STRs) are hotspots of genomic instability that mutate at rates orders of magnitude greater than non-repetitive loci due to frequent replication slippage. Expansions at some STR loci are linked to Mendelian diseases, while variation at other noncoding loci may affect complex traits, possibly by altering transcription factor occupancy of nearby binding sites. Accordingly, some STRs are inferred to be under purifying selection, regardless of their instability. One or more ‘interruptions’, or bases that disrupt the locus’s canonical repeat, significantly decrease an STR’s mutation rate. However, interruptions may themselves be deleterious at constrained loci, particularly at noncoding loci in gene regulatory elements, possibly disrupting the formation of secondary structures key to their function.

We therefore hypothesized that the frequency of interruptions could depend on a locus’s purifying selection, where the fitness effects of both expansions but also interruptions could be more deleterious than at neutral loci. To test this hypothesis, we examined the distribution of interruptions at ~650,000 autosomal 2-6 bp motif STRs. We find that noncoding STRs under purifying selection harbor fewer interruptions than those evolving neutrally. In contrast, purifying selection is positively associated with interruptions in coding STRs. Our findings indicate that the abundance of interruptions may be partially explained at coding STRs by the benefit of a lower mutational burden at their linked loci. In contrast, maintaining a minimum core stretch of uninterrupted repeat may be critical for the function of noncoding STRs that fall within regulatory elements, outweighing the benefits of lowering mutation rate.
Archaic introgression in modern humans modelled by ancestral components
Moisès Coll Macià
Moisès Coll Macià, Mikkel Heide Schierup
Aarhus University (Denmark), Bioinformatics Research Centre (BiRC)

The geographical distribution of archaic DNA in modern humans provides crucial insights into our history of admixture with Neanderthals and Denisovans. However, grouping individuals by ethnicity or location does not always reflect a shared evolutionary history. Take South Americans, their genomes are the result of recent admixtures between divergent Native American, European and African lineages and so to understand their diversity it is crucial to account for the individual-level ancestry proportions. This applies to archaic components as well, since admixture with archaics occurred with the ancestral populations of extant groups. This project proposes a novel approach that models the archaic content of modern individuals with all their ancestral components jointly in a Bayesian framework. Here we analyse individuals from 1000GP and the HGDP (both called by gnomAD) by first obtaining the ancestral components and the archaic fragments with ADMIXTURE and hmmix, respectively. We then incorporate ancestral fractions as predictors in a multiple linear regression to model the Neanderthal and Denisova components of each individual. One intriguing finding that we detect is that the ancestry component which maximizes in South Asians is estimated to harbour similar Neanderthal total sequence than other ancesries, but distributed in much shorter fragments. This potentially hints at shorter generation intervals in their ancestral population over the past 50,000 years than any other Eurasian group. This work demonstrates that describing the archaic legacy in the light of ancestral components reveals a deeper understanding of past admixture events and the demographic history of our ancestral populations.
Spatial Genomic Scale and Determinants of Human Germline Mutation Landscape

Nathan Cramer

Nathan Cramer, Priya Moorjani
Berkeley (USA), University of California

Germline mutations are the ultimate source of genetic diversity within and across species. Understanding the rate and mechanisms by which mutations occur is of paramount importance for studies of medical genetics (to interpret the incidence of de novo and heritable diseases) and evolutionary biology (to date demographic and adaptive events). It is well established that mutation rates vary across the genome, from adjacent base pairs to whole chromosomes. However, the mechanisms underlying these patterns remain elusive. To understand the determinants of local mutation rate, we use low frequency single nucleotide polymorphisms from 76,156 whole genomes in the Genome Aggregation Database (gnomAD.) We focus on putatively neutral mutations by removing conserved regions and exons. We apply wavelet transform to characterize the spatial scale of variation in mutation patterns across the genome. Many factors correlate with local mutation rate variation, including replication timing, recombination rate, germline methylation, and base composition. We quantify the impact of each of these factors at various genomic size scales. By considering different types of mutations – accounting for 6 strand-independent base substitution types and CpG transitions – we explore the contribution of distinct mechanisms impacting the mutation spectrum. We then apply this approach to study mutation rate differences across sexes and populations. Together, this analysis helps quantify the contribution of various genomic features and biochemical processes impacting the spatial landscape of mutation rate in humans.
Mainland Japanese population has distinct three ancestral components from the Jomon hunter-gatherers, the Yayoi farmers, and the latter migrants from the East Asian continent. Population differentiation in the Japanese population has occurred through demographic changes and various environmental changes, including both environmental and cultural transitions resulting from admixture. Focusing on the effect of positive selection, we identified positively selected haplotypes, and then estimated the onset time and the intensity of selection coefficient, in order to assess how selection has shaped the genetic characteristics of the modern Japanese. We detected genome-wide candidate SNPs of positive selection, based on the difference in allele frequencies between Japanese and a genetically close East Asian population. Of them, we examined signatures of ongoing/completed positive selection against each haplotype with >10% frequency by using 2D-SFS, which evaluated the reduction of IAV in the focal haplotype. Here, we present a signature of ongoing positive selection on a haplotype nearby SLC8A1 in the Japanese with a frequency 20% higher than those in other East Asian populations. This haplotype has been selected since 18 kya with selection coefficient of 3%, whereas we could not detect any signatures of selection on any haplotypes in other East Asian populations. This result suggests that selected haplotypes are associated with recent population differentiation within East Asia. In the next step, we plan to classify genomic region into three ancestry-specific derived segments and discuss origins of selected haplotypes, in order to reconstruct the transition of adaptive phenotype in the Japanese.
Pakistan is the fifth most populous country in the world, home to a vast cultural diversity represented in multiple languages and cultural practices embedded in ethnic identities, yet remains under-studied genetically. We analyzed 9 Pakistani ethnic groups from the newly released joint call of the Human Genome Diversity Project and 1000 Genomes Project, a total 279 individuals from Brahui, Balochi, Burusho, Hazara, Kalash, Makrani, Pathan, Punjabi and Sindhi groups. We infer fine-scale genetic structure across Pakistan and in relation to global groups. Amidst the results, Makrani showed an enrichment of African ancestries. To further explore this finding we infer the timing of arrival of African groups and their specific origins using Gnomix, tracts and MAAS-MDS with an African reference panel. Further, we inferred natural selection per SNP and gene using the Population Branch Statistic (PBS). Focusing on the top 99th percentile of PBS values, we found unique as well as shared genes under natural selection across ethnic groups. For example, we found genes like CTXN2 and SLC24A5 (associated with skin pigmentation) highly selected in Pakistan as a whole compared to Bengali and Yorubans, as well as the APPAT lncRNA, associated with atherosclerosis, highly prevalent in Pakistan. We observe the highest PBS value for LINC00319 in Makrani, Kalash and Brahui groups (also in top 99th-percentile of all groups except Punjabi) that is related to 9 cancer types. Another interesting lncRNA found in Balochi, Brahui and Sindhi, LINC00313, is related to cancer but also to osteoarthritis development inhibition.
Discovering the cis-regulatory basis of archaic human-derived phenotypes

Ryder Easterlin

Ryder Easterlin, Simon Fishilevich, Katharina Lange, Nadav Ahituv, David Gokhman

UC San Francisco (USA), Weizmann Institute of Science (Israel)

Our understanding dictates that the phenotypic differences between modern humans and our closest evolutionary relatives, Neanderthals and Denisovans, is largely attributable to genetic variation in noncoding regions of the genome. These regions, which contain enhancers and promoters, play crucial roles in gene regulation. Despite this, the precise impact of lineage-specific changes in noncoding regions on human evolution remains poorly understood. In this study, we utilized massively parallel reporter assays to investigate the influence of genetic variants in putative cis-regulatory elements on the phenotypic divergence between modern humans and archaic humans (Neanderthals and Denisovans). Leveraging high-coverage archaic human genomes from three Neanderthals and one Denisovan, along with catalogs of modern human genetic variation, we identified 57,403 fixed single-nucleotide variants that originated in the archaic lineage. For each variant, we synthesized two sequences: one with the archaic-derived allele and another with the ancestral allele. By testing the effects of each sequence version on gene expression in key cell types relevant to human divergence—osteoblasts, skin fibroblasts, adipocytes, and excitatory neurons—we identified thousands of active regulatory sequences containing archaic-derived variants. Furthermore, we found that hundreds of these variants drive differential gene expression. We pinpointed the genes likely affected by these cis-regulatory changes, shedding light on the organs, biological processes, and phenotypes impacted by these regulatory alterations. Overall, our study offers a comprehensive catalog of the gene regulatory effects of archaic-derived variants in critical cell types, providing valuable and functionally supported insights into recent human evolution.
Worldwide patterns of diversity at the 17q21.31 locus in modern human genomes

Samvardhini Sridharan, Peter James Sudmant
Berkeley (USA), University of California

Among the known 150 large inversions segregating in human populations, one extraordinary example is at the 17q21.31 locus which spans >900 kilobases and comprises ten genes. These genes include the medically relevant MAPT and KANSL1. Several major inversion haplotypes of this locus segregate in modern human populations, each exhibiting distinct frequencies across geographic regions. Some of these haplotypes have been identified as being under selection and linked to fecundity. All of these haplotypes maintain a complex duplication architecture in addition to its inversion status. Consequently, this locus provides a unique system to study the impact of copy number variation on recombination suppression and extended linkage disequilibrium in human populations. Here, we report extensive patterns of temporal and geographic variation at the 17q21.31 locus in several primates and in different modern human populations. We find that while the inverted haplotype (H2) is the ancestral haplotype in humans, it is the less common haplotype in modern-day human samples. We observe an exceptional amount of copy number diversity in direct orientation (H1) haplotypes globally, particularly in understudied modern populations including South Asians. We hypothesize that recombination suppression in large and structurally complex regions of the genome may contribute to the differential accumulation of deleterious mutations. Finally, we observe several unique instances of ancient double recombination events in African populations and date the timing of these events. Together, these results provide valuable insights into the role of chromosomal rearrangements in shaping evolution, diversity, disease susceptibility, and fitness.
Is there any population differentiation at particular SNPs in the human CYP1A2 gene?

Sayaka Chiku

Sayaka Chiku, Jun Gojobori, Yoko Satta
SOKENDAI (Japan)

Human CYP1A2 is a detoxification enzyme that metabolizes xenobiotics such as drugs (caffeine is a well-known example) and is expressed only in human liver. Although CYP1A2 expression and activity levels vary among individuals and populations, the evolutionary background of CYP1A2 polymorphism has not been explored. Therefore, the purpose of this study is to investigate whether certain CYP1A2 SNPs are positively selected. In this study, after LD pruning for quality control, the remaining 215 SNPs from the vcf of 1KGP GRCh37 were used for estimating their frequency and FST among 1KGP populations. The derived allele frequencies at these SNPs showed that they were divided into three characteristics: 136 SNPs occur in “one population”, 73 SNPs in “some populations” include being present only in a particular superpopulation, and 6 in “all populations”. The estimated frequencies show the presence of population- or regional-specific SNPs, some of which are in LD with specific SNPs. Based on the estimated FST values, a total of 7 SNPs were in the top 2.5%. rs762551 and rs57295890 have high FST values between Peru and 9 and 24 populations, respectively. Although both SNPs exist in all populations, the observed high FST values due to unique SNP frequencies in Peru: rs762551 is 36.6% on average but 13.5% in Peru, whereas rs57295890 is 16.4% on average but 68% in Peru. We will examine what makes unique SNP frequencies in Peru by using statistical methods such as Tajima’s D, iHS and 2DSFS.
Uncovering the genetic architecture and evolutionary roots of androgenetic alopecia in African men

Ujani Hazra, Rohini Janivara, Aaron Pfennig, Timothy Rebbeck, Akindele O Adebiyi, Ilir Agalliu, Joseph Lachance

Albert Einstein College of Medicine (USA), Dana-Farber Cancer Institute (USA), Georgia Institute of Technology (USA), University of Ibadan (Nigeria)

Androgenetic alopecia or male pattern baldness (MPB) is a highly heritable trait. However, much of our understanding of the genetics of MPB comes from individuals of European descent. Examining a novel dataset comprising 2,136 men from Ghana, Nigeria, Senegal, and South Africa that were genotyped using a custom genotyping array, we conducted the first African GWAS of MPB and inferred the evolutionary history of genetic variants associated with this complex trait. We found evidence that the genetic architecture of MPB varies across continental populations and identified 51 independent associations in Africa (p-value < 10⁻⁵, r² < 0.2). Although Neanderthal alleles have previously been associated with skin and hair phenotypes, we did not find enrichment for signatures of ancient introgression in European-ascertained baldness hits. Next, we tested for signatures of polygenic selection on MPB using integrated Haplotype Scores. Regardless of the ascertainment scheme, autosomal SNPs were not enriched for outlier iHS statistics. However, X-chromosomal SNPs (rs12558842, rs2497911, rs1204041) near the EDA2R and AR genes were found to have large allele frequency differences between continents. These variants were in linkage disequilibrium and multiple divergent haplotypes were found in high frequency in Africa and Europe. Although these patterns are consistent with the Xq12 region having undergone multiple independent selection events we note that Xq12 is located near the centromere in a genomic region with low recombination rates. Because of genetic hitchhiking, divergent haplotype frequencies at Xq12 need not have been caused by selection acting directly on male pattern baldness.
Curbing the accumulation of deleterious mutations: the roles of weak epistasis and compensatory beneficial mutations

Ulises Hernández Martín del Campo

Ulises Hernández, Joseph Matheson, Walid Mawass, Joanna Masel
University of Arizona (USA), University of California San Diego (USA)

Under realistic human genome-wide deleterious mutation rates Ud>1, new deleterious mutations appear faster than selection can purge them. This can cause fitness decline and eventual population extinction. Using a novel simulation framework that efficiently handles genome-wide linkage disequilibria across many segregating sites, we found that large effect beneficial mutations can compensate for the accumulation of many smaller-effect deleterious mutations. Indeed, populations might even adapt faster following increased mutation rate. However, in the absence of epistasis, the number of beneficial mutations required can exceed Haldane's "cost of selection" limit created by finite reproductive excess. When mutations have larger fitness effects in individuals of lower fitness (global negative epistasis), the variance in fitness increases as the population degrades, increasing the rate at which selection purges mutations, sometimes making the rate of further deleterious fixations negligible. We combined expressions from the Price equation with expressions from a one locus two allele finite Markov chain model to derive a novel solution for mutation-selection-drift metastable balance under high Ud and global epistasis. Very weak negative epistasis is sufficient for population metastability in large but not in small populations. Indeed, weak epistasis combined with rare large effect beneficial mutations allow population persistence even at Ud=10. Given humans' high Ud>2, conventional models suggest that we are on the edge of continual genome degradation - our results suggest that under weak global epistasis and beneficial substitutions, human populations might even tolerate further increases in mutation rate.
Dynamic Rates and Patterns of Nucleotide Substitutions in Ape Telomere-to-Telomere Genomes: Substantial Effects of Sex Chromosomes

Xinru Zhang

Xinru Zhang, Saswat K Mohanty, Robert S Harris, Francesca Chiaromonte, Kateryna D Makova, The T2T Primates Consortium

Center for Medical Genomics, Department of Biology, Department of Statistics, Huck Institutes of the Life Sciences, National Human Genome Research Institute, National Institutes of Health (USA), The Pennsylvania State University (USA), University of Washington School of Medicine (USA)

Mutation is a major evolutionary force that generates genetic diversity for the other evolutionary forces to operate on. Mutation rates vary among species, over time, and across different genomic regions. DNA sequence contexts also substantially affect mutation rates. Yet, the extent of how sequence context influences nucleotide substitutions remains understudied. Leveraging the recently released Telomere-to-Telomere (T2T) whole-genome assemblies for most extant great ape species, we were able to construct the most comprehensive tri-nucleotide substitution spectrum in apes. This has allowed us to compare nucleotide substitution spectra across seven ape species and examine its variation among different phylogenetic branches and genomic locations. Our analyses revealed chromosomal heterogeneity in the distribution of substitutions. Specifically, chromosomes Y and X demonstrated unique substitution patterns, as compared to those for the autosomes. Comparing nucleotide substitution spectra between the two sex chromosomes, we found transversions to be significantly more abundant on the Y than on the X, and transitions vice versa. These findings are broadly consistent with sex-specific signatures of de novo mutations from other studies; C>A, C>G and T>G were shown to be enriched in paternal de novo mutations, whereas C>T mutations—in maternal de novo mutations. C>G might be related to meiotic double-strand breaks in the male germline. Additionally, we compared substitution rates and spectra in the X and Y pseudoautosomal regions (PARs) and on the autosomes. Our results allow us to formulate hypotheses about molecular mechanisms of mutagenesis that should be tested in subsequent wet-lab experiments.
Comparing Neanderthal introgression maps reveals substantial heterogeneity across algorithms, populations, and assumptions

Yaen Chen

Yaen Chen, John A. Capra
San Francisco (USA), University of California

Motivation: Statistical methods to identify Neanderthal ancestry rest on different assumptions and inputs. Nonetheless, most studies of Neanderthal introgression use only a single method. Due to a lack of “ground truth,” we have a limited understanding of comparative strengths, weaknesses, and accuracy of these methods. Results: We performed large-scale comparisons of introgressed variant maps from 12 representative Neanderthal introgression detection algorithms. These span methods using archaic and African human reference genomes (ArchaicSeeker2, CRF, DICAL-ADMIX), only archaic genomes (S*, Sprime, HMM, SARGE, ARGWeaver-D), only African human reference genomes (IBDmix), or neither (ArchIE). Overall, ~66% of the genome is called as introgressed by at least one method, and 15.7% of introgressed variants are unique to a method. We identify high pairwise overlap between most methods, excluding ArchIE and ARGweaver-D, which identified substantially less introgression. Hierarchical clustering reveals that algorithms built with similar motivations do not always have the most similar predictions. Regions unique to a single method often show enrichment for known Neanderthal and modern human differences: for example, loci unique to ArchIE are significantly enriched for long clavicle formation. Thus, unique predictions may not reflect errors, but rather different strengths of the methods. Loci identified by all 12 algorithms are significantly enriched for iron homeostasis, motor coordination, and skeletal phenotypes. Conclusion: Our results highlight substantial heterogeneity in commonly used Neanderthal introgression maps. We make integrated prediction sets available to enable further understanding of the legacy of Neanderthal introgression in modern humans and ensure robustness of future results.
Polygenic adaptation leads to a higher reproductive fitness of native Tibetans at high altitude

Yaoxi He

Yaoxi He, Bing Su
Chinese Academy of Sciences (China), Kunming Institute of Zoology

The adaptation of Tibetans to high-altitude environments has been studied extensively. However, the direct assessment of evolutionary adaptation, i.e. the reproductive fitness of Tibetans and its genetic basis remain elusive. Here, we conduct systematic phenotyping and genome-wide association analysis of 2,252 mother-newborn pairs of indigenous Tibetans, covering 12 reproductive traits and 76 maternal physiological traits. Compared to the lowland immigrants living at high altitudes, indigenous Tibetans show better reproductive outcomes, reflected by their lower abortion rate, higher birth weight, and better fetal development. The results of genome-wide association analyses indicate a polygenic adaptation of reproduction in Tibetans, attributed to the genomic backgrounds of both the mothers and the newborns. Furthermore, the EPAS1-edited mice display higher reproductive fitness under chronic hypoxia, mirroring the situation in Tibetans. Collectively, these results shed new light on the phenotypic pattern and the genetic mechanism of human reproductive fitness under extreme environments.
S6 - Molecular evolution through metagenomics.
Host adaptation and genomic change in intestinal eukaryotes revealed from sequencing microbial mixtures

Abigail Lind

Abigail Lind, Nathan McDonald, Elias Gerrick, Ami Bhatt, Katie Pollard
Georgia Institute of Technology (USA), Gladstone Institutes (USA), Stanford University (USA), Stanford University School of Medicine (USA)

The most prevalent microbial eukaryote in the human gut is Blastocystis, an obligate commensal protist also common in other vertebrates. Yet it is unclear how Blastocystis adapted from a free-living stramenopile ancestor to thrive within its wide range of hosts. To enable comparative genomics interrogation of changes throughout Blastocystis evolution, we cultured and sequenced six strains spanning the phylogenetic and host diversity of the genus. As establishing axenic cultures of Blastocystis is failure-prone and laborious, three strains were grown xenically alongside other microbes. To overcome the challenge of generating contiguous and contamination-free genomes from xenic cultures, we established a metagenomic assembly pipeline utilizing long-read, short-read, and Hi-C DNA sequencing, plus genome annotation leveraging RNA sequencing. Investigating differences in gene content and genome organization, we observe a transition within human Blastocystis towards minimal introns and intergenic regions, loss of canonical stop codons, gain of genes related to host temperature, and diversification of subtype-specific host-interfacing genes. In contrast, Blastocystis isolated from an herbivorous tortoise contained plant carbohydrate metabolizing enzymes, some horizontally acquired from bacteria, likely reflecting fermentation within the host gut. Finally, our analyses indicate that the common ancestor of all Blastocystis lost nearly all ancestral genes for flagella, resulting in its spherical morphology. Our findings reveal adaptations in different Blastocystis lineages for life in the gut of diverse vertebrates. The long-read hybrid assembly methods used here are effective tools for establishing genomics of eukaryotes recalcitrant to axenic culture.
The herbicide glyphosate inhibits EPSPS enzyme involved in shikimate biosynthesis in plants, bacteria, archaea and fungi. This herbicide is applied in higher doses to those recommended in Yucatan. To identify the effects of glyphosate doses variation in bacteria and fungi communities of a black Leptosol, we conduct an exploratory greenhouse experiment and environmental DNA using a metagenomic approach. The DNA was extracted from soil treated with commercial dose (x), 0.5x, 1.5x and 0x glyphosate samples. The V3-V4 16S rRNA and ITS2 regions were amplified and sequenced using NovaSeq 6000 Illumina. QIIME was used to process uncut sequences, unmerged, classified and quantify different taxa among samples. Phyloseq and DESeq2 was used to compare the microbiota enrichment of glyphosate treated samples. Haloferula helveola (bacteria) and Arxiella dolinchedrae (fungi) were enriched in 1.5x glyphosate tratment. Those species are common in soils, and if they are glyphosate resistant maybe used in bioremediation processes. Colletotrichum truncatum (fungi) was also enriched in 1.5x treatment but is potentially pathogenic. A low enrichment was observed in 1x doses with no taxa remarkable in both cases. Future studies, including isolation of colonies by selective cultures and metabolomics and shotgun sequences would help to understand the species potential in bioremediation projects and their function in the soil health.
Magellanic penguin feather-dependent ectosymbiont bacteria discovered by genomes assembled from metagenomes

Brenda Paulina Vázquez Ramírez

Brenda P. Vázquez, Manuel Ochoa-Sánchez, Ulises Rodriguez-Cruz, Valeria Souza
Institute of Ecology, Licenciatura en Ciencias Genómicas, UNAM (Mexico)

The microorganisms associated with the Magellanic penguin plumage are believed to be acquired through the environment with which they come into contact, which would imply that they do not depend on the penguin to subsist. An alternative interpretation is that the microbiome is feather-dependent and therefore characteristic of the penguin species. To discern between these possibilities, assembled genomes were obtained from shotgun metagenomic samples and subsequently analyzed for the metabolic pathways present in the microorganisms found. All genomes had at least one essential pathway absent. These results suggest a Magellanic penguin-dependent ectosymbiont microbiome.
Genomic characterization of novel Treponema species from the oral microbiome of Aboriginal Australians

Davide Bozzi

Davide Bozzi, Samuel Neuenschwander, Gabriel Yaxal Ponce-Soto, David M Lambert, Eske Willerslev, Nicolas Rascovan, Anna-Sapfo Malaspinas

Griffith University (Australia), Institut Pasteur (France), Swiss Institute of Bioinformatics (Switzerland), University of Bremen (Germany), University of Cambridge (United Kingdom), University of Copenhagen (Denmark), University of Lausanne (Switzerland)

Metagenomic analyses, over the last decade, have yielded a wealth of genomic information regarding microbial communities. The assembly of novel genomes from metagenomic data is rapidly outpacing the culture-based approaches in generating new genetic knowledge about unknown microbial species. In this study, we focused on Treponema, a genus of bacteria encompassing various environmental and host-associated taxa, including human symbionts and pathogens. We investigated novel high-quality metagenome-assembled genomes (MAGs) from the genus Treponema identified within the oral microbiome of Aboriginal Australian populations. Some of these MAGs represent species not found in previously studied human populations. We characterized their functional potential and we compared them from a phylogenetic and functional perspective to other species of the same genus to gain insights into their evolution and adaptation processes.
Recently, we have discovered amazing relations between the microbiome and the evolution and ecology of their hosts. For instance, the influence of the gut microbiome in health and behavior of non-human primates. Despite this, little is known about the composition of the gut microbiome in wild populations of New World monkeys. México has three species of wild primates: two howler monkeys (Alouatta palliata, A. pigra) and one spider monkey (Ateles geoffroyi). Howlers are a model for gut microbiome studies in neotropical primates. Their microbiome varies naturally and is influenced mostly by age, diet, and seasonality. Notably, there is rather little information for Mexican howler populations and less so for spider monkeys. We evaluated the diversity and abundance of gut bacterial communities of 155 individuals from wild populations of these three species, encompassing most of its distribution range. We extracted DNA from fresh stool samples and amplified and analyzed the 16S rDNA V4 region. Results suggest that phylogeny is the principal factor in the structure of these species gut microbiome, while diet and geographic distribution are also important. The most abundant microbial Phyla are shared among species but differing at finer taxonomic levels. Ateles geoffroyi has the most differentiated microbial communities, harboring more rare taxa. We also identified significant differences in relation with sex in Alouatta pigra, but not in A. palliata. The gut microbiome abundance and diversity patterns per species are consistent with their diet. Our study provides the first overall microbiome data for these three Mexican primate species.
Nitrogen cycling in symbiotic bacterial communities associated with ancient American plants

Edder Daniel Bustos-Diaz

Edder Daniel Bustos-Diaz, Francisco Barona-Gomez, Angelica Cibrian-Jaramillo
CINVESTAV Irapuato (Mexico), Leiden University (Netherlands), Leiden University (Netherlands), Naturalis Biodiversity Center (Netherlands)

Both the angiosperm family of Gunneraceae and the gymnosperm order of Cycadales can form nitrogen fixing symbiotic relationships with Nostocales cyanobacteria using specialized organs. Multiple studies, mostly on cycads, have shown that other bacterial taxa can also be found in the symbiotic organ, but not much information is known about the functional role of these associated bacterial communities. To address this, we sampled multiple cycads and Gunnera plants at different locations in order to inoculate co-cultures which could sustain the bacterial communities. These co-cultures where then sequenced through shotgun metagenomes in order to obtain metagenome assembled genomes (MAGs). Using these we were able to identify Hyphomicrobiales in the bacterial communities from different plants and locations are capable of partial denitrification, similarly to what has been found in the bacterial communities of the fern Azolla, another cyano-centric symbiotic system. Moreover, the phylogenetic analysis of these MAGs suggest that these Hyphomicrobiales could be under co-selection along with the Nostocales for the symbiosis, which might help to explain why some of these bacterial cyano-centric bacterial communities can jump from host to host with relative ease, even though each systems have their own particular symbiotic organs. The shared taxonomic and functional signals found in these communities suggest that even though Nostocales are key players in the symbiosis, the communities also play a role in the success of these symbioses.
Phylogeny-aware strain profiling limits false positive detections and quantifies divergence time of novel strains

Evan B Qu, Jacob S Baker, Laura Markey, Tami D Lieberman
Massachusetts Institute of Technology (USA)

Reference-based strain profiling methods are well-understood to be limited in the event of novel diversity but provide advantages over reference-free methods in samples with low coverage and multiple strain colonization. We present PHLAME, a novelty-inclusive method to profile bacterial strains up to their most recent shared ancestor along a reference phylogeny. PHLAME leverages the read distribution across clade-specific SNPs and a Bayesian mixture model to distinguish between lowly abundant strains represented in the reference set and highly abundant strains that diverge from the known phylogeny. By masking known clades from the phylogenies of different bacterial species, we show that posterior estimates of divergence time inferred from metagenomics track well with true branch lengths, indicating that many fixed mutations are acquired sequentially along branches of a tree. Using a ground-truth dataset of 4,055 isolate genomes and 33 metagenomes generated from the same samples, we show that existing strain profiling methods infer more false than true positives as the number of reference genomes and profiling resolution increases, whereas PHLAME limits the false detection rate to <0.05 across profiling resolutions while retaining comparable recall to other methods. We demonstrate the utility of PHLAME in detecting novel phylogenetic associations using public skin metagenomes, including the identification of recently emerged clades of C. acnes that are geographically restricted yet highly prevalent, suggesting limitations to the rate of spreading of this bacteria. PHLAME will enable researchers to conduct rigorous epidemiology and evolution studies from rapidly accruing metagenomic data.
Clues to the Evolution of Antibiotic resistance from bacteria and ancient communities from pristine sites in Cuatrociénegas, Coahuila, Mexico.

Gabriela Olmedo Álvarez

Presented by self
Cinvestav, Unidad Irapuato (México)

Our research on antibiotic resistance takes a unique approach, revealing how community competitive interactions have influenced microbial evolution. We leverage the ecological dynamics of bacteria and metagenomics from ancient microbial communities of Cuatrociénegas, Coahuila, a pristine site in Mexico, to offer new insights into the applied aspects of the One Health concept. Unlike most studies that focus on a narrow taxonomic window of human pathogens shaped by antibiotic selection, we step outside the clinical framework to explore the origins of antibiotic synthesis and selection processes in microbes. We analyzed a set of 78 bacteria, identifying competitive traits that drive the evolution of antibiotic synthesis and resistance. Through genomic analysis, we unveiled a constraint in horizontal gene transfer of antibiotic resistance genes (ARGs) across different bacterial lineages. Furthermore, we discovered in the Bacillacea a large set of ß-lactamases that were thought to be restricted to Gammaproteobacteria. Phenotypic and genotypic associations between antibiotic resistance and ARGs further highlight these dynamics. Additionally, our metagenomic analysis of microbial mats and stromatolites, coupled with observations of abundant resistance in cultured bacteria, supports the hypothesis of the evolution of antibiotic synthesis and resistance dating back to ancient microbial communities. By retelling the story of antibiotic resistance from the competitive interactions within microbial communities, we emphasize the pivotal role of microbial interactions in shaping ARGs content. This perspective offers a deeper understanding of antibiotic resistance evolution and provides valuable insights for combatting this global health challenge.
The evolution of Arabidopsis thaliana-associated Pseudomonas and Sphingomonas

Haim Ashkenazy

Haim Ashkenazy, Or Shalev, Talia Karasov, Derek Lundberg, Manuela Neumann, Detlef Weigel

Current address: Swedish University of Agricultural Sciences (Sweden), Max Planck Institute for Biology Tübingen (Germany)

Members of the bacterial genera Pseudomonas and Sphingomonas are often highly abundant in the leaf-associated plant microbiota. While some species or isolates have been described as pathogenic, others are beneficial to host plants or merely commensals. In this study, we focus on the evolution of a local collection of plant-associated Pseudomonas and Sphingomonas strains, to better characterize the factors governing plant colonization by members of these taxa. We collected 1,524 Pseudomonas isolates and 425 Sphingomonas isolates from wild Arabidopsis thaliana plants, over several years and sites near Tübingen, Germany. We sequenced and assembled the genomes of these bacterial isolates and inferred their pan-genome, based on ortholog groups. Out of the 72,397 and 56,923 ortholog groups specified for Pseudomonas and Sphingomonas, respectively, only 1.3% and 6.9% belonged to the bacterial 'core genome', while 36.3% and 40.7% were unique to individual isolates - suggesting high variability in isolates’ gene content. Next, we grouped closely-related strains (determined by their genome-wide average nucleotide identity, ANI) and compared the gene content within and between groups. By comparing these closely-related isolates, we discovered genomic islands that varied among isolates and we characterized their function and evolutionary dynamics. We further identify genomic islands shared between Pseudomonas and Sphingomonas isolates. Taken together, our results demonstrate that focusing on the evolution of closely-related strains, which are co-colonizing the same plant population, can shed light on the complex host-microbe and microbe-microbe interactions shaping the plant microbiome.
De novo assembly of ancient human metagenomes improves community diversity estimates and reveals varied microbial taxon-specific evolutionary trajectories

Irina M Velsko

Irina M Velsko, Keri Burge, Alexander Hübner, I-Ting Huang, Colleen Cavanaugh, Christina Warinner
Harvard (USA), Harvard University (USA), Max Planck Institute for Evolutionary Anthropology (Germany)

Ecological diversity estimates for ancient oral microbiomes have appeared highly similar to those of modern oral microbiomes when profiled with large and diverse taxonomic databases. However, artifacts in the profiling results have hinted at the presence of microbes that remain undetectable because they have no representative genome sequences in the databases used for taxonomic profiling. Hence, diversity estimates for ancient human microbiomes may be artificially low, potentially masking signals of evolutionary change and muddying comparisons to their modern counterparts. To improve our ability to detect and investigate oral microbiome evolution, we directly reconstructed genomes from modern (n=42) and ancient (n=455) oral metagenomes with de novo assembly to generate metagenome-assembled genomes (MAGs, n=8696 medium/high-quality) spanning 100,000 years of human history and from globally diverse populations on six continents. These MAGs reveal substantial undescribed microbial diversity present across ancient and modern human microbiomes (n=562 unnamed taxa), and enable us to better resolve microbiome community composition and to correlate community-level changes with human cultural, social, and demographic history. In addition, specific investigation of individual taxa reveals that phylogenetic patterns mirroring human migratory patterns are taxon-specific, and are stronger in novel taxa than clinical taxa that are well-described, as well as stronger in taxa that are rarely detected in heavily studied populations in Europe and the US. These results highlight the utility of incorporating ancient metagenomic data into human microbiome studies for revealing evolutionary signals at the levels of individual taxa as well as the entire microbiome.
Prokaryotic genomes constantly undergo gene flux via lateral gene transfer, generating a pangenome structure consisting of a conserved core genome surrounded by a more variable accessory genome shell. Over time, flux generates change in genome content. Here we measure and compare the rate of genome flux for 5,655 prokaryotic genomes as a function of amino acid sequence divergence in 36 universally distributed proteins of the informational core (IC). We find a clock of gene content change. Long-term average rate of gene content flux is remarkably constant across all higher prokaryotic taxa sampled. A linear regression predicts between 0% (Chlamydiae) and 30-33% gene content differences (Alphaproteobacteria, Gammaproteobacteria, Clostridia) for conspecific genome pairs having identical IC sequences, corresponding to the accessory genome proportion for well-sampled species. The data suggest that pangenome structure is a general feature of prokaryotic genomes and that it has been in existence since the divergence of bacteria and archaea.
The Influence of Aspergillus Starter Cultures on Microbes Found in Miso
Kimberly Lynette Acevedo

Kimberly L. Acevedo, John G. Gibbons
University of Massachusetts Amherst (USA)

This research explores the intricate interplay within the microbial community of miso fermentation, focusing on the co-culture of A. sojae and challengers, a combination of various different prokaryotes and Saccharomyces cerevisiae. Aspergillus sojae (AS), renowned for its proteolytic prowess in soy fermentation, serves as a parallel model to A. oryzae/A. flavus, presenting an intriguing hypothesis of AS as a distinct population within Aspergillus parasiticus (AP), adapted to the food environment. This study investigates species interactions between wild and domesticated Aspergillus starter cultures and the challengers commonly found in miso fermentation. Through pair-wise dual assays, we measure the growth and inhibitory effects of Aspergillus strains when co-cultured with these challengers. Identification of differences between domesticated and wild cultures prompts a dual-RNA-seq analysis in strains displaying variations. Our hypothesis posits that the absence of secondary metabolite production in A. sojae will lead to heightened growth of challengers in co-culture. This study provides a glimpse into the unexplored territory of microbial dynamics in miso fermentation. Join us at the poster presentation to unravel the collaborative journey of A. sojae and challengers, shedding light on the sensory and physical characteristics of the final fermented product. The outcomes not only contribute to our understanding of microbial interactions but also offer insights into optimizing miso production and enhancing the nuances of this cherished fermented food.
COMPARISON OF THE MICROBIOTA OF A RED SOIL UNDER DIFFERENT DOSES OF GLYPHOSATE.

María alejandra Ocaña Ek

María Alejandra Ocaña, Miriam Monserrat Ortega, Hector Medina
Universidad Autonoma de Yucatan (Mexico), Universidad Autonoma de Yucatán (Mexico)

Glyphosate inhibits the enzyme 5-enolpyruvyl shikimate-3-phosphate synthase (EPSPS), present in plants, bacteria, fungi, and protozoa. This enzyme is vital in these organisms due to its involvement in amino acid biosynthesis and metabolic processes. The exposure to glyphosate causes metabolic disruption in several soil microbial communities. To evaluate the effects of this herbicide on the microbiota, an exploratory study was conducted using a metagenomic approach. Glyphosate was applied using a commercial dose (x), 0.5x, 1.5x and 0x on pots with red Leptosol in a open greenhouse. DNA was extracted and amplified for Illumina sequencing of V3-V4 16SRNA and ITS2 regions. QIIME was used to process uncut sequences, unmerged, classified and quantify different taxa among samples. Phyloseq was used to compare the microbiota enrichment of glyphosate treated samples with Deseq2. The enriched bacteria species for the 1.5x dose of glyphosate were Clostridium paraputrificum, Gemmiger formicilis, Catenibacterium mitsuokai that has been found in vertebrates species and are potentially pathogens. Enriched fungi were for the 1.5x dose of glyphosate were Trichoderma asperellum, and unknown species from the genus Trechispora and Microascus that are associated with plants, mosses, and soil. Fungi and bacteria may be resistant to glyphosate and has the potential to be used in restoration programs. For this and future studies, these enriched organisms can be isolated and cultured and their functions evaluated using omics tools such as shotgun and metabolomics.
Time-lapse of the microbial composition during agave fermentation using Meta-HiC

Mariana G Guerrero-Osornio

Mariana G Guerrero-Osornio, Maritrini Colón-González, Agnès Thierry, Alexander de Luna, Eugenio Mancera, Martial Marbouty, Lucía Morales
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Mexico), Institut Pasteur (France), Instituto de Ecología (Mexico), Laboratorio Internacional de Investigación sobre el Genoma Humano (Mexico), Posgrado en Ciencias Biológicas, UNAM (Mexico), Unidad de Genómica Avanzada (Langebio) (Mexico)

In traditional Mexican distilled spirit production, the fermentation step is conducted in open tanks, promoting microbial interchange with the surrounding environment. This creates a dynamic succession of microbial communities, which, until now, have not been characterized at the genome-wide, especially with species-level resolution. Previous research indicates dynamic shifts in microbial diversity throughout the agave juice fermentation process—starting with a higher diversity and ending with a lower one as fermentation concludes. The objective of this study is to map a genome-wide and species-level resolution of the succession of microbial communities during agave juice fermentation by utilizing the Meta-HiC technique, at seven consecutive time points throughout fermentation. We detail the relative abundance of Metagenome Assembled Genomes (MAGs) across different life domains. Predominantly, MAGs are linked to bacteria and yeast, with a notable prevalence of lactic acid bacterial groups. Intriguingly, we noted an uptick in the relative abundance of MAGs possibly originating from viruses in the fermentation’s latter stages. Consistent with metagenomic studies, some data remain still unclassified, though this portion decreased over time. The Meta-HiC technique has facilitated the assembly of bacterial and yeast genomes at each phase of fermentation, allowing for an analysis of species’ relative coverage and genomic diversity. These findings enhance our understanding of the microbial dynamics involving bacteria, yeast, and viruses in this unique fermentation process, providing essential insights into the ecological interactions driving this microbial succession.
In many species there is a great deal of uncertainty regarding the rate of substitution, implying that the molecular clock is poorly calibrated. Historically, molecular clock calibration has required species divergence times to be known from the fossil record or from geologically dated events, limiting the use of molecular clocks for many species, including bacteria. Sedimentary ancient DNA (sedaDNA) is a rich source of genetic data that captures broad taxonomic diversity over long time transects. These transects of ancient DNA across time can, in theory, be used to calibrate the molecular clock for a large variety of species, including bacteria. However, such analyses are challenged by the fragmented and damaged nature of short read ancient DNA. Here, we introduce ratePlacer, a phylogeny-based method for analyzing sedaDNA that can combine the information from many short reads in a sample while accounting for DNA damages to provide maximum likelihood estimates of sample ages. Furthermore, the method can be used to calibrate the molecular clock when sample ages are known. Using a birth-death process to simulate ancient DNA, we show that ratePlacer is able to accurately estimate the sample age across time. Applying ratePlacer to eighteen well-dated sedaDNA samples spanning eighteen thousand years and comprehensive COI and 16S databases, we calibrate the COI and 16S molecular clock across various taxonomic groups. Our new calibrations of the molecular clock will be useful for dating split times in phylogenies in many taxonomic groups that previously were lacking information about molecular clock rates.
Pervasive selective sweeps within and across human gut microbiomes

Nandita Garud

Presented by self

Los Angeles (USA), University of California

The human gut microbiome is composed of a highly diverse consortia of species that are continually evolving new genetic variants within and across hosts. These genetic variants are known to impart important phenotypes to the host, including the ability to digest foods and evade antibiotics. While much recent work has shown that microbiome adaptations can rapidly arise within hosts on timescales of just a few days and months, little is known about how commonly these adaptations are shared across hosts. Yet, the ability to identify adaptations common to many hosts would not only reveal shared selection pressures across hosts, but also key drivers of functional differentiation of the microbiome that may affect community structure and host traits. Here, we develop a novel selection scan statistic, named the integrated linkage disequilibrium score (iLDS), that can detect the spread of adaptive haplotypes across host microbiomes via migration and horizontal gene transfer. Application of the statistic common commensal gut species from a large cohort of healthy, Western adults reveals pervasive spread of selected alleles across human microbiomes. Among the candidate selective sweeps recovered by iLDS is an enrichment for genes involved in the metabolism of maltodextrin, a synthetic starch that has recently become a widespread component of Western diets. In summary, we demonstrate that selective sweeps both within and across host microbiomes are a common feature of the evolution of the human gut microbiome.
Quantifying negative selection on introgressed fragments in human gut bacteria

Sophie Jean Walton

Sophie Jean Walton, Dmitri A Petrov, Benjamin H Good
Stanford University (USA)

Introgression is ubiquitous across the tree of life, enabling the exchange of genetic information between otherwise isolated groups of individuals. In sexually reproducing organisms, previous work has shown that negative selection against introgressed fragments can contribute to the genetic isolation of species and subspecies. However, the evolutionary forces acting on introgressed fragments in bacteria, which reproduce clonally and exchange genetic material through horizontal gene transfer, remain poorly characterized. Recent work has demonstrated that the rates of genetic exchange can be reduced between groups of ecologically and genetically similar strains. However, it remains unclear to what extent negative selection or other molecular barriers to recombination (e.g. bacterial defense systems) are driving this genetic isolation. To distinguish between these mechanisms, we developed a new approach for quantifying the strength of negative selection on horizontally transferred fragments, which compares the frequency spectra of individual transfer events and synonymous mutations on the core genome. We detect these events by leveraging a hidden Markov model that infers the site specific ancestry of genomes that recombine via horizontal gene transfer. We applied this approach to a collection of Phocaeicola vulgatus genomes from hundreds of human hosts, which were previously shown to be divided into two genetically isolated clades. Our preliminary results suggest that horizontally transferred fragments from the other clade tend to be deleterious in their new genetic background. This provides new evidence that selective forces (e.g. genetic incompatibilities) may contribute to the genetic isolation of this prominent member of the human gut microbiome.
The emergence of a resistance mechanism to a synthetic antimicrobial in the resistome: when genomic mobilization is enough

Stella Cellier-Goetghebeur

Stella Cellier-Goetghebeur, Claudèle Lemay-St-Denis, Maxime St-Aubin, Joelle N. Pelletier
CGCC, Center in Green Chemistry and Catalysis (Canada), Department of Biochemistry and Molecular Medicine, Department of Chemistry, Engineering and Applications (Canada), Function, PROTEO, The Québec Network for Research on Protein, Université de Montréal (Canada)

While resistance to natural antibiotics is ancient, this is not the case for the emergence of mechanisms of resistance to antimicrobials of synthetic origin. Trimethoprim is one such antimicrobial: this synthetic molecule introduced in the 1960s targets the ubiquitous and essential bacterial dihydrofolate reductase enzymes (FolA). However, Type B dihydrofolate reductase (DfrB) enzymes were identified in the 1970s as intrinsically providing resistance to trimethoprim. When FolA is inhibited by trimethoprim, DfrB acts as a substitute to maintain active metabolism. Although FolA and DfrB catalyze the same reaction, they are evolutionary unrelated. The origin of DfrB enzymes is unknown; here, we examine the mechanisms that led to DfrB enzyme evolution and emergence in the clinical resistome. Our recent work reveals that DfrB enzymes are present not only in pathogenic bacteria but also in environmental bacteria not subjected to anthropogenic pollution. Here, through curation and analysis of a dataset of 82 experimentally characterized DfrB enzymes identified in a wide variety of environments such as permafrost and the rhizosphere, we demonstrate that the emergence of DfrB enzymes in bacteria predates the introduction of trimethoprim in clinical settings. Our genomic analyses reveal that DfrB-mediated trimethoprim resistance was recruited at least once to the clinical setting, from agricultural bacteria. This led to the emergence of a distinct population of dfrB genes defined by “clinical traits” rather than “environmental traits”. Our observations allow proposing a model that recapitulates evolutionary trajectories leading to the emergence of DfrB enzymes in the modern resistome.
Horizontal gene transfer is a ubiquitous force in microbial evolution. Previous work has shown that the human gut is a hotspot for gene transfer between species, but the more subtle exchange of variation within species -- also known as recombination -- remains poorly characterized in this ecosystem. Here, we show that the genetic structure of the human gut microbiome provides an opportunity to measure recent recombination events from metagenomic sequences of fecal samples, enabling quantitative comparisons across diverse commensal species that inhabit a common environment. By analyzing recent recombination events in the core genomes of 29 human gut bacteria, we observed widespread heterogeneities in the rates and lengths of transferred fragments, which are difficult to explain by existing models of ecological isolation or homology-dependent recombination rates. We also show that natural selection helps facilitate the spread of genetic variants across strain backgrounds, both within individual hosts and across the broader population. These results shed light on the dynamics of in situ recombination, which can strongly constrain the adaptability of gut microbial communities. Furthermore, we discuss the implications of these observations on the coevolution (or "codiversification") of the gut microbiota with modern human populations.
S7 - Epigenetic inheritance: from molecular mechanisms to evolutionary consequences.
The transition from natal downs for heat conservation to juvenile feathers for simple flight is a remarkable environmental adaptation process in avian evolution. However, the underlying epigenetic mechanism for this primary feather transition is mostly unknown. Here we conducted time-ordered gene co-expression network construction, epigenetic analysis, and functional perturbations in developing feather follicles to elucidate four downy-juvenile feather transition events. We discovered that LEF1 works as a key hub of Wnt signaling to build rachis and converts radial downy to bilateral symmetry. Extracellular matrix reorganization leads to peripheral pulp formation, which mediates epithelial - mesenchymal interactions for branching morphogenesis. ACTA2 compartments dermal papilla stem cells for feather cycling. Novel usage of scale keratins strengthens feather sheath with SOX14 as the epigenetic regulator. We found this primary feather transition largely conserved in chicken (precocious) and zebra finch (altricial) and discussed the possibility that this evolutionary adaptation process started in feathered dinosaurs.
Evolutionary study on DNA methylation in the tree shrew brain and its role in X chromosome inactivation

Dongmin Son

Dongmin Son, Lei Shi, Soojin V. Yi

Cellular and Developmental Biology, Department of Ecology, Department of Molecular, Evolution and Marine Biology, Key Laboratory of Animal Models and Human Disease Mechanisms of Yunnan Province, Key Laboratory of Genetic Evolution & Animal Models, Kunming College of Life Science, Kunming Institute of Zoology (China), Kunming Institute of Zoology?Chinese Academy of Sciences (China), Neuroscience Research Institute, University of California Santa Barbara (USA), University of Chinese Academy of Sciences (China)

Tree shrew is an emerging model organism to study human evolution due to several advantages, notably their evolutionary proximity to primates. Here we analyzed the first genome-wide, single-base-resolution methylome and transcriptome from tree shrew brains. We identified a clear correlation between gene expression and methylation level in both the promoter and gene body regions. We examined the X chromosome DNA methylation and observed a significant global reduction (hypomethylation) of DNA methylation across the entire X chromosome in females. This pattern resembled what was observed in koalas (a marsupial) and differed from that in humans. Our comparative analysis indicated that the degree of differential DNA methylation between the female and male promoters was dependent on CpG contents. Conversely, differences in CpG contents between species could explain divergent patterns of DNA methylation. Moreover, we show that female X hypomethylation does not directly contribute to the silencing of the inactivated X chromosome, nor does it significantly drive sex-specific gene expression, suggesting the existence of a unique role of DNA methylation in XCI. We also newly annotated the Xist locus in tree shrew, and demonstrate that it is strongly regulated by DNA methylation, influencing its extremely sex-specific expression. This study provides a foundation for future investigations into the DNA methylation system in tree shrews, facilitating the exploration of evolutionary aspects of primate DNA methylation systems and the role of DNA methylation in the X chromosome.
MBD2/3 lost its methyl-CpG binding ability in multiple families of Holometabola

Elisa Israel

Elisa Israel, Zoe Marie Länger, Joachim Kurtz, Sonja Prohaska
Leipzig University (Germany), University of Münster (Germany)

DNA methylation is sparse in insects, this reduction is even more pronounced in Holometabola. In some Holometabola, DNA methylation is lost entirely. Methyl-CpG-binding domain (MBD) proteins bind methylated DNA and are therefore important readers of these epigenetic marks. We hypothesize that evolutionary reduction of genome-wide methylation results in changes to MBD proteins. Only one MBD family member, MBD2/3, is known in insects. Two isoforms have been identified in Bombyx mori. The long isoform MBD2/3-L contains a complete MBD domain spanning the first two exons, whereas the short isoform MBD2/3-S lacks the second exon and therefore half of its MBD domain. It has been reported that only the MBD2/3-L isoform is able to bind methyl-CpGs. In this study, we analyzed transcriptomic and genomic sequence data from holometabolan orders to identify MBD2/3 genes and isoforms. Furthermore, we analyzed transcripts of representative species from each insect order. Overall, MBD2/3 is highly conserved in sequence and gene structure, with splice sites being identical within each order. Both isoforms are present in most insect orders. However, in Coleoptera and Hymenoptera, the longer isoform MBD2/3-L, capable of binding methylated DNA, has been lost entirely. This loss is also observed in several other holometabolan families and species. The results suggest that MBD2/3 has lost its binding ability in Coleoptera and Hymenoptera. Furthermore, we propose that MBD2/3-L has been lost several times independently but exclusively in Holometabola. The reduced levels of DNA methylation may be linked to these losses.
Genome doubling results in N6-Methyladenosine and physiological response divergence enhanced tetraploid Hordeum bulbosum tolerance to salt stress

Genlou Sun

Genlou Sun, Margaret Scarrow
Saint Mary's University (Canada)

Polyploidy has played a critical role in increased tolerance to environmental stress compared to diploid progenitors. N6-Methyladenosine (m6A) is suggested to play an important role in stress response. However, whether genome doubling affects m6A to increase autopolyploids stress tolerance is still unclear. This study aims to compare physiological and m6A changes between autotetraploid and diploid wild barley (Hordeum bulbosum) in response to salt (NaCl) stress. Our results showed that: (1) Relative water content was significantly higher and water loss was significantly lower in autotetraploid H. bulbosum than diploid at 0 days (control), 14 days (salt stress) and 28 days (recovery); (2) proline was significantly higher in autotetraploid barley at day 0 but through treatment and recovery proline was higher in diploids. Both autotetraploid and diploid barley showed a significant increase at treatment and recovery when compared to controls; (3) H2O2 was significantly higher in diploids compared to tetraploids; (4) Chlorophyll A was higher in autotetraploids, however diploids showed a more significant drop under stress than autotetraploids, while chlorophyll B appeared higher in autotetraploid controls except for controls observed at recovery; (5) m6A in total RNA showed no significant difference between ploidies in controls, but during treatment and recovery autotetraploids were significantly higher than diploids. Both had a significant increase in m6A for treatment and recovery when compared to controls. These results suggest that increased m6A might be one of molecular mechanisms that increases salt tolerance in autotetraploid H. bulbosum compared to diploids, which enhances the adaptation of autopolyploids.
The role of CTCF in the transcriptional activation of the hematopoietic-specific gene HBA2 is evolutionary conserved

Gustavo Tapia

Gustavo Tapia-Urzúa, Hober Nelson Núñez-Martínez, Martin Franke, Sylvia Garza-Manero, Ángel Josué Cerecedo-Castillo, Carlos Alberto Peralta-Álvarez, Georgina Guerrero, Félix Recillas-Targa

Andalusian Centre for Developmental Biology (CABD) (Spain), Institute of Cellular Physiology, UNAM (Mexico)

CTCF is a DNA binding protein that regulates chromatin organization and gene expression, by interactions with transcription factors and co-factors, and by promoting long-range interactions of specific regulatory elements. While the architectural role of CTCF has been further characterized, its relationship with DNA methylation has been less explored. Here, we show that CTCF binding to tissue-specific regulatory elements is dynamic during erythroid differentiation and that a subset of bindings sites delimits differentially methylated regions. CTCF enrichment, therefore, protects against propagation of DNA methylation, promotes chromatin accessibility, and binding of the erythroid-specific transcription factor GATA-1, in particular loci. In agreement, deletion of a CTCF binding site within the promoter of the chicken globin gene HBAD, and its orthologous in humans HBA2, results in CpG methylation propagation to the promoters, which is coupled to transcription factor binding reduction and lower transcriptional levels of HBAD and HBA2, respectively. These changes were not explained by disruption of interactions between cis-regulatory elements of the chicken ?-globin domain, since no major changes in chromatin contacts were observed. Therefore, there are relevant CTCF mechanisms beyond of its architectural role that are conserved in specific loci among vertebrates.
The architectural protein CCCTC-binding factor (CTCF) has emerged as an essential regulator of gene expression in cell differentiation. However, the relationship and its regulatory effect on IncRNA genes in hematopoiesis remains elusive. We demonstrated that CTCF targets and regulates the IncRNA DUBR, which was highly expressed in human hematopoietic stem and progenitor cells (HSPC) but depleted in erythroblasts. By CRISPR-Cas9-mediated genome editing and genome-wide analysis of transcriptome as well as chromatin landscape, we assessed that DUBR depletion enhances the expression of myeloid-erythroid cell differentiation genes by coordinating genome-wide enhancer activation and reducing the activity of the transcriptional repressor HES1 at gene promoters. Our results indicate that attenuation of IncRNA DUBR coordinates the establishment of a myeloid-erythroid gene expression program.
Exposure to early-life stress induces developmental shifts in behaviour and DNA methylation in a small prey fish

Janay A. Fox

Janay A. Fox, Simon M. Reader, Mélanie F. Guigueno, Rowan D. H. Barrett
McGill University (Canada)

Early-life experiences can predict the environments experienced later in life, giving individuals an opportunity to develop behaviour that matches the environment (‘adaptive programming’). Epigenetic mechanisms such as DNA methylation (DNAm) offer a link between the environment and the genome that could underlie mechanism such as developmental plasticity. The goals of this study were to investigate the impact of early-life exposure to predation stress on behaviour and DNAm in the brain of Trinidadian guppies (Poecilia reticulata), a small freshwater fish. We exposed guppies throughout development to either alarm cue (conspecific skin extract), inducing predation stress, or a control cue. We then gave them two behavioural assays, an open-field and a shoaling test, before performing whole-genome bisulfite sequencing on whole brains. Guppies exposed to alarm cue during development exhibited increased shoaling in adulthood compared to those exposed to the control treatment, but there were no impacts detected on boldness or exploratory behaviour. The shift in shoaling behaviour displayed by guppies exposed to alarm cue matches that seen in naturally occurring populations, suggesting that developmental plasticity could be used to adaptively program behaviour to match predation environment. Additionally, we identified stable shifts in DNAm in the brain in response to developmental alarm cue exposure. These shifts in DNAm were in genes involved in neural signaling and responses to stimulus. In conclusion, this study has shown how early-life predation stress can induce developmental plasticity and that shifts in neural DNAm could be an underlying mechanism of developmental behavioural plasticity.
Wolbachia induces cellular differentiation in Drosophila cell culture cells following stable infection.

Jodie Jacobs

Jodie Jacobs, Samuel Sacco, William Seligmann, Cade Michanandi, Merly Escalona, Richard Green, Shelbi L Russell
Santa Cruz (USA), University of California

Wolbachia is an endosymbiotic bacterium prevalent in arthropods and nematodes, renowned for its manipulation of host fertility. The phenotypic effects of Wolbachia implicate core developmental mechanisms prompting an investigation into its potential role in mediating cellular differentiation in Drosophila melanogaster. Through sequencing the transcriptome of Drosophila melanogaster grown in vitro we have identified many genes with differential expression between uninfected cultures and those with stable Wolbachia infections. Cluster analysis of these data suggest alterations in cell state associated with Wolbachia infection. Furthermore, we observed significant changes in the genome-chromatin contact map in infected cells compared to their uninfected counterparts. These alterations in chromatin structure and gene expression propose a potential mechanism through which Wolbachia may influence host cellular differentiation. We employed Gene Ontology (GO) analysis to examine both genes with differential expression and genetic elements within regions of the genome with enriched contacts. The results revealed significant genome-chromatin contact enrichment in immune response pathways, including genes like bonamin, serpin, and tetraspanin in stable Wolbachia infections. GO analysis of the host transcriptome revealed that Wolbachia infections induce significant enrichments in the ecdysteroid kinase-like gene family and folate transporter-like family while suppressing gene expression in the metalloexopeptidases gene group. These findings provide new insights into the molecular mechanisms underlying Wolbachia niche formation and maintenance in Drosophila melanogaster.
Tissues represent a fundamental evolutionary interface at the junction of phenotype and genotype. Tissues differentiate by epigenetic mechanisms, which are of core interest to developmental and evolutionary biology. Epigenetic studies are however restricted by access to pure tissues, and underlying assumptions about the locations and distributions of epigenetics markers. Advanced surgical techniques cannot typically be applied to the foci of evolutionary studies in the field, nor can prior assumptions about the predisposition for and consistency of methylations in CpG context be safely made in most non-model organism. Epigenetic modifications are nonetheless linked to consistent patterns of somatic inheritance. As such, generalized evolutionary methodologies may allow for the efficient deconvolution of mixed tissues across eukaryotes. Here we apply somatic site-frequency information to deconvolute 5mC methylation patterns though analysis of mutational accumulation. We use simulations and archival reads to show that somatic mutations are common and detectable in next-generational sequencing data. We then use somatic-site-frequency changes to accurately derive the relative mixtures of bisulfite reads in in silico mixtures of known proportions. Using this information, we are able to perform unbiased and accurate deconvolution of mixed tissue methylations in CpG and non-CpG context. We are ultimately able to recover 20-30% of differentially-methylated sites, and approximately 50% of differentially-methylated CpG islands and gene bodies in any cytosine context at contamination levels up to 90%. Our findings highlights the utility of evolutionary theory across scales, and expand the accessibility of epigenetics to studies.
Defining the Role of Cis-Acting DNA Sequence in Evolutionary Divergence of the Bivalent Chromatin State in Mammals

Kimberly Griffin

Kimberly Griffin, Kira Marshall, Kenny Thai, Bluma Lesch
Yale (USA)

Changes in gene regulation play a major role in evolution. Bivalency is an important developmental gene regulatory state defined by the presence of the activating histone mark H3K4me3 and repressive mark H3K27me3 on the same nucleosomes. Bivalency is found in cells with stem-like characteristics and is located in the promoters of transcriptionally silent genes important for somatic differentiation during embryonic development. Bivalency is present in spermatogenic cells throughout mammalian evolution and is proposed to poise developmental genes to be efficiently activated or repressed after fertilization as cells differentiate in the embryo. It is currently not understood how the bivalent state is regulated. We are using a comparative evolutionary approach to determine what sequence changes are important in causing a gene to gain bivalency in a species lineage. The gene Traf6 is bivalent in mouse embryonic stem cells (mESCs) and spermatogenic cells, while it is only marked by H3K4me3 in other mammalian species. To determine what genomic sequence designates Traf6 as bivalent in mouse but not in other species, we generated a humanized reporter by replacing the promoter of mouse TRAF6 with the corresponding region from the human genome in mice and mESCs. We inserted GFP after the humanized sequence, allowing us to track transcriptional consequences of changes in chromatin state. We found that murine germ cells carrying the humanized Traf6 promoter recapitulate the human regulatory state, providing us with a system for dissecting how changes in promoter sequence affect bivalency at this model locus and potentially impact phenotype.
Novel insights into X-Chromosome Inactivation through Peromyscus Leucopus

Maria De Lourdes Andrade Ludeña

In placental mammals, X-chromosome inactivation (XCI) is necessary for ensuring gene expression balance between XY males and XX females by silencing one X chromosome in females. XCI is initiated by the long non-coding RNA (lncRNA), X inactive specific transcript (Xist). However, the exact mechanism of Xist remains debated. The rodent Peromyscus leucopus has recently emerged as a model organism given its high levels of genetic diversity within the species. We identified an Xist homolog in P. leucopus that shares both human and mouse Xist features. Using RNA FISH, we visualized the localization of Xist lncRNA in the nucleus of P. leucopus fibroblast cells. Contrary to current knowledge that Xist is exclusively expressed in female cells, we detected low levels of Xist expression in male P. leucopus fibroblasts. This suggests a different mechanism of X-inactivation where the single X chromosome in males may undergo gene silencing. We performed multi-species sequence alignments to define the six RNA motifs (Repeats A-F) within Xist homologs and identified one, Repeat C, that heavily varies in copy number and nucleotide sequence across species. Repeat C is heavily degraded in P. leucopus and humans, but expanded in mice, indicating divergent Xist RNA functions during evolution. To test our hypothesis that Xist sequence changes underlie male Xist expression, we use CRISPR-mediated gene editing in male and female P. leucopus cells. Comparative analysis among rodent species will further our understanding of the evolutionary and functional differences in Xist regulation, and how lncRNAs evolve.
Exploring the evolutionary dynamics of epigenetic switches and beyond: Insights from gene regulatory circuit models

Mariana Gómez-Schiavon

Mariana Gómez-Schiavon
Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), UNAM (Mexico)

Epigenetic switches have long fascinated evolutionary biologists as key players in adaptation to fluctuating environments. However, our research now extends beyond their traditional role. We investigate the broader landscape of dynamic properties in gene regulatory circuits, utilizing mechanistic models to explore evolutionary processes shaping these circuits. Through in silico evolution, we uncover how dynamic properties emerge, propagate, and persist under natural selection, shedding light on the structural dependencies within circuits and their cellular contexts. While maintaining a focus on epigenetic switches, we delve into other emergent properties such as plasticity and oscillations, examining their selective advantages and competition both among themselves and with classical genetic adaptation. Our findings reveal fundamental trade-offs, such as the balance between adaptation time and phenotypic robustness to biochemical noise, driving the selection of epigenetic switches in fast fluctuating environments. This research not only enriches our understanding of epigenetic mechanisms but also provides insights into broader biological phenomena. Our investigations hold implications for diverse fields, from bacterial antibiotic resistance to plant survival strategies in extreme environments. Through this work, we aim to contribute new perspectives on the complexity of life and the underlying principles that govern it.
Contribution of epimutation to evolution in C. elegans and beyond

Peter Sarkies

Presented by self
University of Oxford (UK)

Transgenerational epigenetic inheritance is becoming increasingly well documented as a process whereby information beyond the DNA sequence can be passed on between generations. The mechanism of transgenerational epigenetic inheritance has been best characterised in the nematode C. elegans, where small non-coding RNAs and changes in the chromatin environment combine to transmit epigenetic information, sometimes for hundreds of generations in the absence of DNA sequence alterations. These long-term epigenetic differences between individuals, known as epimutations, thus have the potential to contribute to evolutionary processes such as natural selection or drift; however, this possibility remains largely untested. I will describe how my laboratory has designed and performed laboratory evolution experiments to test the occurrence, stability and genome-wide distribution of epimutations in C. elegans. I will explain how our results indicate that epimutations have limited stability, lasting between 4-6 generations on average, and how specific subsets of genes are prone to acquire epimutations. I will then discuss how these epimutations, despite limited duration, may contribute to evolutionary processes, and explain the possible implications of these findings for understanding the development of antihelminthic agents in parasitic nematodes.
Transgenerational epigenetic inheritance (TEI) has garnered increasing interest, yet this additional system of inheritance remains highly controversial. In plants, several examples of true epiallelic variants (i.e. that are inherited independently of any DNA sequence changes), have been described and all involve DNA methylation changes over transposable element (TE) sequences. However, the precise determinants of TEI are still unknown in plants, and so are its extent and functional impact in nature. To address this issue, we have undertaken a stringent and comprehensive genome-wide analysis of DNA methylation variation at TE sequences in an experimental population where severe hypomethylation was experimentally-induced as well as in hundreds of natural strains of the plant Arabidopsis thaliana. Collectively, our results indicate that TE-associated epiallelic variation is relatively abundant in nature where it is shaped by an antagonism between RNA-directed DNA methylation limiting its inheritance on the one hand and transcriptional activity favoring it on the other, and as a result is found preferentially over divergent and low copy-number TE sequences located near stress-responsive genes. In addition, we will discuss the different mechanisms through which the environment also exerts a major influence and their consequences for understanding the adaptive role that TEI may play in plants.
Exploring the potential for epimutations to drive cancer cell adaptation

Sito Torres-García

Presented by self
Cambridge Institute (UK)

The emergence of therapy resistance in metastatic cancer – wherein tumour cells from a primary site progressively colonize distant organs – impedes curative therapy and results in patient relapse. It is known that resistance can be caused by genetic mutations, but in some cases, there is no clear genetic basis, suggesting that epigenetic mechanisms might drive resistance. Epigenetic mechanisms may drive therapy resistance in cancer cells through formation of epimutations changes in gene expression that are independent of changes in DNA sequence. Such epimutations may provide initial tolerance, allowing a fraction of tumour cells to survive and later acquire secondary genetic mutations that drive disease progression to relapse. Epimutations have been proposed to facilitate adaptive phenotypic responses to external insults, but experimental evidence is scarce. Our previous research demonstrated that epimutations can drive environmental adaptation in fission yeast, but whether similar mechanisms play a role in complex mammalian systems such as cancer remains elusive. We are investigating the potential for epimutations to prime adaptive evolution in metastatic breast cancer by establishing a novel profiling method capable of detecting epimutations in individual cancer cells exposed to therapy. Our project aims to unravel conserved epigenetic mechanisms that drive therapy resistance in metastatic cancer, but also general adaptation to stress in mammalian cells. Understanding these mechanisms could shed light on broad principles of eukaryotic adaptation and reveal potential therapeutic opportunities to improve patient outcome.
Constructing a cattle and sheep pan-epigenome to elucidate epigenetic impacts upon gene expression and phenotypic variation

Stephanie McKay

Stephanie McKay, Shangqian Xie, Morgan Stegemiller, Gabrielle Becker, Kimberly Davenport, Katie Shira, Denise Konetchy, Patricia Villamediana, Brenda Murdoch

South Dakota State University (USA), University of Idaho (USA), University of Missouri (USA), Washington State University (USA)

Phenotypic variation is driven by both genetic and epigenetic variation, yet understanding the extent of epigenetic modifications and their impact on phenotype remains largely unexplained. Moreover, the onset of pangenomics remains disconnected from pan-epigenomics and hence inhibits a fully comprehensive view of diversity. In order to facilitate dissection of epigenetic contributions towards phenotypic variation, we have begun generation of a cattle and sheep pan-epigenome. Ultimately, an exhaustive characterization of epigenetic conservation and diversity will be examined with 5-methylcytosine, 5-hydroxymethylcytosine, histone marks (H3K4me3, H3K27ac, H3K4me1 and H3K27me3), open chromatin, chromatin accessibility and N6-methyladenosine in three tissues from each of three different breeds of cattle and sheep. In this study we have performed Whole Genome Bisulfite Sequencing (WGBS) and Methylated RNA Immunoprecipitation Sequencing (MeRIP-Seq) to investigate the effects of DNA 5-methylcytosine and RNA N6-methyladenosine on gene expression. After characterization and comparative analysis, results indicate that RNA methylation has a greater influence on gene expression compared to DNA methylation. Subsequently, gene interaction networks were constructed to demonstrate the interactions of DNA methylation, RNA methylation and gene expression. Specifically, we identified two gene interaction networks whose gene expression was substantially impacted by RNA methylation. Interestingly, both of these gene interaction networks are known to influence transcription. These findings enhance our understanding of genome-wide dynamics and the complex interactions among DNA and RNA methylation with gene expression, offering new perspectives and opportunities to further epigenetics research.
Many organisms show sexually dimorphic phenotypes that are regulated by differential expression and gene splicing. The mealybug Planococcus citri has extreme sexual dimorphism whereby males and females differ dramatically in morphology, lifespan and the way they inherit genetic material. Mealybugs do not possess sex chromosomes and as such males and females are genetically identical, suggesting this extreme sexual dimorphism is epigenetically regulated. Males exhibit a form of genomic imprinting known as paternal genome elimination (PGE) whereby the haploid paternal genome is silenced in somatic cells and eliminated from the germline. The molecular mechanisms behind this are still under speculation, however the involvement of DNA methylation seems likely. Mealybugs show distinct sex-specific DNA methylation profiles and considering the known involvement of DNA methylation in imprinting in mammals and plants, it is possible that it also plays a role in this species. In addition to this, mealybugs have independently evolved promotor methylation associated with gene silencing, more similar to the methylation profile seen in mammals than in other insects. To further understand the role of DNA methylation in PGE, we have combined RNA-seq and DNA methylation data with parental SNP information to ascertain whether differential methylation occurs on the maternal or paternal strand of DNA, and how this is linked to sex-biased gene expression as well as the differential behaviour of paternal chromosomes in males and females. Together these results will give insights into epigenetic regulation of sexual dimorphism and genomic imprinting in an insect with unique biology.
S8 - Clustering of human cohorts beyond race and ancestry: Towards relational thinking in genomics.
Population structure reflects shared population histories. Identifying structured subgroups in biobanks is key to exploratory analyses, visualization, and downstream study of biomedical, demographic, or geographic relationships. One approach—identity-by-descent (IBD) clustering—detects chromosomal segments that are shared identical-by-descent between at least two individuals, and finds communities in networks of shared IBD. Another is density clustering on variant data that has been dimensionally-reduced. While numerous dimensionality reduction approaches are applied to biobanks, we focus on uniform manifold approximation and projection (UMAP), which preserves high-dimensional topology of data. IBD clustering and UMAP do not rely on social/overly coarse labels such as continent, race, or ethnicity to generate clusters. In past applications, we have found that UMAP-identified subgroups correlate strongly with demographic history, environmental, and phenotypic data and identify fine-scale clusters similar to IBD. However, no explicit connection has been drawn between the approaches. Such connections can inform researchers of relationships between coalescence and clustering, aid in interpretation of non-linear dimensionality reduction, and provide alternative computationally cheap tools to study clusters of structure in large biobanks. We study the relationships between these approaches and measures such as time to most recent common ancestor (TMRCA). Using 1000 Genomes Project, CARTaGENE, UK biobank, and Allofus data, we also study relationships between population structure clusters and distributions of environmental variables and geographic data. We present different approaches to visualization, offer best practices on using and visualizing clusters via different methods, and highlight the impact of data processing and algorithm parametrization.
Quantitative Interpretations of Principal Component Analysis
Benjamin M Peter

Benjamin M Peter, Divyaranat Popli
Max Planck Institute for Evolutionary Anthropology (Germany), University of Rochester (USA)

Principal Component Analysis (PCA) is a widely used method to assess population structure. However, interpreting PCA-plots remains difficult, and is often done qualitatively. In contrast to previous approaches that interpret the principal components themselves, we develop a framework that uses PCA to simplify the geometry of the genotype space, allowing us to link PCA to underlying population genetic models. We show that population trees, admixture graphs and general migration models all leave clear geometric patterns in a PCA. These patterns can be used for statistical inference and also to directly assess and visualize deviations from the model. Our results have numerous practical implications. First, we find that different versions of PCA should be used for different purposes; probabilistic PCA enables us to isolate population structure from sampling noise, and is best used to visualize or correct for population structure. In contrast, ordinary PCA does not make this distinction and is best suited for investigating statistical artifacts and biases. Second, our results give suggestions on how the results of a PCA should be plotted, how many PCs should be used and how the axes should be scaled. Finally, building on our framework, we are able to generate fast, accurate and flexible tests for admixture, as we can use the full data set rather than a small subset for inference. In summary, while most population genetic models assume a discrete population structure, the geometric perspective on PCA illustrates how models can be extended to populations that vary in a more continuous manner.
Evaluating accuracies of forensic analyses across genetic ancestries

Elena I Zavala, Rori Rohlfis, Priya Moorjani
Berkeley (USA), University of California, University of Oregon (USA)

DNA analyses have transformed forensics applications from identification of missing individuals to solving cold cases. In these analyses, profile matching between DNA recovered from a biological sample or piece of evidence and an individual is commonplace. The significance of this match, which is of particular importance for partial or mixture profiles, is typically calculated based on reference allele frequencies of a handful of markers. Recently, investigative leads have been provided using varying qualities of sequencing data—i.e. coverage, number of sites, etc.—to match individuals or genetic relatives. Given the potential impact of these results on people living today, this raises questions as to how new methods should be tested to understand accuracies across diverse populations. In this study, we discuss the utility of using different statistics to describe individual DNA profiles in place of discrete population identifiers (i.e. heterozygosity and f-statistics). We then evaluate the impact of degraded DNA on genotyping accuracy for forensic genetic analysis across diverse populations using five different SNP panels. The outputs of this study demonstrate the potential for performing human identification with degraded DNA at low coverages (<5X), but highlights critical considerations. In particular, it is important to consider genotype refinement, choice of reference panels, and interpretations of resulting DNA profiles. We discuss the critical need to move beyond conventional population definitions in analysis of diverse individuals for forensic applications.
What is in a discrete category? Clustering genomics cohorts beyond race, ethnicity—and ancestry
Hussein Mohsen

Presented by self
Memorial Sloan Kettering Cancer Center (USA)

Since their inception, classification systems of human biodiversity in modern science reflected socio-politically enforced categorical lines and recurrently adjusted in the service of systems of power and disenfranchisement. Further, normalized discrete classification systems have been muddled with arbitrariness and often relied on fragmented data. In this talk, I will first provide a historical critique of the use of race(-ialized) and ethnicity categories as proxies of variation in genomics. With respect to genetic ancestry categories, which vary in type yet are also heavily politically demarcated, I will then describe quantitative lenses that transcend their use in studying genomics cohorts with a focus on detecting somatic-germline interactions in cancer.
Non-parametric representations of structure in genetic ancestry through time

John Novembre

Universitat Pompeu Fabra (Spain), University of Chicago (USA)

Understanding population history often requires the assumption of discrete, bounded populations even when the underlying history is more complex. Currently, there is a need for tools that can reveal the structure of genetic ancestry in a time-varying way, without assumptions about whether the structure has discrete units or continuous gradients. Here, we present progress on this problem by describing structure in genetic ancestry through time, leveraging reconstructed Ancestral Recombination Graphs (ARGs). Specifically, we assess how individuals sampled at present descend from sets of genetic ancestors in the past. This approach naturally lends itself to revealing discrete population groupings when they are present and to revealing continuous gradients when the history is more complex. Attractively, the method produces flexible representations of genetic histories, in effect, adding a time dimension to popular low dimensional summary approaches such as PCA. The method is also helpful for revealing the timescale of genetic structure in a way that is more instructive about how genetic structure impacts allelic sharing than many population history models. The method is individual-based and does not rely on population descriptors, though labels might be used to facilitate interpretation in light of historical processes if that is a focus. Overall, the method is a promising alternative to single snap-shot representations of population structure (e.g. PCA) and to demographic history tools that assume discrete populations. It also has potential as a teaching tool to help reveal structure in the "cloud of gene histories" (W. Maddison) underlying genetic variation.
Human genetic variation is continuous and yet categorical labels are still commonly used in population genetics analyses to describe groups of individuals and to facilitate matching to reference samples. Such approaches necessarily exclude individuals to maintain specificity, result in possibly suboptimal groupings, and lead to decreased power for analysis. Two particular applications that currently adopt discrete groupings are calculating allele frequencies in a reference sample, e.g. for genetic diagnosis, or linkage disequilibrium (LD) matrices, e.g. for fine-mapping. Here, we propose a method to move beyond discrete genetic ancestry groupings towards a continuous model based on genetic similarity to improve accuracy in both these applications. Our method computes local allele frequencies and LD patterns using a distance-weighted method in a low dimensional representation for samples in a genetic reference panel. Using cross-validation, we compared our method to more traditional categorical approaches by evaluating the error and bias from estimated allele frequencies to the known frequencies of groups of individuals. We show an improvement in predicting allele frequencies and LD patterns in hold-out sets of individuals, and we show how the accuracy of our method outperforms discrete methods at a range of genetic divergences. We deploy this method in the largest public aggregated database of human exomes and genomes (gnomAD) which is widely used for genetic diagnosis. Our method provides a critical first step in abandoning the use of discrete labels in population genetics and simultaneously improves downstream applications, including fine-mapping and rare disease investigations.
Multiple periods of admixture and isolation during and after the Transatlantic Slave Trade on the island of São Tomé.

Marta Ciccarella

Marta Ciccarella, Romain Laurent, Zachary Szpiech, Etienne Patin, Françoise Dessarps-Freichey, José Utgé, Laure Lémée, Armando Semo, Jorge Rocha, Paul Verdu

BIOPOLIS Program, C2RT, CIBIO-InBIO, CNRS, CNRS UMR2000 (France), Departamento de Biologia, Department of Biology, Faculdade de Ciências, Human Evolutionary Genetics Unit, Institut Pasteur, Institut Pasteur (France), Institute for Computational and Data Sciences, MNHN, Penn State University (USA), Plate-forme Technologique Biomics, UMR7206 Eco-anthropologie, Universidade do Porto (Portugal), Université Paris Cité, Université Paris Cité (France)

Human admixture history is rarely a simple process in which distinct populations, previously isolated for a long time, come into contact once to form an admixed population. In this study, we aim to reconstruct the complex admixture histories of the population of São Tomé, an island in the Gulf of Guinea that was first settled by the Portuguese in the 15th century, was the site of the first slave-based plantation economy, and experienced successive waves of slavery and post-slavery migration from Africa. We examined 2.5 million SNPs newly genotyped in 96 São Toméans and found that geography alone cannot explain the observed patterns of genetic differentiation within the island. We defined five genetic groups in São Tomé based on the hypothesis that individuals sharing the most haplotypes are more likely to share similar admixture histories. Using IBD and different local ancestry inference methods, we inferred shared ancestries between São Toméans and 70 African and European populations, and we used ancestry tract distributions to reconstruct the admixture histories of each São Toméan genetic group separately. We identify admixture events between admixed groups that were previously isolated on the island. This study demonstrates how complex admixture and isolation histories during and after the Transatlantic Slave-Trade shaped extant individual genetic patterns at a local scale in Africa; it explores cases in which recently admixed populations are themselves at the sources of other admixture events; and it further shows the need for extreme caution when assigning anthropological labels to genetic clusters.
Genetically informed representations of ancestry and ethnicity

Simon Gravel

Presented by self

McGill University (Canada)

Individuals inherit different parts of their genomes from different ancestors. Understanding this variation in genetic ancestry is key to understanding the distribution of genetic variants across individuals. In humans, genetic ancestry is often correlated with ethnicity and geography. Ancestry and ethnicity are complex concepts that cannot be easily summarized with simple labels. Yet, we need ways to succinctly explain broad patterns in population datasets. In this talk, I will discuss strategies to investigate the complex relationship between ancestry, geography, and ethnicity. I will focus on low-dimensional representations of genetic data including PCA, Admixture, UMAP, as well as topological clustering approaches. I will discuss benefits and limitations of these approaches, present a few new methodological developments that allow us to investigate finer patterns, and discuss how these methods can be used to identify and represent relevant patterns in the data, whether these are related to geography, ancestry, ethnicity, or sampling protocols.
S9 - Human population demography and adaptation signals in the Americas.
Evidence for genetic continuity, contacts, and changing social organization in the Central Andes from ancient DNA
Alison R Barton

Alison R Barton, Weston C McCool, Valda G Black, Lars Fehren-Schmitz, Jennifer L Kennedy, David E Reich
Binghamton University (USA), Harvard Medical School (USA), Harvard University (USA), Howard Hughes Medical Institute (USA), University of California Santa Cruz (USA), University of Utah (USA), Washington State University (USA)

Previous work with populations from the Central Andes has suggested a high level of genetic continuity in the region but has also highlighted evidence of movement of peoples and changing kinship structures. In this work, we examine 315 new samples from ancient individuals from modern Peru and Bolivia dating from 8600-200 BP. We find similar evidence of genetic continuity in the Andes based primarily on North-South geography. However, identity-by-descent (IBD) analysis supports the hypothesis of genetic contacts between these groups locally and in a few cases with peoples in what is today Colombia and Ecuador, with at least 3 individuals sharing greater than 50 cumulative cM of IBD segments longer than 8 cM with individuals from these areas. Moreover, we find additional support for the claim that consanguineous pairings increased during the Late Intermediate Period (LIP, 1000-1476 CE) likely due to changing social structures centered around the ayllu system that were reported historically. We observe in increase (Fisher’s exact test, p = 0.00061) from 17% (4/23 samples) to 65% (90/161) during the LIP that subsequently drops (Fisher’s exact-test, p = 1.3E-07) after this period to 9% (3/35). We also present the first large pedigrees from this time period from 3 valleys in the Nazca highlands representing samples from 10 different chullpas. Five of these tombs with the densest sampling suggest burial practices were strongly tied to biological kinship and illustrate family units in which we appear to see this social structure in place.
Ancient Mitogenomes from Belize Provide Insights into the Mobility and Migration of Ancient Communities in Mesoamerica and the Caribbean.

Celia D Cleary

Celia D Cleary, Angelina J Locker, Lauren C Springs, Diane Z Chase, Arlen F Chase, Adrian SZ Chase, Melissa M Badillo, Adela P Vallejos, Genara Cano, Roy Rodriguez, Rick WA Smith, Austin W Reynolds

Baylor University (USA), Belize (Belize), Department of Anthropology, Department of Comparative Cultural Studies, Department of Integrative Biology, Department of Sociology and Anthropology, Department of Women and Gender Studies (USA), Faculty of Native Studies, George Mason University (USA), Indigenous Science, Institute of Archaeology, Mansueto Institute for Urban Innovation and Department of Anthropology, Technology, To'one Masehualoon NGO (Belize), University of Alberta (Canada), University of Chicago (USA), University of Houston (USA), University of Texas at Austin (USA), Vanderbilt University (USA), and Society Lab

Genomic research regarding paleodemography in Mesoamerica to date has often been conducted as part of large-scale studies of North and South America. Poor preservation of ancient DNA in the neotropics, among other difficulties, has limited ancient genomic studies of mobility in this region, with paleogenomic-scale studies of population history of Mesoamerica becoming a sustained focus in the past 5 years. Here, we present mitogenomic data of seven ancient Maya Ancestors recovered from interments in the ancient Maya city of Caracol, located in the Maya Mountains of present-day Belize, and 54 contemporary individuals from Corozal, in Northern Belize. These mitogenomic data were combined with 63 previously published ancient mitogenomes recovered from sites spanning Mesoamerica and the Caribbean to reconstruct paleodemographic and mobility patterns of maternal lineages through time. This work was conducted with authorization from the Belize Institute of Archaeology in and collaboration with Maya partners in Corozal. Haplogroup frequencies and median joining haplogroup networks were generated to understand the sharing of haplotypes across time and space in this region. Bayesian fixed molecular clock models were used to estimate matrilineal phylogenetic relationships between community members included in this study. Bayesian Skyline Plots were generated to infer dynamics in matrilineal effective population size. Models resulting from these 124 mitogenomes are used here to reconstruct paleodemographic events temporally and geospatially to gain a more robust understanding of the demographic history of the region.
The genetic history of early Americans

Cosimo Posth

Presented by self

University of Tübingen (Germany)

Until recently, the genetic history of ancient Native Americans was inferred through the generation and analysis of genetic data from present-day indigenous populations across the Americas. In the last ten years, ancient genomic data from this continent is becoming increasingly available with around a thousand ancient individuals having been sequenced to date. Analysis of these data has revealed an early population split into the ancestors of Native Americans leading to paleo-Alaskan groups, as well as a subsequent split into two lineages that contributed to non-arctic populations. One lineage survived in North America whereas the other represents the primary ancestry found in ancient and present-day South Americans. However, at least three genetic connections between North and South Americans were identified, suggesting a model of multiple gene-flow events between the two sub-continents. The settlement history of South America resulted from a rapid star-like radiation, followed by long-term continuity in multiple regions. This scenario is supported through the analysis of Early Holocene hunter-gatherers across different regions of South America who lack significant allele sharing among each other. The generation of additional genome-wide data from Early to Middle Holocene individuals is currently helping clarify the modes and tempos of genetic exchange between North and South American populations. In addition, localized genomic time transects are providing a more nuanced description of the demographic history of specific American regions until present-day populations. The study of their genomic transformations in comparison with archeological and linguistic evidence is helping to better delineate the complex population history across the Americas.
All humans carry a small fraction of archaic ancestry across the genome, the genetic legacy of Neanderthals and Denisovans. While the effects of Neanderthal ancestry on human fitness have been thoroughly explored, there are fewer examples of adaptive introgression from Denisovans. Here, we propose a candidate gene for adaptive introgression, MUC19, for which some modern humans carry the Denisovan haplotype. MUC19 encodes a member of the gel-forming mucin protein family related to lacrimal and salivary gland function in humans and has been previously proposed as a top candidate for positive selection in Indigenous American populations. We find the Denisovan-like haplotype of MUC19 at its highest frequency in admixed American populations. Interestingly, the Denisovan harbors nine missense mutations in this region, all of which are fixed in modern-day individuals harboring the introgressed Denisovan-like haplotype. Furthermore, after conducting formal tests of selection, we find that the Denisovan-like haplotype of MUC19 is responsible for the signals of positive selection in admixed American populations. Lastly, we find that late Neanderthals also carry the Denisovan-like haplotype in the heterozygous state, while the earlier Altai Neanderthal is more diverged than expected from the Denisovan in this region. Our results suggest a complex pattern of multiple introgression events throughout human history at MUC19, first from Denisovans into late Neanderthals, followed by Neanderthal introgression into modern humans, and then selection for the introgressed haplotype in the Americas, which may have played a unique role in the evolutionary history of Indigenous American populations.
Uncovering South American genomic diversity in the Southern Cone

Epifanía Arango Isaza

Epifanía Arango-Isaza, Patrick F. Reilly, Marcos Araújo Castro e Silva, Tábitha Hünemeier, Felipe I. Martínez, Serena Tucci, Kentaro K. Shimizu, Chiara Barbieri
Pontificia Universidad Católica de Chile (Chile), Universitat Pompeu Fabra (Spain), University of Cagliari (Italy), University of Zurich (Switzerland), Yale University (USA)

South America's rich cultural and historical diversity contrasts with its remarkably reduced genetic diversity, caused by a recent bottleneck that occurred during the initial peopling. The extent to which this lack of genetic variation limits the power of population genetic analysis is debated. However, available genetic data from Native American ancestries in South America are mostly restricted to uniparental markers and SNP chips, the latter often designed with inadequate representation of target SNPs from the continent. South America's genomic diversity, therefore, still lies largely unexplored, leaving gaps in our understanding of South American demographic history. To fill this gap we generated 20 high-coverage genomes from populations in Central-Southern Chile, focusing on the Southern Cone of South America, and analyzed them with other high-quality ancient and modern genomes. We observed that population profiles inferred from whole genomes differ from directly comparable SNP chip data, highlighting the ascertainment bias intrinsic in the SNP chip. High within-population genetic diversity in the Southern Cone is indicative of a sustained larger population size in contrast to Amazonia, where European colonization has exerted a more pronounced impact on population decline. We also detect a minor pulse of gene flow from ancient migration sources which appears specifically in the Native American ancestry of Central-Southern Chile. Our study contributes to the growing number of genomes from South America, uncovering previously understudied genomic diversity and connecting present-day indigenous ancestries with the ancient population dynamics of the region's first migrants.
Adaptation in Ancient Maya

Esther Brielle

Esther S. Brielle, Vera Tiesler, Keith M. Prufer, Gabriel D. Wrobel, Douglas J. Kennett, Takeshi Inomata, Travis Stanton, Richard George, Aubree Marshall, David Reich

Harvard (USA), Michigan State University (USA), Riverside (USA), Santa Barbara (USA), Universidad Autonoma de Yucatan (Mexico), University of Arizona (USA), University of California, University of New Mexico (USA)

Our study analyzes ancient DNA from 421 Maya individuals, spanning 200 to 9500 years before present, to identify genetic adaptations to environmental and societal shifts. The consistent genetics of the Maya over millennia allow a general linear model to analyze genotypes over time. Preliminary results reveal significant selection in genes related to body weight, muscle mass, height, and metabolic traits, including glucose regulation and blood flow. Notably, chromosomes 6, 7, 9, and 18 exhibited strong selection signals, with chromosome 6 being particularly significant for genes such as E2F transcription factor 3, associated with body fat and muscle, CDK5 regulatory subunit associated protein 1-like 1, linked to diabetes and metabolic diseases, and neurensin 1, related to growth and various protein and blood-related measurements. Among the identified loci, cytochrome P450, family 3, subfamily A on chromosome 7, recognized for its role in hormone processing that results in increased strength and muscle mass, and nuclear factor, erythroid 2-like 3 pseudogene 1 on chromosome 18, related to heart size and ethanol metabolism, were also highlighted. These findings suggest adaptations to changes in diet and lifestyle associated with the transition from foraging to farming, urbanization, and long-distance trade, without reliance on domesticated pack animals. We are further investigating differences between elite and nonelite groups, highland vs. lowland adaptations, and the timing of these genetic adaptations, potentially in response to dietary shifts and microbial changes. This study will enhance our understanding of the Maya’s evolutionary responses to their changing environment.
Human migration leads to the contact of populations that have been previously separated, often exposing them to different environments. The gene flow resulting from migration is an opportunity for beneficial alleles, previously adapted in a source population, to be introduced into the derived admixed population, leading to adaptation. Since the last centuries, America has been a target for numerous migration events, leading to modern admixed populations. In this study, we analysed ten thousand participants from the Biorepository and Integrative Genomics (BIG) initiative, that enrolled over 35k participants with electronic health records. Admixture deconvolution from genetic data indicates that 50% of the participants have non-European origins; this includes 20% African and 30% with admixed origins, with two-, three-way, and multiple admixture patterns. The most frequent two-way admixture (23%) includes that of Europeans with Africans, Indigenous from the Americans, or East Asians. Leveraging the lengths of ancestry tracts, we detected a strong signal of positive selection (six standard deviations above the mean, measured by iDAT) in the Indigenous-European admixed group. This signal is located in a 10Mb region on chromosome 1 that includes the DARC locus, a well-known example of human adaptation to malaria, among other factors. We used approximate Bayesian computation to demonstrate that this signal is compatible with a demographic scenario that implies recent admixture (12-13 generations ago) and natural selection (s=0.02-0.06). Our results highlight the importance of further studying the impact of recent gene flows in America.
Ancient genomes from the northern Southern Cone document thousands of years of continuity of a deep South American lineage

Javier Maravall López, Josefina M. B. Motti, Nicolás Pastor, Rodrigo Nores, David Reich

(Argentina, 5000, 5000 (Argentina), 7631 (Argentina), Broad Institute of Harvard and MIT (USA), CONICET, Departamento de Antropología, Department of Genetics, Department of Human Evolutionary Biology, Facultad de Ciencias Sociales, Facultad de Filosofía y Humanidades, Harvard Medical School (USA), Harvard University (USA), Howard Hughes Medical Institute, Instituto de Antropología de Córdoba (IDACOR), NEIPHPA-CONICET, Universidad Nacional de Córdoba, Universidad Nacional de Río Cuarto (Argentina), Universidad Nacional del Centro de la Provincia de Buenos Aires

The northern Southern Cone of South America is underrepresented in modern and ancient genomic surveys. To address this gap, we assembled new genome-wide data from more than 250 ancient individuals from this world-region, spanning a time transect up to 8000 years before present. We conclude that most of the ancient individuals we report in this study are descended from a deep lineage strongly represented in the central region of present-day Argentina that was already defined by 7600 years BP and persisted as the main ancestry component of these populations for thousands of years after. We also find evidence of substructure, significant differences in community sizes and mating patterns, and comparatively small subsequent contributions of other sources, including one related to the present-day Arhuaco people and one associated with ancient California-Channel Islands individuals, replicating a similar signal previously found in Central Andes ancient individuals. Apart from the main Central Argentina-associated component, individuals from the northwesternmost region of the country carried a large proportion of Central Andes-related ancestry, and individuals from the Gran Chaco region could be modeled as being entirely descended from a source related to the present-day Wichí people. Our findings highlight the remarkable genetic continuity of this understudied world-region, while providing fine-grained insights into networks of contacts of ancient groups within it. We present this work on behalf of a wider group of scholars who collected, generated, and analyzed material for this study.
Local adaptive retention of high amylase copy number haplotypes in Peruvian potato farmers at historical time scale

Kendra Scheer

Kendra Scheer, Luane Landau, Kelsey Jorgensen, Charikleia Karageorgiou, Lindsey Siao, Angelis Morales-Rivera, Tom Brutsaert, Abigail Bigham, Omer Gokcumen
Los Angeles (USA), Recinto de Río Piedras (Puerto Rico), Syracuse University (USA), Universidad de Puerto Rico, University at Buffalo (USA), University of California

Starch digestion has a profound but variable impact on the evolution of human populations. The amylase gene (AMY1), which encodes for the namesake starch-digesting enzyme, shows remarkable copy number variation among humans, linked with metabolic diseases, microbiome, and adaptation to starch-rich diets. Recent studies have resolved the complex structural haplotypes and hypothesized that haplotypes harboring increased AMY1 gene copy numbers are increased in allele frequency due to recent soft sweeps in agricultural populations. To test this hypothesis, we utilized multiple bioinformatic approaches and digital PCR to show that Pima from northern Mexico and Quechua from Peruvian Andes harbor the highest number of AMY1 gene copies among world populations. To investigate the genetic variation in Andeans we sequenced 10 Peruvian individuals using Oxford Nanopore long-read sequencing to obtain 20 amylase haplotypes. Population genetic analyses suggest that the increased amylase gene copy number variation is a genome-wide outlier between Peruvian and neighboring populations (top 0.0004761653 percent), suggesting recent local selection. Previous work has shown that Peruvians relied heavily on starch-rich potatoes for their caloric needs following their domestication in the Andes around 6000-10000 BP. Prior work has tied this reliance to selection at the MGAM locus, another starch-digesting enzyme, in both ancient and modern Andeans. Thus, we hypothesize this strong dependency on potatoes has affected both enzymes in the starch digestion pathway in Peruvian populations. Collectively, our work illustrates how local dietary traditions can facilitate the adaptive retention of specific structural variations at historical timescales.
Ancient genomic data reveals multiple genetic shifts in the demographic history of Colombia
Kim-Louise Krettek

Kim-Louise Krettek, Lilian Andrea Casas Vargas, Alex Mas-Sandoval, Leonardo Arias Alvis, Ella Reiter, William USAquen, Jose-Vicente Rodriguez Cuenca, Cosimo Posth
Archaeo- and Palaeogenetics Group, Department of Archaeogenetics, Department of Evolutionary Genetics, Dipartimento di Scienze Statistiche "Paolo Fortunati", Grupo de Genética de Poblaciones e Identificación, Grupo de Investigación en Antropología Biológica, Institute for Archaeological Sciences, Leiden University Centre for Linguistics (Netherlands), Max Planck Institute for Evolutionary Anthropology (Germany), Senckenberg Centre for Human Evolution and Paleoenvironment, Universidad Nacional de Colombia (Colombia), University of Tübingen (Germany), Università di Bologna (Italy)

The first peopling of America via the Beringian land bridge has been extensively researched. Scholars have proposed that the ancestral population that first reached North America split into two main lineages, one persisting in the northern sub-continent and the other spreading southwards to reach the southern tip of South America by at least 14,500 years ago. Recent archaeogenetic studies have begun outlining the genetic history of various South American regions such as the Andes, the Atlantic coast, and the Southern cone. However, Colombia, the entry point into the southern sub-continent, has so far been omitted from ancient genomic studies. To illuminate Colombia’s crucial role in the initial settlement of South America and reconstruct the demographic history of local populations, we generated genome-wide data from 21 ancient individuals spanning the last 6,000 years. Our genomic time-transect, encompassing a spatially confined area on the central Colombian plateau, reveals at least two shifts in genetic ancestry over time, one associated with Mesoamerican populations and the other with present-day indigenous Colombian groups. Moreover, we unveil a complex interplay between ancient populations from Colombia and the circum-Caribbean region. This diachronic investigation into demographic transformations within Colombia holds promise for uncovering previously unknown population movements in northern South America. Additionally, our newly generated data enriches the sparse ancient genomic record from the tropics, fostering further population genetic analyses across the American continent.
Identifying signals of natural selection in unadmixed individuals of the Mexican Biobank

Maria Fernanda Miron Toruno

Maria F Miron-Toruno, Ram González-Buenfil, Santiago Medina-Muñoz, Andrés Moreno-Estrada, Austin W Reynolds
Baylor University (USA), Cinvestav (Mexico), Unidad de Genómica Avanzada

Over the past decade, the exponential increase in computational power and the development of next-generation sequencing technologies have facilitated the identification of genomic regions that appear to be shaped by natural selection. Studies of natural selection represent a genomic approach to understanding human evolution and differences in disease risk across many populations worldwide. However, the limited availability of diverse cohorts with genomic data impairs the effective implementation of selection analyses in many underrepresented regions, such as Latin America. Recently, Mexico generated the Mexican Biobank (MXB), one of the largest genomic resources of nationwide coverage in the Americas. The MXB comprises over 6,000 individuals from all 32 states and 898 sampling localities across Mexico, including over 1,500 unadmixed individuals of indigenous ancestry, genotyped for 1.8 million genome-wide SNPs. In this project, we aim to investigate the genetic footprint of recent adaptive pressures on Indigenous Mexican populations by analyzing a subset of ancestry-masked samples with >90% Indigenous American ancestry from the MXB. To do this, we will compute different selection statistics, including iHS, XP-EHH, nSL, and PBS. Moreover, we will be able to connect putatively selected alleles to molecular phenotypes reported in the MXB to better understand how selection has shaped traits of biomedical relevance. To date, this project represents the most extensive selection study in a national-scale Mexican cohort and the first to pair selection analyses with biomedical data to gain insight into the genetic basis of Mexican health-related traits.
FUNCTIONAL GENOMIC INSIGHTS INTO THE EVOLUTION OF THE ARSENIC TOLERANCE IN ANDEAN POPULATIONS

Mario Andres Apata

Mario A Apata, Sofia Acosta, Wyatt Comin, Klevis Lasku, Bridget Diviak, Ricardo Verdugo, Mauricio Moraga, Anne Stone, Jeanne Wilson-Rawls, Melissa Wilson

Arizona State University, Center for Evolution and Medicine, Chile. (Chile), Departamento de Antropología, Evolutionary Biology Graduate Program, Facultad de Ciencias Sociales, Facultad de Medicina, Institute of Human Origins, Instituto de Investigación Interdisciplinaria, Molecular and Cellular Biology Graduate Program, Programa de Genética Humana, School of Human Evolution and Social Change, School of Life Sciences, Tempe, USA (USA), USA. (USA), Universidad de Chile. (Chile), Universidad de Chile. Santiago de Chile. (Chile), Universidad de Talca

Arsenic causes severe toxicological effects in humans, yet some human populations have evolved tolerance. Many Andean people including those in the Camarones valley in northern Chile, as well as others in northwest Argentina, have been exposed environmentally to arsenic in their food and water supply over thousands of years. These populations carry conserved single nucleotide mutations in noncoding regions of the gene for a key enzyme known as the Arsenic (+3 oxidation state) Methyltransferase or AS3MT. AS3MT metabolizes arsenic in the liver; however, these Andean specific mutations have an unknown impact on the expression of this gene. Therefore, our aim was to evaluate experimentally whether some of these mutations might alter AS3MT gene expression in the liver. To address this, we selected four candidate mutations based on (1) whether they have been evolving under positive selection, (2) population genomic analyses comparing the Camarones and other indigenous populations from South America to characterize the allele frequency of these mutations, and (3) the intersection with potential regulatory regions. Then, we cloned small genomic DNA fragments carrying the Andean and the reference alleles (European) into a luciferase reporter plasmid. HepG2 cells, a human hepatocyte tumor cell line, were transfected with the different Luciferase reporter clones and treated with different arsenic concentrations. We observed significant differences between the Andean and European alleles when cells were exposed to arsenic. Our work contributes to an integrated perspective between evolutionary genomics and molecular genetics to understand how Andean people evolved in an arsenic-rich environment.
Colombian Arawak genomes shed light on population demography of South America

Marisol Naydu Espitia Fajardo

Marisol N Espitia Fajardo, Daniela Orozco Perez, Juan P Peña Sanchez, Angela M Zamudio Jacome, Yuberlis Gonzalez Ipuana, María D Ortega, Guillermo Barreto, Andres Moreno Estrada
Centro de investigación y de estudios avanzados del IPN (Cinvestav) (Mexico), Universidad del Valle (Colombia)

Population genetic studies in South American indigenous groups are essential to better understand their complex demographic history. However, genomic data available for these populations remain scarce, which makes it difficult to reconstruct their population dynamics, dispersal patterns, ancestral origins and migratory routes within the continent. As the main gate to South America, present day Colombia plays a key role in understanding such evolutionary dynamics. To fill this gap, we genotyped 120 Colombian indigenous individuals belonging to the Arawak linguistic family, the largest in South America, together with individuals of the Chibcha-Paezen family. Genomewide data for 1.8 million SNPs were generated using MEGA arrays and combined with previously published data from Native American individuals. We performed analyses of global ancestry and population structure, including PCA, ADMIXTURE, as well as IBD sharing and F statistics to infer the genetic history of these populations. Arawak populations showed genetic affinities with both Andean and other Amazonian populations, which could be explained by the extended range of ecoregions inhabited by Arawak speaking groups across South America. This suggests that the Arawak are not strictly substructured into genetic clusters from the Andes and the Amazon, but rather show patterns of shared ancestry throughout their expansion range. This also implies that the Arawak language did not spread exclusively through cultural transmission, as we found genetic connections between Arawak speaking communities over long geographic distances across South America. Our study contributes to the population demography reconstruction of overlooked indigenous cultures that have shaped present-day genetic diversity of South America.
Skin pigmentation is a complex trait influenced by genetic, environmental, and evolutionary factors. The Mayan population of the Yucatan Peninsula exhibits a diverse range of skin tones, reflecting their unique genetic ancestry and environmental exposures. Understanding the genetic variants underlying skin pigmentation in this population can provide insights into the evolutionary history and adaptation to local environmental conditions. In this study, we conducted a comprehensive genetic analysis in two Mayan communities of the Yucatan Peninsula to establish the degree of association with pigmentation phenotypes of six variants previously reported in the literature (HERC2-rs12913832*A, OCA2-rs1800404*C, BNC2-rs2153271*T, SLC45A2-rs16891982*C, IRF4-rs12203592*T and TYR-rs1042602*A). The ancestral composition of individuals was estimated using the Human Origins Axiom1 microarray. Our analysis revealed a complex genetic structure in the Mayan individuals, associated with demographic dynamics of the Mayan population with Spanish and African populations during colonial times, involving pigmentation-related genes. Overall, our study provides insights into the genetic basis of skin pigmentation variation in the Mayan population of the Yucatan Peninsula, highlighting the role of genetic ancestry. This research contributes to a deeper understanding of human diversity and adaptation in the context of indigenous populations and sheds light on the evolutionary history of pigmentation traits in the Americas. As well as in the use of categories that fail to reflect the reality of genetic diversity and evolutionary dynamics.
Reconstructing Western and Northern Mexico’s Past
Roslyn Curry

Roslyn Curry, Jakob Sedig, Jose Luis Ponzo
Harvard University (USA), Instituto Nacional de Antropología e Historia (Mexico), n/a (USA)

Though ancient DNA has revolutionized our understanding of the past, America/Mexico remain understudied. The Proyecto de investigación de poblaciones antiguas en el norte y occidente de México (PIPANOM) addressed this imbalance through the examination of over 400 ancient individuals from central, western and northern Mexico. Data from these individuals have revealed population structure that is broadly associated with geographic distance across what is now modern Mexico, Belize, Guatemala, and the southwest United States. These data have also provided new insights into long-standing questions about migration and interaction of different archaeologically defined cultures in key eras of Mexico’s past, such as the movement of people during the Postclassic period. Along providing new insights in an area that has been understudied with aDNA, PIPANOM has helped correct institutional imbalances in archaeological and paleogenetic research. PIPANOM is an archaeologist-led project that has brought together more than 20 collaborators. As part of the project, there have been numerous workshops, trainings, and visits bringing together Mexican and American archaeologists, geneticists, researchers, analysts, and students. As such, PIPANOM provides a roadmap for equitable large-scale projects that produce robust scientific results.
Inferring fine-scale demography of indigenous lineages from 500 whole genomes across Mexico

Santiago G Medina-Muñoz


Advanced Genomics Unit (UGA), CINVESTAV (Mexico), Department of Cross-Cultural and Regional Studies (Denmark), Department of Integrative Biology, GenomeAsia 100K consortium (Singapore), Nanyang Technological University (Singapore), National Laboratory of Genomics for Biodiversity (LANGEBIO), Singapore Centre for Environmental Life Sciences Engineering, The Asian School of the Environment, University of Copenhagen, University of Wisconsin-Madison (USA)

Genetic studies of population history in the Americas have traditionally focused on early settlement and dispersal, often overlooking the local dynamics in areas with a rich Prehispanic heritage, like Mexico. This oversight misses critical insights into the genetic landscape shaped by these regional movements. To deepen our understanding of Mexico’s demographic history, we sequenced 500 present-day genomes at 30x coverage, capturing a broad spectrum of indigenous ancestries nationwide. This effort represents the most comprehensive sequencing project of indigenous ancestries by leveraging fine-scale sampling of the MX Biobank. By integrating these genomes with those from indigenous populations in South America and additional diverse global ancestries, we assessed a series of demographic hypotheses. We use gene-genealogy based methods for exploratory analysis and linkage disequilibrium statistics to infer demographic models. Our demographic reconstruction is further enriched by insights drawn from linguistics, as well as historical and archaeological records. We found previously unreported genetic relationships among Mesoamerican populations, particularly among the Nahuatl-speaking peoples who led one of the major expansions within Mesoamerica. We identified a shared genetic component among central Nahua populations, while also uncovering more distant connections: West Nahua groups exhibit a mixture of central and northern Mexican ancestries, whereas East Nahua groups blend central Nahua with Southeast Mayan ancestries. Additionally, our findings highlight genetic connections with South American populations, emphasizing the intricate demographic interplays shaping modern Mexico. This research provides new insights into Mexico’s genetic landscape, with significant implications for historical narratives and genomic studies in indigenous populations.
Evolutionary Changes Throughout Human History in America
Tábita Hünemeier

Presented by self
University of São Paulo (Brazil)

The American continent is geographically vast, spanning all northern and southern latitudes, and was populated relatively late in the expansion of Homo sapiens. However, the internal spread across the continent was remarkably rapid following the Out of Beringia movement. Various evolutionary processes have guided the unique history of American settlement, including extreme and subsequent bottlenecks, forced and voluntary migrations, natural selection, gene-culture coevolution, and finally, changes in population structure caused by European invasion and subsequent colonization of the continent. Genomic data from Indigenous American populations and populations derived from Indigenous peoples (admixed Americans) are rare and underrepresented in both medical and population genomic studies. To fill this gap, we generated thousands of high-coverage genomic data from Indigenous and admixed populations from America. The results presented here aim to elucidate their evolutionary history by detailing the processes that led to the formation of the current American population and identifying the prevalent evolutionary factors and processes that shaped this human group from the initial settlement around 15,000 years ago to the impact of contact with Europeans and Africans centuries ago.
Picuris Pueblo oral history and genomics reveal genetic continuity in North American Southwest
Thomaz Pinotti

Centre for GeoGenetics, Department of Archaeology, Department of Zoology, GLOBE Institute, Laboratório de Biodiversidade e Evolução Molecular (LBEM), Musée de l’Homme, Picuris Pueblo (USA), Southern Methodist University (USA), Universidade Federal de Minas Gerais (Brazil), University of Cambridge (United Kingdom), University of Copenhagen (Denmark), Université de Paris (France)

Archaeological, ethnographic, and anthropological evidence, along with oral tradition, establishes a direct connection between Ancestral Puebloans and present-day Indigenous Puebloans. However, policymakers often overlook or downplay this evidence when making decisions about Ancestral Puebloan archaeological sites and heritage. Despite the potential of ancient DNA to support these claims, decades of exploitative research ethics have led to mistrust among Indigenous peoples of North America, and the few genetic studies conducted so far have been driven by the necessity of repatriation claims rather than a desire to understand tribal histories. Here we report the first study in which an Indigenous group, the sovereign nation of Picuris Pueblo in New Mexico, initiated collaboration with geneticists and archaeologists to fill in gaps in traditional knowledge and further their understanding of their population history and ancestry. We generated 29 genomes from 16 ancient individuals spanning the last millennium and 13 present-day members of Picuris Pueblo. We show genetic continuity between the ancient and present-day Picuris, and more broadly between Picuris and other ancient individuals from Ancestral Puebloan sites in the region. Haplotype-based methods show no indication of a population decline before European contact, indicating that the depopulation of some Pueblo sites represented a reshuffling of peoples in the region rather than a demographic collapse. Our collaboration illustrates an approach that equally privileges and complements oral traditional, archaeological, ethnographic, and genetic lines of evidence while also ensuring the Indigenous community’s control of their genetic data.
The analysis of high-coverage sequences of two archaic humans: Neanderthal and Denisovan, have revealed that they interbred with anatomically modern humans (AMH). Recent studies in European and East Asian cohorts have shown that Neanderthal and Denisovan introgression facilitated adaptation and contributed to complex traits associated with pigmentation, the resistance to pathogens and drug metabolism. However, the impact of archaic variation on complex traits remains under-explored in admixed populations from the Americas. Here, we leverage genome-wide genotype data from the Mexican Biobank Project (MXB) to investigate the genomic landscape of archaic variation in Mexico. We inferred a map of 113,542 Neanderthal introgressed variants (NiVs) and 8,374 Denisovan introgressed variants (DiVs) identified in Mexican individuals. We conducted and validated a rigorous pipeline to estimate the contribution of Neanderthal introgressed variants to the heritability of 31 distinct phenotypes measured in 5,753 individuals in MXB. We find that Neanderthal introgressed variants do not significantly contribute to trait variation (on average, NiVs explain ~ 0.23% of trait variation, pval > 0.05). Consistent with previous studies (Wei et al., 2023), we find that Neanderthal introgressed variants are depleted of heritability for all traits considered. Our study provides novel evidence on the impact of Neanderthal introgression on complex trait variation in Mexicans and represents a valuable resource for future studies characterizing archaic introgression in underrepresented populations in genomic studies.
Unraveling transpacific immigrations to Peru and the Philippines between the mid-16th and early 20th centuries

Zara Paulina Martínez-Sánchez


Big Data Institute, CINVESTAV (Mexico), Centro de Ciencias Genómicas (CCG), Department of Biomedical Data Science, Laboratorio Nacional de Genómica para la Biodiversidad (UGA-LANGEBIO), Li Ka Shing Centre for Health Information and Discovery, Stanford University (USA), The Wellcome Centre for Human Genetics, Universidad Nacional Autónoma de México (UNAM) (Mexico), University of Oxford (United Kingdom)

The migratory stories of people who crossed the Pacific, thanks to maritime routes that connected Latin America and Asia since the mid-16th century with the Manila Galleon have been overlooked despite often carrying a narrative of discrimination and even slavery. This study dives into the history of some of these migrations by combining historical and computational methods for ancestry inference such as Admixture and Gnomix, MAAS-MDS to marcate specific origins, and Tracts to time migrations, using the 1000 Genomes Project, Human Genome Diversity Project, Oceania Genome Variation Project and a Philippine dataset. Particularly, we focus on migrations from East Asia to Peru, estimating two pulses of migration that occurred in the mid-19th and early 20th centuries that are consistent with the records of migrations from China and Japan to Peru, in addition to a third migration pulse estimated to occur in the early 18th century. We also map Asian ancestries in these individuals to China, Japan and Myanmar. Focusing on the west side of the Pacific, we also study migrations from Latin America to the Philippines based on historical information of transpacific migrations through the Manila Galleon during the Colonial Period (Mawson, 2016), finding the presence of American Indigenous ancestry in varying proportions among individuals from different regions of the Philippines, larger than those observed in other East Asian cohorts. This way, our work investigates and makes visible a piece of the migratory history that occurred between Latin America and Asia in the last 500 years.
Zehra Köksal, Claus Børsting, Graciela Bailliet, Germán Burgos, Elizeu Carvalho, Adriana Castillo, Carlos David Neyra Rivera, Danielle Malheiros, Marília Brito Gomes, Beatriz Martínez, Humberto Ossa, María Laura Parolin, Alfredo Quiroz, Ulises Toscanini, Irina Velázquez, Carlos Vullo, Leonor Gusmão, Vania Pereira

The Native Americans’ (NAMs) migration routes through South America are not entirely understood. Phylogenetic studies on Y-chromosomal haplogroups defined by SNPs, give a unique view on male geographic dispersals and can be used to trace back migrations in historic times. The two main NAM founding lineages Q1b1a1a1-M848 and Q1b1a2-Z780 harbor valuable information on some of the long-lost NAM migration routes in South America. Here, we introduce the custom-designed massively parallel sequencing panel NAM-Q-Y, which targets 1,569 variants of NAM haplogroup Q sublineages. Data were generated in 422 admixed individuals of NAM male ancestry from different regions of South America. Sequencing data from the 422 individuals were combined with 222 publicly available genomes from Central and South American individuals, which further resolved the phylogenetic tree of the NAM haplogroup Q lineages, introducing more than 100 novel or refined branches. The distribution of selected sublineages within Q1b1a1a1-M848 and Q1b1a2-Z780 allowed analyses of regional migration patterns that supported previous hypotheses on (1) southward movements along the Andes, (2) independent westward and eastward movements from the north of South America, and (3) southward Andean spread followed by eastward movements, where the population split into a northward (to Northeast Brazil) and a southward (to East Argentina) migrating subpopulation. Furthermore, the first clear patterns of inland movements in the Southern Cone were reported based on Y-chromosomal data. The presented findings support and complement previously hypothesized movements, and increase our understanding of the NAMs migration routes through South America.
Admixture has played a critical role in shaping the genetic diversity among human populations. Exploring the genomes of diverse and deeply divergent present-day human populations such as the Khoe-San populations of Southern Africa and Central African rainforest hunter-gatherers (RHG) is key to understanding our evolutionary history as humans. Although several studies have previously suggested contributions from unknown deeply diverging lineages into the ancestors of different present-day African populations, there is no coherent picture due to differences in datasets, methods and frameworks considered. This problem is particularly difficult since no source genomes from archaic populations are available. Using methods that do not require the reference archaic genomes, we analyzed 25 high-coverage genomes from five different Khoe-San populations and a set of genomes from different RHG populations, in the context of African genomic diversity. We explored two models of early human population structure: the reticulation of multiple stems migrating between early human populations, and archaic hominin admixture from deep ghost lineages. We found that a model with introgression from an unknown source population into the ancestors of Khoe-San populations better explains patterns of outliers in terms of private polymorphisms than the weakly structured stem model. We identified genomic segments that represent putative fragments of ghost introgression in the Khoe-San individuals.
A Rich Resource of African-Descent Genomes (CAAPA2)
Brenna M. Henn

Presented by self
Davis (USA), University of California

The symposium illuminates the important role of genomics in deciphering the complex history and enhancing the healthcare of African and African-descendant populations. In recent decades, genetic research has crucially pinpointed Africa as the origin of anatomically modern humans, provided clarity on the "Out-of-Africa" migration and recent intra- and trans-continental migrations, and elucidated the intricate genetic landscape of the continent's diverse peoples. Studies using genetic information from mitochondrial DNA, Y-chromosomes, autosomal markers, and ancient DNA have revealed the genetic sub-structure of African populations, offering insights into prehistoric human dynamics and contemporary genetic diversity. Crucially, this genetic mosaic has been shaped by subsistence strategies, cultural backgrounds, environmental factors, and diseases, informing not just anthropological understanding but also contemporary health. With increasing genomic datasets from various African populations, including ancient DNA, researchers are poised to unravel further secrets of human population genetics and evolution. This symposium will therefore bring together novel findings of genetics to chart the history, diversity, and selection in African and African-descendant populations, promising significant contributions to the global understanding of molecular biology, human health and evolutionary history. Session contributions will include a global assembly of researchers whose studies span the breadth of African genetic research. The overarching goal is to foster interdisciplinary dialogue, integrate diverse research perspectives, and discuss the future of genomics in Africa and beyond.
Genetic signatures from the Dutch East India slave trade in southern Africa

Dana R. Al-Hindi, Stacy Edington, Austin W. Reynolds, Neus P. Font, Brenna M. Henn
Baylor University (USA), Davis (USA), University of California, University of Colorado (USA)

People who self-identify as South African Coloured (SAC) are often described as five-way admixed, harboring majority of their genetics from indigenous Khoe-San (avg. 46%) populations, with other ancestries from equatorial African populations, Europeans, and South and East Asians. Our primary focus are the equatorial African, and South and East Asian ancestries that were introduced after 1652 through the Dutch East India slave trade. About half of the enslaved individuals brought to the Cape Colony between 1652 and 1808 were from South and East Asia, but nearly all of these were from the Indonesian archipelago (23%) or the Indian subcontinent (26%). Previous studies have frequently used the Han Chinese as a reference for the South and East Asian signals among the SAC people, but given the region's slave trade and historical record, we hypothesize populations from Indonesia and/or the Indian subcontinent would be more appropriate. Here, we fine tune which South and East Asian populations best serve as source population for those genetic components in the SAC. The SAC admixed individuals in our analysis stem the Northern and Western Cape (n=619) of South Africa. Reference populations were extracted from the 1000 Genomes project, alongside 162 whole genome Indonesian samples from the Indonesia Genetic Diversity Project. We combine the use of highly accurate local ancestry inferences and ancestry specific MDS to establish if populations other than the Han Chinese are more representative of the Asian components in the SAC.
Impact of Khoe-San gene flow in phenotypic variation of anthropometric and cardiometabolic traits
Dhriti Sengupta

Presented by self
University of the Witwatersrand (South Africa)

A defining feature of many Southern African populations is gene flow from the Khoe-San (K-S) hunter-gatherers. Our previous work based on data from the AWI-Gen study (a pan-African population cross-sectional cohort that includes ~5000 Southeastern Bantu speaking participants) has demonstrated this unique ancestry to not only impact population structure but also various population-genetic estimates and even imputation performance and accuracy. Employing multivariate regression analyses that includes K-S ancestry (KSA) proportions along with age, sex, socio-economic status (SES), education attainment, physical activity (PA), HIV, and smoking status, we have investigated the relative contribution of KSA to various anthropometric and cardiometabolic traits. The results for anthropometric traits show KSA proportion to be strongly associated with standing-height (SH), hip-circumference (HC), Waist to Hip Ratio (WHR), and Body Mass Index (BMI). Of these, BMI and HC were increased when KSA proportions were higher, whereas SH and WHR were decreased. Sex stratified analyses further demonstrated that the impact of KSA was generally stronger in females. We also detected evidence of interactions between KSA and factors such SES, PA, HIV for some of the anthropometric traits. Similar impact and interactions were also observed for cardiometabolic traits such as Diastolic blood pressure. Our findings suggest that K-S gene flow and the interaction between KSA and the environment could be a key factor in the unique body composition and disproportionately high incidence of obesity and hypertension in Southern Africa, especially in women. We also highlight the value of incorporating population-level genetic estimates in epidemiological studies, especially from Africa.
Sudanic, Central, and Western African ancestry contributions to ancient Ugandan Populations

Esther Brielle

Esther Brielle, Peter Schmidt, John Krigbaum, Jeremy Choin, Kendra Sirak, Shamam Waldman, Daniel Tabin, Jonathan Walz, Jackline Besigye, David Reich

Harvard (USA), School for International Training (USA), Uganda Museum and University of Pretoria (South Africa), University of Florida (USA)

Our analysis of ancient DNA from Uganda’s Ndali Crater Lakes region uncovers a complex admixture history and genetic diversity shaped by millennia of migration and interaction. Examining 15 ancient individuals from 400 to 1600 CE revealed a dynamic interplay of Sudanic African, Central African, West African, and eastern African pastoral ancestries. Early admixture between Sudanic and Central African groups was evident in the Ugandan western Kansyore population from approximately 1500 BP, indicating admixture occurred in the first millennium BCE, predating the individuals examined. Subsequent significant gene flow from West African populations, likely Bantu speakers, in the first half of the first millennium CE, contributed to a genetic makeup of approximately one-third western Kansyore-associated and two-thirds West African-associated ancestry in several ancient Ugandans. Additionally, modern Ngumba and Tikar South populations from Cameroon, despite dense rainforest barriers, share genetic links with ancient Ugandans, suggesting historical interactions that contributed to the broader genetic mosaic of Central Africa. Using principal component analysis, f-statistics, qpWave, qpAdm modeling, and modified ChromoPainter, we reconstruct the admixture sequence, providing insights into the genetic history of the Western Rift and Lake Victoria to elucidate the relationships between ancient genetic ancestries from across Africa and their legacy in both ancient and present-day populations in Uganda and beyond.
Evolutionary History of Transient Receptor Proteins in Sub-Saharan Africa

Jasmin Rees, Shaohua Fan, Jeffrey Spence, Yuanqing Feng, Matthew Hansen, Jonathon Terhorst, Alessia Ranciaro, Jibril Hirbo, William Beggs, Neil Thomas, Thomas Nyambo, Sununguko Wata Mpoloka, Alfred Njamnshi, Charles Fokunang, Gurja Belay, Yun Song, Sarah Tishkoff, TOPMED CONSORTIUM

Addis Ababa University (Ethiopia), Fudan University (China), Kampala International University in Tanzania (Tanzania, Stanford University (USA), The University of Yaounde ? (Cameroon), United Republic of), University of Botswana Gaborone (Botswana), University of California (USA), University of Michigan (USA), University of Pennsylvania (USA)

Transient receptor potential (TRP) channels are expressed across a number of different skin cells, including sensory neurons, keratinocytes, melanocytes and skin-resident immune cells. As such, skin expressing TRP channels not only act as chemical, thermal or mechanical sensors, but also play a key role in the growth and development of skin and hair, as well as in mediating immunological and inflammatory responses of the skin. Various TRP genes have hence been implicated across a wide range of skin disorders and diseases, including atopic and pruritic dermatitis, psoriasis, alopecia, rosacea, and skin cancer. Using a large number of high-quality whole-genome data from diverse sub-Saharan Africans, together with publicly available non-African genomes, to trace the evolutionary history of TRP genes across sub-Saharan Africa. We demonstrate that TRP genes contain alleles that are highly differentiated amongst African populations, including within and between major languages groups, and have likely been targets of positive selection. We evaluate the role of abiotic factors in driving the genetic variation of TRP genes across sub-Saharan Africa, considering the potential thermal and UVB selective pressures acting on TRP genes with known thermoregulatory or melanocyte function. Finally, using large-scale BioBank data, we identify associations between common and selected variants within TRP genes and disease traits, specifically skin-based disorders. With thanks to the TOPMED Consortium and funding sources ADA 1-19-VSN-02, NIH grants 1R35GM134957, R01DK104339, and R01AR076241, and 1X01HL139409-01.
Genomic Insights into the population history of southern-central Africa
Jeremy Choin

Jeremy Choin, Maggie Katongo, Kendra Sirak, Elizabeth Sawchuk, George Mudenda, Potiphar Kaliba, Lovemore Mazibuko, Menno Welling, Alan Morris, Kathryn de Luna, Nadin Rohland, Mary Prendergast, Jessica Thompson, David Reich
Amsterdam University of the Arts (Netherlands), Department of Anthropology, Department of Genetics, Department of History, Department of Human Biology, Department of Human Evolutionary Biology, Georgetown University (USA), Harvard Medical School (USA), Harvard University (USA), Livingstone Museum (Zambia), Malawi Department of Museums and Monuments (Malawi), Rice University (USA), Stony Brook University (USA), Yale University (USA), the University of Cape Town (South Africa)

The present-day genomic, cultural, and linguistic landscapes of southern-central Africa (Malawi and Zambia) were drastically changed by a rapid population expansion from western to southern Africa that started after ~3500 years ago. These Bantu speakers brought new technologies and new forms of subsistence to regions that were previously occupied by hunter-gatherers. The details of these interactions are especially poorly resolved in southern-central Africa, which was also historically involved in long distance trading networks. Here, we generated ~70 new genome-wide data from the south-central Africans who lived within the last ~1500 years. We investigated admixture history using an extended version of Chromopainter (PHCP) adapted to work on unphased pseudo haploid data. First, we demonstrated that PHCP could serve as a viable alternative/complement to the commonly employed qpAdm software. Utilizing PHCP, we found that the genomes of southern-central Africans were mostly composed of Bantu-related ancestry and varying proportions of hunter-gatherer-(5%-15%), Nilotic pastoralist-(0%-6%) and Afroasiatic-related(0%-2%) ancestries. One historical individual buried at the fishing site of Kalala Island in the Kafue River, Zambia, presents significantly more hunter-gatherer-related ancestry (~45%)—most likely from Later Stone Age foragers with ancestry similar to that found in ancient individuals from Malawi/Zambia (16,000 - 2,000BP) and to a lesser extent from Central Africans—showing that this foragers ancestry profile, ubiquitous until ~2000 BP, persisted in the region. This supports local oral history that details encounters between Bantu-speaking peoples and resident foragers.
Ancient mitogenomes highlight legacy of Transatlantic Slave Trade in Peru
Laura N Pott

Laura N. Pott, Jaime Zolik, José Luis Santa Cruz Alcalá, Claire Maass, Maria A. Nieves-Colón
APACHETA Consultores (Peru), Museo Virtual Afroperuano de San Luis de Cañete (Peru), SNA International (USA), Universidad Nacional Mayor de San Marcos (Peru), University of Minnesota (USA)

Contemporary Latin American populations have complex admixture patterns due to colonization, indentured servitude, and the Transatlantic Slave Trade. However, the population history of Afro-descendant communities remains understudied. As part of a community-engaged research project and the first ancient DNA (aDNA) study of an Afro-diasporic community in South America, we use aDNA to investigate patterns of mitogenomic diversity among individuals buried in the cemetery of the 18th century Hacienda La Quebrada sugar plantation in Peru. We isolated aDNA and prepared single-stranded and double-stranded libraries for 30 individuals, followed by shotgun sequencing and mitochondrial enrichment. We recovered 36 mitogenomes (average depth of 190.5X), retaining 32 mitogenomes (average depth of 214X) representing 28 individuals for further analysis after considering coverage and contamination. We successfully obtained 32 mitochondrial haplogroup assignments with an average quality of 93% using Haplogrep 3 and a hard-filtering variant coverage threshold of 5X. 75% of the individuals carry haplogroups L0, L1, L2, or L3 (associated with African populations); 21.4% carry A2, B2, C, or D1 (associated with Indigenous American populations); and 3.6% carry H1bw (associated with Iberian populations). Median joining networks were generated using popART to visualize phylogenetic relationships. These diverse results highlight several mitolineages that have not been commonly reported among enslaved populations in the Americas, including Eastern Africa and the Iberian Peninsula. They also suggest that contemporary admixture patterns in Afro-Peruvians are longstanding, providing insight into the demographic history and genomic impacts of slavery in Afro-diasporic communities in Latin America.
Heterogeneous genetic architectures and evolutionary genomics of prostate cancer in Sub-Saharan Africa

Rohini Janivara

Rohini Janivara, Wenlong Carl Chen, Ujani Hazra, Timothy Rebbeck, Joseph Lachance
Dana Farber Cancer Institute (USA), Georgia Institute of Technology (USA), Harvard T.H. Chan School of Public Health (USA), University of the Witwatersrand (South Africa)

Men of African descent face disproportionately high rates of prostate cancer (CaP) incidence and mortality, yet research into its evolutionary underpinnings in African populations remains limited. Additionally, CaP risk predictors built using inherited variants often exhibit poor performance across populations due to the inherent genetic differences shaped by human evolutionary history. Here, we explore the evolutionary foundations of CaP associations within sub-Saharan African populations. Leveraging genomic data from 3,963 CaP cases and 3,509 controls across Ghana, Nigeria, Senegal, South Africa, and Uganda, we unravel ancestry-specific genetic architectures and fine-map disease associations. Our analysis reveals fifteen independent associations, including four novel ones, at 8q24.21, 6q22.1, and 11q13.3, all reaching genome-wide significance. Strikingly, many lead disease associations are effectively monomorphic outside of Africa. Furthermore, we ascertain that the genetic architecture of CaP varies significantly across Africa, with effect size differences playing a more prominent role than allele frequency disparities. Integrated haplotype score statistics indicate no discernible selection acting on CaP within African populations. Comparing population branch statistics of CaP variants between Europe and Africa, we find that heterogeneity in CaP’s genetic architecture predominantly arises from genetic drift rather than natural selection. Importantly, allele age estimates highlight the contributions of both African polymorphisms that are due to recent mutations and older variants that predate the out-of-Africa migration. Collectively, our findings underscore how evolutionary history has contributed to differences in the genetic architectures and hereditary disease burdens of a complex trait across human populations.
Geographic and Environmental Impact on Genetic Variation in Khoe-San Descendant Communities in South Africa

Stacy L Edington

Stacy L Edington, Dana R Al-Hindi, Neus Font-Porterias, Justin W Myrick, Caitlin Uren, Paul J Norman, Marlo Möller, Brenna M Henn, Austin W Reynolds

Baylor University (USA), Stellenbosch University (South Africa), University of California Davis (USA), University of Colorado School of Medicine (USA)

While recent studies have focused on reconstructing the population history of Bantu-speaking populations of southern Africa, very little work has been done on the recent population history of Khoe-San communities. In particular, there has been limited research on the effects of European colonization and the Dutch-East India slave trade on the population structure of Khoe-San and Khoe-san descendant (K-SD) communities. We have sampled over 3,000 Khoe-San and K-SD individuals from South Africa, including 150 newly genotyped for this study to better understand how geographic and environmental factors may be impacting genetic variation in these communities across South Africa. We first used MAPS to infer effective migration rates across space using IBD tract information. Our MAPS results show relatively more migration along the southern coastal plain of South Africa and within the arid interior region, with barriers to migration between these regions. These observed decreases in effective migration rates are consistent with the location of the Cederberg and Roggeveld mountains that divide the coastal plain from the interior of the country. We next applied SPRUCE, a machine learning approach to integrate genetic and environmental data, to determine which factors may be influencing the migration patterns in our study region. Our results show that for the past 30 generations, landscape steepness and precipitation variation are the most important factors explaining effective migration patterns. Together, the environmental variables included in our model explain ~33% of the variation in effective migration in our dataset.
S11 - Spatial population genetics: where are we now?
A space-time model for jointly inferring gene flow and natural selection
Alessandro Lopez-Hernandez

Alessandro Lopez-Hernandez, Vicente Diego Ortega-Del Vecchio
Laboratorio Internacional de Investigación sobre el Genoma Humano, N/A (Mexico), Universidad Nacional Autónoma de México (Mexico)

The computational methods developed to infer past demographic events and the impact of natural selection usually employ statistical models that do not use ancient DNA data and do not take spatial information into account. The use of ancient DNA data is particularly important to infer past demographic events that could not be observed from the analysis of present-day genomic information. On the other hand, the use of spatial information is important to perform inferences of local adaptation events and to do a geographical characterization of migration events. Here we present a new likelihood-based methodology employing a statistical model that uses temporal genomic information and spatial data to infer past demographic events and the impact of natural selection acting on new alleles. Our approach uses partial differential equations to model changes in allele frequencies considering parameters that model gene flow, genetic drift and the impact of natural selection. First, we perform simulations to show that our approach can be used to infer the amount of gene flow across a geographical landscape and to infer the selective coefficient of one advantageous allele or a set of deleterious alleles. Then, we apply our method to genomic data from the “Allen Ancient DNA Resource” of 3672 European individuals from different temporalities that span the present up to 10,000 years ago. We show that our method can infer the amount of gene flow and the impact of natural selection acting on putative deleterious and advantageous alleles from European individuals.
Spatial drivers of disease progression in bacteria and viruses

Alison Feder

Presented by self
University of Washington (USA)

The intra-host pathogen evolution that drives disease progression does not happen in well-mixed test tubes, but rather in our complex, heterogeneous bodies. These environments mediate biotic interactions among pathogens and impose heterogeneous challenges for their growth. In this presentation, I’ll discuss two vignettes in which understanding the spatial context of intra-host pathogens reveals new insights into their pathogenesis. First, I will examine how the opportunistic pathogen Pseudomonas aeruginosa persists in the lungs of people with cystic fibrosis through new CFTR modulator therapies. Through the analysis of spatially-resolved, longitudinal genomic data, we identify that certain lung regions serve as sanctuary sites in which bacteria persist, and these sites can enable bacterial recolonization via migration in other lung regions. Second, I will present data from a rare form of measles in which it enters the central nervous system and adapts to its new environment. Spatially-resolved RNA sequencing reveals the presence of two major viral genotypes that appear to co-transmit and maintain balanced allele frequencies throughout the brain. Multiple recurrent mutations and back mutations related to viral manipulation of host cell fusogenicity suggest that frequency-dependent selection may be acting on viral cell-to-cell transmission. These vignettes illustrate the evolutionary and medical insights that become available when examining pathogen genomic data through a spatial lens.
In nature, organisms live on continuous, heterogeneous landscapes that fundamentally shape how they interact and reproduce over both evolutionary and ecological timescales. Including these important spatial complications in population genetic methods and models will result in a fuller understanding of how genetic variation originates and is maintained. For instance, heterogeneous population densities and source-sink dynamics shape genetic variation over space, but there are few tools currently available that can deal with these ubiquitous complexities. In this presentation I will discuss new supervised machine learning tools that we have introduced (disperseNN, disperseNN2, and mapNN) that use geo-referenced genotype data in combination with deep neural networks to provide estimates of dispersal and density over a landscape. With relatively small samples of SNP data we are able to both infer heterogenous maps of dispersal and density and to do so more accurately than competing methods. The methods are flexible because they rely on simulations rather than mathematical models, and so can incorporate complexities for which no analytical approximations are available. I will discuss applications to empirical data as well as future prospects for the next generation of spatial population genetic tools.
Larval dispersal and connectivity in the marine environment are important for biodiversity, ecosystem function and population well-being. Invertebrate marine species around Aotearoa New Zealand vary tremendously in factors that influence larval dispersal, including the amount of time larvae spend in the ocean and the time of year in which larvae are released. We have previously shown both of these to significantly impact larval dispersal. Here we examine the output of dispersal simulations depending on these factors with a genetic dataset of multiple species. We expect that genetic patterns should reflect those of larval dispersal if larval dispersal drives gene flow between populations. By comparing the results of larval dispersal simulations from the Lagrangian tracking software “OpenDrift” with a nationwide circulation model with population genetic metrics, we assessed the power of dispersal driven by ocean currents in predicting genetic patterns. Across a broad range of species, we found limited linear correlation between dispersal and genetic metrics, which suggests that processes outside of oceanic currents alone impact genetic patterns. Using a variety of dispersal metrics, we estimated how many generations would be required to connect distant locations by dispersal only and which locations are sinks and sources. Locations on the east coast of the South Island were sources, while those in central locations were sinks. These findings highlight the use of a multidisciplinary approach in disentangling what factors are involved in causing contemporary genetic patterns and the general understanding of marine connectivity and dispersal.
In space no one can hear you sweep: a novel signal of selective sweeps in continuous geographic space

Clara T Rehmann

Clara T Rehmann, Peter L Ralph, Andrew D Kern
University of Oregon (USA)

As organisms adapt to environmental changes, natural selection modifies the frequency of non-neutral alleles. For beneficial mutations, the outcome of this process may be a selective sweep, in which the allele rapidly increases in frequency and perhaps reaches fixation within a population. Selective sweeps have well-studied effects on patterns of local genetic variation in panmictic populations, but much less is known about the dynamics of sweeps in continuous space. In particular, because limited movement across a landscape leads to unique patterns of population structure, spatial dynamics may influence the trajectory of mutations under positive selection. Here, we use forward-in-time, individual-based simulations in continuous space to track a beneficial mutation as it sweeps through a population. We describe the interconnected effects of selection, allele frequency, allele age, and continuous spatial structure on the dynamics of a selective sweep. With realistic simulation, we show that selection changes the joint distribution of allele frequency and geographic range and demonstrate that this signal can be used to identify selective sweeps. We then leverage this signal to identify in-progress selective sweeps within the malaria vector Anopheles gambiae, a population under strong selection pressure from vector control measures. By considering space, we identify multiple previously-undescribed variants with potential phenotypic consequences, including mutations impacting known IR-associated genes and altering protein structure and properties. Our results demonstrate a novel signal for detecting selection in spatial population genetic data that may have implications for genomic surveillance and understanding geographic patterns of genetic variation.
Evolution of phenotypic plasticity owing to migration
Davorka Gulisija

Davorka Gulisija
University of New Mexico (USA)

Phenotypic plasticity is one of the pivotal mechanisms that allow populations to persist in the face of environmental change. For plasticity to evolve and persist, populations need to continuously experience selection either due to changing environments or an influx of maladaptive phenotypes. Continuously changing environments are limited to specific eco-zones and organisms, but not universal, making immigration a likely contributor to the wide prevalence of phenotypic plasticity in nature. However, the conditions under which plasticity arises due to migration are not clear, particularly if plasticity is costly. Here, we propose that plasticity will arise in populations across a broad range of conditions, including costs of plasticity, if migration rates are sufficiently large. We support our argument with analytic derivation and extensive simulations assuming a two-locus two-habitat population genetics model. We examine the properties of the evolution of the plasticity modifier allele across adapted Wright-Fisher populations assuming a range of migration rates, plasticity effects with and without the cost of plasticity, recombination rates, mutation rates, and genetic architectures. We find the evolution of plasticity due to migration is robust to the genetic architecture of the plasticity effect and recombination rates, and occurs across a wide range of migration scenarios.
Sexual organisms exist in geographic space, a fact that restricts the available sexual partners with whom an individual can reproduce. Over generations, genes travel over the landscape, creating meaningful patterns of genetic diversity across space. Inferring rates of this spatial gene flow has applications across many fields; for example, gene flow rates can explain how ecological features contribute to genetic differentiation, or reveal cryptic subpopulation structure. Here we implement and extend Lundgren and Ralph\'s 2019 gene flow rate inference method, and demonstrate its use in an empirical scenario. The method uses a flexible discrete-deme model of subpopulations, and uses a coalescence-time framework to allow for anisotropic gene flow surfaces. We implement Bayesian inference of the gene flow rates in the open-source computing platform Stan, which is well-suited to high-dimensional inference tasks, and allows for greater reproducibility of our analyses. We apply our method to Anopheline data from the Ag1000G project.
The number of individuals in a population, or census size, is one of the most important pieces of information for species that we want to conserve, control, or harvest, but reliably estimating it can be very difficult. Widely used genetic capture-recapture methods require capturing the same individual multiple times, which can be impossible for elusive species or species that are sampled lethally. Close kin mark-recapture (CKMR) is a promising new method that does not require recaptures of the same individual and can make use of a single temporal sample. It uses genetic data to identify related pairs of individuals, and then estimates census size from these pairs. However, a major shortcoming of current CKMR methods is that they do not account for spatial population structure or spatial heterogeneity in sampling, patterns that strongly affect estimates. We developed a spatially explicit CKMR method using a convolutional neural network trained on simulated populations. The method uses as input maps of parent-offspring pairs on the landscape and maps of sampling intensity, and is able to accurately estimate census size, even when dispersal is limited and some areas of the landscape are more intensely sampled than others. Parent-offspring pairs can be identified from as little as hundreds of SNPs. We are extending our method to use more distantly related individuals, and to estimate a map of population density rather than a single census size. These novel methods have the potential to inform and evaluate conservation efforts for many species.
Using biological invasions to understand rapid adaptation to new environments: a genomic reconstruction of the house sparrow global spread

Jack Harper

Jack Harper, Mark Ravinet, Britt Heidinger, Simon Griffith, Lynn Martin
Macquarie University Sydney (Australia), North Dakota State University (USA), University of Nottingham (United Kingdom), University of Oslo (Norway), University of South Florida (USA)

The well-known house sparrow (Passer domesticus) is a small bird renowned for its charismatic nature and close association to humans. While the native range of the house sparrow is Eurasia, due to multiple intentional and accidental introductions from the mid-19th century onwards, the species is also invasive in Australasia, Southern Africa, and the Americas. In modern North America, the house sparrow now exists from Southern Panama all the way to the Northwest territories of Canada, attesting to the species’ ability to survive in a diverse range of environments. This research aims to understand evolutionary success in biological invasion by reconstructing the parameters of the house sparrow invasive spread. Using whole genome resequencing of 500 birds, we demonstrate substantial population structure present within the introduced range, suggesting high levels of divergence between different introduction sources. We also studied parallel latitudinal clines within the United States and Australia to quantify the extent of rapid adaptation to novel environments from bottlenecked founder populations, suggesting that certain genome regions and traits may be more beneficial for evolutionary success. Invasive species are an excellent model for understanding complex evolutionary processes, and this research will inform future work on understanding the genomics of successful adaptation to new environments.
Reconstructing the locations of genetic ancestors for a recombining sequence
James Kitchens

James Kitchens, Puneeth Deraje, Matthew Osmond, Graham Coop
Davis (USA), University of California, University of Toronto (Canada)

Present-day spatial patterns of genetic relatedness among organisms reflect the past movements of their ancestors. The ability to untangle these histories has improved dramatically thanks to advances in ancestral recombination graph (ARG) methods. By extending spatial methods previously applied to trees, we generalize a model of Brownian dispersal to work on ARGs, thereby accounting for correlations along a chromosome when computing the likelihood-based estimates of dispersal rate and locations of genetic ancestors. Further, we develop an efficient algorithm that allows for the application of the method to complex ARGs, scalable to thousands of samples. We evaluate our method’s ability to reconstruct spatial histories using simulations. Surprisingly, despite using the fullest information available in the data, we find that our dispersal estimates are biased, uncovering friction between the histories of recombinant lineages and Brownian dispersal models. We identify potential resolutions to this problem based on relaxing the constraints that ARGs place on the movement of lineages and show that ARG-based spatial inference can be used to effectively track the geographic history of admixed individuals. Approaches like this will be key to understanding the interplay of migration, recombination, drift, and adaptation in geographically spread populations.
Exploring the behavior of two-locus statistics in continuous space

Lloyd Kirk

Lloyd Kirk, Simon Gravel, Peter L Ralph, Jerome Kelleher, Aaron P Ragsdale
Big Data Institute, Department of Human Genetics, Department of Integrative Biology, Institute of Ecology & Evolution, Li Ka Shing Centre for Health Information and Discovery, McGill University (Canada), University of Oregon (USA), University of Oxford (United Kingdom), University of Wisconsin-Madison (USA)

Succinct tree sequences have transformed the way we store and analyze genomic variation in both simulated and empirical datasets. Tskit is the widely-used reference implementation of tree sequences and the leading toolset for analyzing tree sequence data. We have developed a generalized framework for computing two-locus statistics in tskit, implementing a broad range of two-locus statistics. This work enables the flexible computation of LD matrices (as well as LD scores and LD decay) from observed mutations and expectations for those statistics from the underlying genealogical structure. Our framework extends to any single or multi-population statistic between sites with arbitrary numbers of alleles. We use the robust ecosystem of simulation tools that integrate with tskit to study how two-locus statistics behave under realistic demographic scenarios. Most population genetic models of geographic dispersal are based on assumptions of discrete, homogeneous populations. However, individuals in fact live across continuous geography, and recent advances enable simulation of population genetic processes in continuous space. We use these new methods, along with our framework for computing two-locus statistics to investigate the impacts of neighborhood size on local LD structure. We aim to uncover the effects of mate choice and dispersal that occur at fine scales and how they impact the dynamics of two-locus statistics in the presence of continuous geographic structure. In addition, we investigate the effect of spatial sampling of individuals on the estimation of different measures of LD and estimate the impact of sampling strategy on downstream inferences.
Spatial ancestry reconstruction using parsimony on tree sequences
Michael Grundler

Michael Grundler, Jonathan Terhorst, Gideon Bradburd
University of Michigan (USA)

Many statistical methods for quantifying ancestry average over the ages of shared ancestors in a sample, effectively "flattening" the temporal component of the genealogy that connects all individuals within a species. This flattening has the effect of painting a static notion of ancestry, rather than one that changes as it proceeds backwards in time. Here, we describe a method that capitalizes on the rich genealogical information encoded in genomic tree sequences to infer the geographic locations of the shared ancestors of a sample of sequenced individuals. After validating our method with simulations, we illustrate its application to an empirical dataset and show how a spatiotemporal summary of ancestors offers an intuitive and informative path toward understanding ancestry, demography, and major population movements through time.
Determining concordance and drivers of spatial population structure through genogeographic clustering
Shane Lavery

Vanessa Arranz, Rachel Fewster, Shane Lavery
University of Auckland (New Zealand), University of Barcelona (Spain)

In recent years there have been great strides made in the sophistication and detail of both the collection and analysis of spatial population genetic data within single species. However, the methods of comparing patterns of population structure across species have been much slower to develop. This is problematic, as it is becoming clear that an understanding of the patterns of concordance in population structure across multiple species is necessary to better comprehend community patterns, and thus provide insights into the common drivers of those patterns. Here, a novel approach, "genogeographic clustering", was used to determine concordance of spatial population structure in a community of 21 species of coastal marine organism in New Zealand waters. The approach involved: 1) use of genetic divergences from previously published studies to quantitatively describe patterns of population structure within each species as a fitted spline curve, 2) quantitatively clustering species by their similarity in geographic pattern using a dendrogram of curve similarities, and 3) then testing whether known life-history and ecological traits are associated with the species sharing similar genetic patterns, using distance-based regression. This approach shows promise in helping us determine both the concordance in spatial patterns among species in a community, and the potential processes driving those patterns.
Population structure in genetics arises from an evolutionary history of local genetic drift that is offset by gene flow. On one extreme, gene flow can be continuous through time, as might be expected for gene flow between neighboring populations, or it might involve instantaneous bursts of rare long-range events across the habitat. Disentangling the effects of both of these modes of gene flow in a single modeling framework is a largely open challenge. Here, we present a fast and accurate method called FEEMSmix (an extension to FEEMS from Marcus, Ha et al 2021) that infers long-range gene flow events using single nucleotide variant data in a background of spatially heterogeneous diffusive migration. The method uses observed deviations from a baseline spatially heterogeneous isolation-by-distance model to detect plausible pairs of populations that were involved in long-range gene flow events. Then in a coalescent-based framework, the method infers the magnitude, timing, and direction of these events assuming an instantaneous pulse in the last few generations. The method produces geographic maps annotated with long-range edges corresponding to gene flow events and maps of uncertainty in the geographic location of a particular source. We validate the method using simulations. When we apply this framework to existing data sets from North American Canis lupus (gray wolves), Homo sapiens (humans) from Afro-Eurasia, and Phylloscopus trochiloides (greenish warblers) in Central Asia, we find the method paints a fuller picture of the genetic structure and evolutionary history of each sample.
Phylogeographic analysis of the invasive Asian hornet, Vespa velutina, to reconstruct its colonization of Europe from Asia

Xuhua Xia

Xia C Xia
University of Ottawa (Canada)

The Asian hornet, Vespa velutina, is an invasive species that has established successfully in multiple European countries and incurred financial costs by damaging local apiary. Did the colonization occur only once or multiple times from Asia to Europe? How did the population expand and differentiate after the initial colonization? Can ecological models successfully predict its habitat suitability in Europe and its biogeographic expansion? Is it possible to estimate the genetic variation of the founding queens? How can a geophylogeny help us visualize genetic variation over time and space and identify geographic barriers to gene flow? How would human activities facilitate the dispersal of the hornets? I present preliminary results to address these questions.
S12 - Exploring the Evolutionary Effects of Admixture.
Complex neutral processes drive the evolutionary fate of Neanderthal alleles in 30,780 admixed genomes with African-like and European-like ancestry

Aaron Pfennig

Aaron Pfennig, Joseph Lachance
Georgia Institute of Technology (USA)

Following introgression, Neanderthal DNA was initially purged from modern human genomes but has remained constant at ~2% in Europe during the last 45,000 years. Here, we leverage whole-genome sequences of 30,780 unrelated, recently admixed individuals with African-like and European-like ancestry in All of Us to test the evolutionary fate of Neanderthal alleles. In these admixed genomes, Neanderthal variants were recently introduced into an African genetic background, leading to a reevaluation of their present-day fitness. Surprisingly, we observe more Neanderthal ancestry in such admixed genomes than expected under neutrality and a simple demographic model. However, this enrichment is driven by segments also observed in African genomes, due in part to incomplete lineage sorting. When focusing on introgressed segments uniquely contributed by European-like ancestors, the observed enrichment disappears. Nevertheless, we identify 405 and 359 genomic regions significantly enriched or depleted for Neanderthal ancestry, respectively. Preliminary functional and population genetic analyses of these regions reveal enrichments for GWAS hits associated with known Neanderthal phenotypes and pleiotropic eQTLs but do not support strong recent selection. Simulations under a plausible demographic model suggest that complex neutral processes can explain the initially observed enrichment and outlier regions, underlining the importance of appropriate demographic models as in population genetics. We also identified four emerging introgression deserts, whose analyses may provide insights into the evolutionary dynamics leading to previously reported archaic deserts (i.e., simple negative selection vs. hybrid incompatibilities). Altogether, Neanderthal DNA appears to behave largely evolutionary neutral in extant human genomes.
Nonhuman Primate Models for Hominin Introgression and Genetic Admixture
Alaina L Brenner

Alaina L Brenner, Timothy D Weaver, Brenna M Henn
Davis (USA), University of California

Hybridization and genetic admixture occur across many ecological taxa past and present, including isolated lineages of late Pleistocene hominins. However, little is known about the dynamics of these genetic exchanges in the first few generations of admixture. Therefore, this project investigates the mechanisms of hybridization and admixture in the genomes of contemporary nonhuman primates as a model to understand hominin introgression events in human evolution. Here, we use 8 generations of low coverage whole genome sequencing (WGS) to assess local and global ancestry of an admixed colony of 140 macaques once housed at the California National Primate Research Center (CNPRC). Indian and Chinese rhesus macaques are especially relevant due to their similar divergence time in generations to humans and Neandertals. These macaque subspecies have accumulated substantial genetic variation since the divergence and subsequent isolation of the two lineages ~160,000 years ago. During our first step in this ongoing project, we modified the GATK variant calling pipeline for macaques and identified 52,207,528 non reference alleles on average per individual. After assessing local and global ancestry in the WGS, we observed large discrepancies between the WGS admixture estimates and the previous estimates based on the CNPRC pedigree and limited microsatellite data. Specifically, ~40% of the sample have ancestry from both subspecies which is lower than expected under random mating. We can then use simulations to test whether this pattern is consistent with purifying selection against hybrids. This model will be used to illuminate human and Neandertal admixture within the first 10 generations.
Disentangling the effects and history of human admixture from the Americas to Oceania

Alexander Ioannidis

Presented by self
Santa Cruz (USA), University of California

The underrepresentation of diverse populations in genetic research hinders our understanding of the genetic basis of disease. This is particularly true for Oceanian populations, who account for less than 0.02% of samples in the NHGRI-EBI GWAS catalog. I will discuss an analysis of individuals from Polynesia, Micronesia, and Melanesia using a number of selection statistics together with local ancestry analyses. Our analyses have identified several genetic risk loci specifically associated with these populations. We also show that 20 HLA genotypes vary in frequency among Melanesian, Micronesian, and Polynesian populations, providing new evidence of selection among specific Pacific Islander communities. Comparison between these results and populations from another underrepresented region of the world, namely the Americas, will also be highlighted together with methods for disentangling admixture histories related to these signals.
Alternative genomic and biochemical pathways to red carotenoid pigments in birds
Alexander Kirschel

Alexander N. G. Kirschel, Sifiso Lukhele, Matteo Sebastianelli, Sophia C. Hayes, Loïs Rancilhac, Alan Brelsford, Bridgett M. vonHoldt
Princeton University (USA), University of California Riverside (USA), University of Cyprus (Cyprus), Uppsala University (Sweden)

Red pigments in animals are typically shaped by sequestration of carotenoids acquired from the diet and enzymes that catalyze the conversion of dietary yellow carotenoids to red ketocarotenoids. Several candidate genes have been found to function in this pathway, including CYP2J19 and BCO2, but a large-scale study in nature of the relative effect sizes of candidate genes is lacking. Furthermore, the relative effect on phenotypes of different dietary carotenoids present in the diet is little known. To investigate these questions, using 202 whole genomes, we conducted a genome-wide association analysis to identify the genes underpinning variation between yellow and red forecrown feathers across a hybrid zone of Pogoniulus tinkerbirds. We identified a locus of large effect spanning several genes including CYP2J19, CYP2J40, HOOK1, FGGY and NF1A that explained 80% of the variation in forecrown color, with heterozygotes at this locus all converting carotenoids and exhibiting red hues. We then compared the extent of gene expression in liver tissue across the contact zone to confirm the relative effects of genes within this locus. Moreover, we investigated the pathways resulting in red forecrown feathers in disjunct populations of tinkerbirds across Africa. We determined the extent to which red forecrown color is ancestral, has evolved in parallel, or might be explained by another pathway involving a different dietary carotenoid distribution identified using Raman spectroscopy and high performance liquid chromatography. Our study reveals how red pigmentation in birds may result from alternative genomic and biochemical pathways of carotenoid metabolism.
Ancient DNA screening at Denisova Cave to explore Bos/Bison admixture
Alexandre Gilardet

Alexandre Gilardet, Katerina Douka, Peter Heintzman, Love Dalén
Centre for Palaeogenetics (Sweden), Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna (Austria)

Nuclear genomes can usually provide a more nuanced story than mitogenomes. Previously, mitochondrial sequences have shown that the yak (Bos grunniens) clusters with the American bison (Bison bison) while the European bison (Bison bonasus) clusters with cattle (Bos taurus) (Marsolier-Kergoat et al. 2015, Zeyland et al. 2012). However, nuclear data showed the American and European bison to be closer to each other than other relatives (Verkaar et al., 2004). Additional and more ancient nuclear data from both yak and bison can potentially resolve these discordant phylogenies. Denisova Cave lies in Siberia and both Bos and Bison sp. remains have been morphologically identified there. This site is thus a good candidate to resolve the common evolutionary history of Bos/Bison. The available bone fragments collection spans several glacial and interglacial periods up to the Middle Pleistocene and could therefore hold answers about the possible climate-mediated admixture history of Bos/Bison. We are now screening the collection for those samples with good endogenous DNA preservation. Already, we are able to retrieve mitogenomes at a high enough coverage for species assignment. The later generation of high coverage ancient nuclear genomes will help test the hypothesis of ancient hybridization and gene flow between the extinct steppe bison (Bison priscus) and wild yak (Bos mutus).
Bayesian Inference of Admixture Graphs on Native American and Arctic Populations
Andrew Vaughn

Andrew H Vaughn, Svend V. Nielsen, Kalle Leppälä, Michael J. Landis, Thomas Mailund, Rasmus Nielsen
Aarhus University (Denmark), UC Berkeley (USA), University of Copenhagen (Denmark), University of Oulu (Finland), Washington University in St. Louis (USA)

Admixture graphs are a widely used tool in population genetics to infer historical relationships from samples of genetic data. Admixture graphs provide a concise description of the historical demographic relationships between genetic samples of populations as a directed acyclic graph representing population splits and mergers. Inferring graph topologies, however, involves a combinatorial search, and since the space of graphs grows super-exponentially in the number of populations and the number of admixture events, an exhaustive search is typically not possible. Instead, existing methods that search for well-fitting graphs utilize greedy algorithms, which are liable to get stuck in local maxima. Furthermore, these methods are unable to report uncertainty in these estimates as they only yield point estimates of graphs. We here improve on these approaches by developing a novel MCMC sampling method, AdmixtureBayes, that can sample from the posterior distribution of admixture graphs. This enables an effective search of the entire state space as well as the ability to report a level of confidence in the sampled graphs. We apply AdmixtureBayes to a set of Native American and Arctic genomes to reconstruct the demographic history of these populations and report posterior probabilities of specific admixture events. In particular, we explore the sources of gene flow into the modern Athabascan and Inuit populations from distinct ancient Siberian and Native American populations.
Pervasive gene flow despite strong and varied reproductive barriers in swordtails

Daniel Powell

Stepfanie Aquillon, Sophia Haase Cox, Quinn Langdon, Theresa Gunn, JJ Baczenas, Shreya Banerjee, Alex Donny, Ben Moran, Carla Gutiérrez-Rodríguez, Oscar Ríos-Cárdenas, Molly Morris, Daniel Powell, Molly Schumer

A.C. (Mexico), A.C. (USA), Centro de Investigaciones Científicas de las Huastecas “Aguazarca”, Centro de Investigaciones Científicas de las Huastecas “Aguazarca” (Mexico), Gladstone Institute of Virology (USA), Howard Hughes Medical Institute (USA), Instituto de Ecología A.C. (Mexico), Ohio University (USA), Stanford University (USA), Stanford university (USA), University of California Davis (USA), University of California Los Angeles (USA)

The formation of new species occurs through the evolution of reproductive barriers. However, recent research has demonstrated that hybridization has been pervasive across the tree of life despite the presence of strong barriers. Swordtail fish (genus Xiphophorus) are an emerging model system for studying the interface between these barriers and hybridization. We document overlapping mechanisms that act as barriers between closely related species, X. birchmanni and X. cortezi, by combining genomic sequencing from natural hybrid populations, artificial crosses, behavioral assays, sperm performance, and developmental studies. We show that strong assortative mating plays a key role in maintaining subpopulations with distinct ancestry in natural hybrid populations. Lab experiments demonstrate that artificial F1 crosses experience dysfunction: crosses with X. birchmanni females were largely inviable and crosses with X. cortezi females had a skewed sex ratio. Using F2 hybrids we identify several genomic regions that strongly impact viability. Strikingly, two of these regions were previously identified as genetic incompatibilities in hybrids between X. birchmanni and its sister species X. malinche. Our results demonstrate that ancient hybridization has played a role in the origin of this genetic incompatibility. Moreover, the incompatibility causes extraordinarily similar survival effects in X. cortezi ? X. birchmanni hybrids as in X. malinche ? X. birchmanni hybrids. Our findings highlight complex evolutionary outcomes of hybridization in the face of varied reproductive barriers and historical hybridization between species.
Differentiating mechanism from outcome for ancestry-assortative mating in admixed populations

Dashiell J Massey

Dashiell J Massey, Zachary A Szpiech, Amy Goldberg
Duke University (USA), Pennsylvania State University (USA)

Non-random mating is an important source of genetic structure in natural populations. Empirical studies across multiple species have reported positive correlations in trait values between mates, potentially confounding genome-wide association studies, selection scans, and demographic inference. Within several recently admixed human populations, empirical genetic studies have reported a correlation in global ancestry proportion between spouses, referred to as ancestry-assortative mating. Here, we combine genetic data analysis and benchmarking of common parental-ancestry inference methods with forward genomic simulations to link observed correlations in ancestry between human spouses to the underlying mechanistic mate-choice process. Indeed, much of the theoretical work on assortative mating derives from single-trait models in non-human species, whereas ancestry proportion is a genome-level trait and human mate-choice is likely shaped by numerous sociocultural, as well as biological, factors. We consider the impacts of potential mate-choice models, parental-ancestry inference methods, and analysis strategies on the identification of evidence for ancestry-assortative mating in human populations. For instance, we find that multiple mate-choice models can produce identical correlations in ancestry proportion between spouses; however, we highlight alternatives to the common parental ancestry correlation plots in which these potential models may be well distinguished. With this work, we aim to sound a note of caution about interpreting correlations in empirical data as evidence for a particular model of human mating practices – as well as to offer suggestions toward development of new best practices for analysis of human assortative mating.
Evolutionary insights from admixed genomes

Emilia Huerta Sanchez

Presented by self
Brown University (USA)

Despite having a small number of Denisovan fossil remains, genomic studies have demonstrated that 1) Denisovan and modern humans interbred, 2) Denisovans exhibit more population structure than Neanderthals, and 3) Denisovans may have inhabited a wide range of geographic regions, potentially spanning from Siberia to Island South East Asia. Denisovans, however, remain a more mysterious population than Neanderthals. In this talk, I will present research that leverages admixed populations from the Americas to help characterize the impact of Denisovan ancestry, and I will discuss new methods that will help gain insight into the history of Denisovan populations.
Selection on hybrid incompatibility in baboons demonstrates Haldane’s rule on time scales similar to human-Neanderthal divergence

Kasper Munch

Erik Sørensen, Kasper Munch
Aarhus University (Denmark), BiRC

With divergences between 700,000 and 1.3 million years, the six baboon species mirror the relationship between members of the Homo genus. Baboons thus present an opportunity to study the emerging reproductive barriers similar to those present when non-African modern humans intermixed with Neanderthals. Exploiting the diversity across 225 full genomes of the six baboon species, we use local ancestry inference to identify admixture tracts along the individual genomes of yellow and olive baboons. Admixture proportions in Tanzania decline rapidly along a population transect away from the hybridization zone, from 21% in the Tarangire population closest to the hybridization zone to 4% in the Serengeti population furthest away. The population admixture proportion in 100kb windows correlates with recombination rate, suggesting that the admixture gradient is maintained by negative selection. Following Haldane’s rule for reproductive incompatibility, this correlation is strongest on the X chromosome. The population admixture proportion in 100kb windows further correlates with the diversity in unadmixed populations, a proxy for the joint effects of background selection and reduced incomplete lineage sorting. The strength of this correlation increases with distance from the hybridization zone as the selection has had more time to act on introgressed segments. In Ethiopia, autosomal admixture between Olive and Hamadryas baboons (only 700,000 years diverged) show similar evidence of selection, but here the correlation to recombination rate is an order of magnitude stronger for X chromosome admixture. Our results thus demonstrate hybrid incompatibility following Haldane’s rule in primates diverged no more than humans and Neanderthals.
Ecological genomic structure of a complex of Myotis bats in the Baja California peninsula and western Mexico
Laura Alejandra Najera Cortazar

Laura A Najera Cortazar, Sergio Ticul Alvarez Castaneda, Ian Carr, Kathryn Maher, Helen Hipperson, Winifred Frick, Simon J Goodman
BIOPOLIS Association - Research Centre in Biodiversity and Genetic Resources (BIOPOLIS-CIBIO) (Portugal), Centro de Investigaciones Biologicas del Noroeste S. C. (CIBNOR) (Mexico), Santa Cruz (USA), University of California, University of Leeds (United Kingdom), University of Sheffield (United Kingdom)

Cryptic species diversity within bats of the genus Myotis creates challenges in understanding their evolution and ecological interactions. Using mitochondrial and genome-wide SNP data, we investigated the population structure and species boundaries of a complex of Myotis bats in the Baja California peninsula, and two continental sites in Mexico. Mitochondrial haplotypes showed that individuals phenotypically identified as M. pensularis, mostly form a distinct haplogroup closely related to M. velifer. However, phylogenetic analysis of mitochondrial sequences returned paraphyletic clades for the species. ddRAD SNP data results also did not reliably partition individuals, and identified evidence of introgression between the two in bats sampled in mid-peninsula and a continental population. Our mitochondrial phylogenetic analysis identified two other potential cryptic Myotis lineages that will need further investigation. Bayesian structure and cluster analyses showed high levels of genetic differentiation and population structure within M. californicus populations, detecting two different clusters, but with evidence for both local and long-distance dispersal. In contrast, admixture was detected among M. yumanensis individuals at several sites, reflecting gene flow with the other three Myotis species in the study. There was weak genetic structure among M. yumanensis individuals, with conflicts between nuclear and mitochondrial population assignments, suggesting female philopatry and long-distance male mediated dispersal. Our study represents an important step forward for the understanding of species boundaries in Myotis bats from the American continent that can inform development of taxonomic assessments and conservation strategies for North American bats.
Divergent ancestry in the TSHR gene is associated with changes in seasonal reproduction during dog domestication
Lauren Hennelly

Lauren M Hennelly, Katia Bougiouri, Laurent Frantz, Cindy Santander, Xin Sun, Nuno FG Martins, Frank Hailer, Renzo FF Balboa, Nucharin Songsasen, Thomas Gilbert, Klaus P Koepfli, Mikkel HS Sinding, Shyam Gopalakrishnan
Cardiff University (United Kingdom), Ludwig Maximilian University Munich (Germany), Smithsonian's National Zoo and Conservation Biology Institute (USA), University of Copenhagen (Denmark)

Domestication has altered many traits in dogs compared to their wild ancestor, the gray wolf. Among these changes is in reproduction, most notably, the loss of reproductive seasonality. To better understand the genetic basis of reproductive changes, we investigated genomic differences in key reproduction-related genes between dogs and wolves. We find one of the most extreme genome-wide signals of differentiation between dogs and wolves is at the thyroid stimulating hormone receptor (TSHR) gene, a key gene related to seasonal reproduction and thyroid hormone regulation. We pinpoint this differentiation at a ~18kb region containing highly divergent ancestry and a ~3kb LINE1 insertion found in dogs. Using ~2,000 canid genomes, we find the divergent haplotype is nearly fixed in dogs, however, it is at low frequency in Australasian dogs. While nearly absent in modern and ancient gray wolves, this divergent haplotype is most prevalent in wolves from India and Southwest Asia. Ancient canid genomes reveal it appeared early in dog domestication ~11,000 years ago and is associated with the secondary source of western dog ancestry. Our findings are consistent with recent work proposing that as dogs expanded into Southwest Asia, gene flow with local wolves acted as a second source of ancestry, which likely introduced new genetic variation for which selection to act upon. Although further work is necessary to quantify the functions of this ~18kb region, our work is the first to detect divergent ancestry at a locus that potentially plays a role in core traits related to domestication.
Different measures of reproductive isolation reflect different stages of species formation

Linda Hagberg

Linda Hagberg, Dörte Neumeister, Linde N Wieringa, José L Bella, Ricardo J Pereira
Ludwig-Maximilians-Universität (Germany), Naturkundemuseum (Germany), Universidad Autónoma (Spain)

Speciation is often thought of as complete reproductive isolation (RI), characterised by the accumulation of reproductive barriers counteracting gene flow. Despite the widespread view of a forward-in-time speciation process, the accumulation of reproductive barriers is not linear or even irreversible, as numerous studies focusing on the admixing effects of hybridisation have found. Instead, the strength of reproductive barriers can fluctuate in cycles of e.g. geographic isolation (blocking gene flow) and secondary contact (allowing gene flow). Tools used to identify RI vary in their underlying assumptions and thus in their implications. Here, I compare the implications of three different methods commonly used for inferring RI with data from two pairs of hybridising populations of the meadow grasshopper Pseudochorthippus parallelus. I focus on RI inferred using (i) divergence time and effective gene flow, (ii) hybrid zone width, and (iii) heterogeneity of introgression across independent genomic regions. I find the strength of RI to be extremely variable among the methods and suggest that the differing results may simultaneously be correct but represent windows into different spatio-temporal and genomic scales of species formation.
Interspecific hybridization events have shaped the genomes of various organisms across Eukarya. Yeast hybrids, recognized for their weak prezygotic barriers and prevalence in human-associated environments, are valuable models for studying the effects of hybridization on speciation, adaptation, and genome evolution. In this work, we isolated and sequenced over 60 hybrids of Saccharomyces cerevisiae and its sister species, S. paradoxus, from traditional agave distilleries and their natural surroundings across Mexico. These hybrids are prevalent in this environment, accounting for up to 10% of the total sequenced isolates and rising to 40% in semi-arid regions, suggesting that climatic factors may influence hybrid formation or survival. We will present comparative analyses of the nuclear and mitochondrial hybrid genomes which revealed numerous distinct genomic architectures, indicating multiple hybridization events and varying stages of genome stabilization. This study contributes to our understanding of the evolutionary dynamics of interspecific hybridizations, providing a valuable resource to study their impact on speciation and adaptation.
Detecting selection in admixed populations can be challenging, especially when the admixture is not considered in the model explicitly. It is possible to mistake selection signatures with haplotypes that exist in high frequency due to admixture instead of selection. To untangle these selection signatures from other biological processes, we suggest extending hapFLK, a test designed to detect selection in population trees using haplotypic information from multiple populations. Estimating the variance-covariance matrix of the allele frequencies is crucial for accurately estimating the distribution under neutrality in this test. In the original test, the variance-covariance matrix was obtained from the branch lengths of the Neighbor-joining population tree, which considers the evolutionary framework. To extend this test to admixed populations, we suggest three different strategies: use hapFLK as it is, based on trees; estimate the variance-covariance matrix using TreeMix; and estimate the variance-covariance matrix empirically. The last procedure could be thought of as a way of estimating branch lengths based on F-statistics. For the empirical estimation, we estimate the frequency of the ancestral alleles first based on the Neighbor-Joining tree and then the covariances of each pair of populations. Our simulations showed that the best strategy is to use the empirical variance-covariance matrix to detect selection. However, we could not achieve the same power as when using the theoretical variance-covariance matrix. Further work should be focused on the matrix estimation. Finally, as previous work has shown, detecting hard sweeps is easier than detecting soft sweeps.
Exploring recent evolution of matrotrophy in swordtails (Xiphophorus)

Nemo Valentin Robles

Nim Robles, Theresa Gunn, Cheyenne Y. Payne, Derek Ly, David Reznick
A.C. (Mexico), Centro de Investigaciones Científicas de las Huastecas “Aguazarca”, National Marine Fisheries Service, National Oceanic and Atmospheric Administration (USA), Riverside (USA), Santa Cruz (USA), Southwest Fisheries Science Center, Stanford University (USA), University of California

Xiphophorus malinche and X. birchmanni are sister species of swordtail fishes residing at various elevations in river systems in México, which have extensive naturally-occurring hybrid zones. Unlike most teleosts, Xiphophorus bear live young (viviparity). One of the evolutionary benefits of viviparity is it allows for post-fertilization maternal nutrient provisioning, known as “matrotrophy”. Matrotrophy increases offspring survival, but can result in reduction of parent survivorship, frequency of reproduction, and lifetime number of offspring. Temperature, predation, and resource availability are known to influence the evolution of matrotrophy, which vary with elevation. Recent work by Payne et. al. (in prep), found that lab crosses between male X. malinche, a high elevation species, and female X. birchmanni, a low elevation species, are largely inviable while the reciprocal cross produced viable offspring. X. malinche live at 1200 m in cold, resource-poor environments. We hypothesize that they evolved matrotrophy to help their young overcome these challenges. X. malinche fry are born much larger than other related species, while X. birchmanni fry are similar in size to other Xiphophorus, suggestive of a recent evolution of matrotrophy within X. malinche. Additionally, X. malinche mothers upregulate prolactin-related genes in ovarian tissue, a key hormone in pregnancy, during embryonic development. With this information, we plan to study matrotrophy during embryonic development in different environments to determine whether matrotrophy may be an adaptation to high elevation. X. malinche provides an exciting opportunity to study ecological and genetic drivers of matrotrophy and its role in interspecies reproductive isolation.
The co-evolution of phylogenomic signal and recombination rate during complex speciation in placental mammals

Nicole Foley

Nicole M Foley, William J Murphy
Texas A&M University (USA)

Recombination-aware phylogenomic studies have reshaped our understanding of the evolutionary histories of select clades across the Tree of Life. However, the widespread application of these transformative methods is limited by the availability of recombination maps that correspond to chromosome-level genome assemblies. Here, we characterized the progressive genome-wide re-patterning of phylogenomic signal that occurs with discordance due to introgression or incomplete lineage sorting across an old vertebrate lineage, placental mammals. We selected 10 mammal clades for which there was 1) a chromosome-level genome, 2) phylogenomic data, and 3) population-genomic data, representing mammalian superordinal clades. Population genomic data was used to generate recombination maps using a machine learning approach, ReLERNN, for 20 mammalian species. Discordant chromosomal patterns of phylogenomic signal generated from genome-wide samples of ML trees could be assigned to four groups displaying progressive discordance patterns. We repeatedly observed an enrichment of the species tree on the X with introgression. With more pervasive introgression, this pattern is highly restricted to a multi-megabase low recombining desert at the center of the X chromosome conserved across multiple mammalian families and orders. Using genome-wide PhyloP scores from a 240 mammalian species alignment we show this region is characterized by mild selective constraint and accelerated evolution. We hypothesize the enrichment for species-specific ampliconic genes and genetic components of the X chromosome inactivation center may act as recurrent components of a supergene in placental mammals.
Genomic stability over 65 million years promotes interspecific hybridization in polyploid Potamogeton

Nikita Tikhomirov

Nikita Tikhomirov, Alison Scott, Alexander Bobrov, Zdenek Kaplan, Polina Novikova

Institute of Botany CAS (Czech Republic), Max Planck Institute for Plant Breeding Research (Germany), Papanin Institute for Biology of Inland Waters RAS (Russian Federation)

Three recurrent whole-genome duplication events produced three species-rich lineages in Potamogeton, the largest aquatic plant genus. Despite diverging millions of years ago, the species of the same ploidy can hybridize and produce fertile offspring. To study the effect of polyploidy and admixture on the evolution of Potamogeton, we generated 14 (and counting) PacBio HiFi-based genomes and resequenced 500 samples covering all main lineages in the genus. We dated two whole-genome duplications by synteny-guided analysis of paralog divergence, yielding ages of 65 Mya (hereafter the “ancient” event) and 13 Mya (the “recent” event). The ancient duplication is shared by all Potamogeton taxa, and the recent by half of the extant Potamogeton species. The third polyploidization is even more recent and is shared by only two species. The ancient polyploids have undergone substantial fractionation, as evidenced by the single-copy state of BUSCO genes and the high divergence of paralogs. Nevertheless, we found long self-collinear blocks formed by paralogs lasting from the ancient whole-genome duplication. The recent polyploids still exhibit many duplicated BUSCO genes and also demonstrate little structural variation. Furthermore, the recent genome duplication has likely happened through the hybridization of different species, resulting in the combination of two dissimilar subgenomes. Despite the expectation that genomic rearrangements follow polyploidization, Potamogeton maintained a highly conserved genome structure. We hypothesize that this genomic stability might enable admixture between species, which could be beneficial under low population sizes and/or mutation rates.
Differential Introgression in a bunting hybrid zone illuminates the genes underlying early genetic barriers between incipient species

Niloofar Niloo Alaei Kakhki

Niloofar Alaei Kakhki, Sajad Noori, Ricardo J. Pereira
Stuttgart State Museum of Natural History (Germany)

Studying hybrid zones offers unique insights into the evolution of reproductive barriers and their persistence amidst gene flow. Hybridization has diverse outcomes, such as the breakdown of reproductive isolation, emergence of adaptive phenotypic variability and the adaptive introgression. We investigated a narrow moving hybrid zone between the black-headed bunting (Emberiza melanocephala) and red-headed bunting (E. bruniceps) located in the northeast of Iran, where species boundaries are putatively maintained by behavioral isolation based on male coloration and new phenotypes arise through hybridization. Our ecological niche models indicate that this hybrid zone is older than the Last Glacial Maximum and reveals strong divergence in environmental preferences between parental taxa. Whole genome sequencing data reveals substantial divergence in mitochondrial lineages, contrasting with reduced average divergence in the nuclear genome, consistent with extensive introgression upon secondary contact. A clear non-random association between mtDNA lineages and male plumage coloration suggests a strong female preference for conspecific males maintaining this hybrid zone. Chromosome painting demonstrates genome-wide admixture, except in genomic areas characterized by high differentiation peaks, consistent with localized barriers to gene flow or regions of low recombination. Spatial sampling will assess the hybrid zone’s movement in the last 50 years, possibly due to adaptive introgression of coloration genes. Our results suggest that these sister species have exchanged genes over thousands of generations, making this hybrid zone an ideal natural laboratory to pinpoint the barriers against the gene flow and understand the role of behavioral isolation in the dynamics of species boundaries.
Tracing the impact of admixture in signatures of selection in the Mexican Biobank.

Ram González-Buenfil

Ram González-Buenfil, Fernanda Mirón-Toruño, Santiago G. Medina-Muñoz, Austin W. Reynolds, Andrés Moreno-Estrada
Baylor University (USA), CINVESTAV (Mexico), Department of Anthropology, UGA-LANGEBIO

Events of admixture (i.e., gene flow between populations) have occurred throughout the human lineage due to migration and other demographic movements that exposed individuals to new environments and different selective pressures. Present-day Mexico is home to one of the greatest admixture events in recent history primarily between Indigenous American and Europeans, which had evolved independently from each other for thousands of years until ~500 years ago. This composition represents a unique setting to study the effects of admixture on adaptation. In Latin America particularly, the extent to which admixture has shaped adaptation processes remains poorly explored. To investigate this, we are analyzing genomic data from the Mexican Biobank (MXB), consisting of 6,057 individuals across all 32 states in Mexico and genotyped for 1.8 million genome-wide SNPs. To detect signatures of selection we are using two novel statistical models, Adaptmix and OHANA, which model allele frequencies in admixed populations to identify selection and distinguish whether it has occurred before or after admixture. We are also employing phenotypic data from MXB to correlate our candidate regions with potential traits under selection. Preliminary findings suggest the presence of both ancestral and post-admixture signatures of selection, with implications for the study of evolution in diverse populations from the Americas. This study leverages the largest dataset of its kind for the Mexican population and marks a significant advance in understanding the genetic basis of adaptation in admixed populations.
Global, asynchronous sweeps at multiple insecticide resistance genes in Aedes aegypti mosquitoes

Thomas L Schmidt

Thomas L Schmidt
University of Melbourne (Australia)

Aedes aegypti (yellow fever mosquito) is a highly invasive pest that confers the majority of the global dengue burden. This species successfully invaded much of the world’s tropical and subtropical regions between the 16th and 19th centuries. Insecticide-based management beginning in the mid-20th century led to the evolution of insecticide resistance in Ae. aegypti, and identical point mutations at the voltage-sensitive sodium channel gene (VSSC) on chromosome 3 have been observed across its global distribution. This study uses a genomic analysis of selective sweeps to infer how this global distribution of mutations was reached, and found a combination of adaptive introgression and repeated substitutions at the same nucleotide position. Surprisingly, a second locus on chromosome 2 was found to harbour variation segregating in line with VSSC genotype, and this locus showed similar patterns of positive selection to those observed at the VSSC gene. This locus contained 15 resistance-associated glutathione S-transferase (GST) epsilon class genes, and we detected three distinct homozygous haplotypes that had introgressed across the Indo-Pacific region, the Americas, and Australia. North American Aedes aegypti showed signs of significantly higher copy number of these genes. VSSC and GST sweeps had similar broad geographical patterns, but local patterns and linkage networks indicate these likely spread at different times. These findings highlight the strong patterns of selection acting on resistance genes globally and the likely key importance of GST genes in resistance evolution.
Dobzhansky-Muller incompatibilities and adaptive introgression facilitate explosive speciation of Lake Baikal amphipods

Valentina Burskaia

Valentina Burskaia, Ilia Artiushin, Dmitry Fedorov, Yakov Korobitsyn, Georgii Bazykin, Hannes Svardal
Kharkevich Institute for Information Transmission Problems (Russian Federation), Lomonosov Moscow State University (Russian Federation), Skoltech (Russian Federation), University of Antwerp (Belgium)

The adaptive radiation of amphipods in Lake Baikal has brought forward more than 340 species (20% of the world’s freshwater amphipods), making it one of the largest species flocks after the famous African cichlid radiations. We analyzed multi-species transcriptomic (60 species) and genomic (15 species) alignments of Lake Baikal Amphipoda. Our results indicate that hybridization between two ancient independent lineages of Baikalian amphipods preceded the period when fast speciation started. It was followed by differential fixation of introgressed loci in all main clades of the species flock. We found intriguing signals of Dobzhansky-Muller incompatibilities as well as positive selection on introgressed loci. These findings provide evidence that hybridization likely facilitated adaptive radiation of Lake Baikal amphipods. Both processes have been shown to occur in other adaptive radiations, but their functional role in rapid diversification is still debated: hybridization may also randomly coincide with speciation. In our set of hybridization tests and selection scans we demonstrate, that introgressed material undergoes different types of selection, which is unlikely to be a consequence of neutral processes.
Genetic effects on complex traits and diseases are similar across segments of different continental ancestries in the Mexican Biobank

Yuridia Selene Posadas García

Yuridia Selene Posadas García, María J. Palma Martínez, Valeria Añorve Garibay, Lourdes García García, Andrés Moreno Estrada, Hou Kangcheng, Pasaniuc Bogdan, Mashaal Sohail

Bioinformatics Interdepartmental Program, CINVESTAV (Mexico), Center for Computational Molecular Biology (USA), Centro de Ciencias Genómicas de la UNAM (Mexico), David Geffen School of Medicine at UCLA (USA), Department of Computational Medicine, Department of Human Genetics, Department of Pathology and Laboratory Medicine, Instituto Nacional de Salud Pública (Mexico), Laboratorio Internacional de Investigación sobre el Genoma Humano de la UNAM (Mexico), National Laboratory of Genomics for Biodiversity (LANGEBIO), UCLA (USA)

A hotly debated area in human genomics is the role of diversity in medicine. In particular, how similar are genetic effects on complex traits and diseases among different ancestral backgrounds? Recently a novel computational approach (admix-kit) was developed to compute correlation of causal effects (radmix) across local ancestries. Published results showed similar effects across ancestries in African-European admixed individuals sharing the same environment, concluding that genetic causal effects have limited contribution to reduced accuracy of polygenic scores. Here, we estimated correlation of effect sizes for 9 binary and 11 quantitative complex traits and diseases among Indigenous, African, and European ancestries by applying admix-kit to 5,833 individuals from the Mexican Biobank. Individual genotypes were imputed using Topmed and local ancestries were inferred using Gnomix. We compared segments of each ancestry against an aggregate of the other two ancestries. We showed that for all analyzed traits and diseases, radmix across ancestral backgrounds was not significantly different from one (p > 0.0025). Meta-analyzing across traits, we estimated radmix values of 0.88 (95% CI: 0.7290-1) for African segments, 0.977 (95% CI: 0.937-1) for Indigenous segments and 0.943 (95% CI: 0.896-0.982) for European segments. To corroborate our conclusions, we will assess the prediction performance of methods integrating GWAS from different populations in a Bayesian approach such as PRS-CSx compared to those estimating effect sizes for different local ancestral backgrounds separately. Our findings hold significant implications for determining the best approach to estimate polygenic scores in admixed individuals.
S13 - Animal paleogenomics beyond higher latitudes.
Compacted hair in broken carnivore teeth reveal dietary prey of historic lions

Alida de Flamingh

Alida de Flamingh, Tom P Gnoske, Julian C Kerbis Peterhans, Vellizar Simeonovski, Nduhiu Gitahi, Ogeto Mwebi, Bernard R Agwanda, Julian M Catchen, Alfred L Roca, Ripan S Malhi
Field Museum of Natural History (USA), National Museums of Kenya (Kenya), University of Illinois at Urbana-Champaign (USA), University of Nairobi (Kenya)

The synergistic advancement of molecular and computational technologies has pushed genomics into a new era; complete nuclear genomes can now be sequenced from minuscule quantities of DNA and from specimens that are more than a million years old. DNA analysis from hair samples is a well-established approach widely used in human forensics and wildlife conservation science. Hair samples are less prone to contamination from external DNA sources, and can be used to identify the species of animal from which the hair originated. Here we used ancient DNA and bioinformatic methodologies optimized for low quantity and quality degraded DNA, to systematically identify the dietary prey species from four individual hairs and three hair clumps compacted in the teeth of the infamous Tsavo lions that are known to have also preyed on humans during the late 1890’s in Kenya. Molecular and bioinformatic analysis of hair DNA identified giraffe, human, oryx, waterbuck, wildebeest and zebra as prey species. DNA preservation allowed for the reconstruction of complete mitogenome profiles for zebra and giraffe. Giraffe mitogenomes are phylogeographically partitioned, and we found that the Tsavo lions ate at least two individuals that belong to a subspecies of Masai giraffe (Giraffa tippelskirchi tippelskirchi) typically found in southeast Kenya. The protocol and approach reported here enable a better understanding of the hunting behaviors, diets, and ecology of historic individuals, populations, and species and holds promise for elucidating these characteristics in extinct populations and species.
Small but mighty: ancient DNA analyses of late Pleistocene and Holocene microvertebrates from Central Texas

David Treviño Ledesma

David T Ledesma, Melissa E Kemp
The University of Texas at Austin (USA)

Fossil identifications are foundational for making robust interpretations of past evolutionary and ecological processes. Morphology has traditionally been the main method for fossil identification; however, ancient DNA (aDNA) has emerged as an exciting new tool for identifying fossils to lower taxonomic levels than what morphology can often provide. The ability to recover aDNA from fossils is known to be dependent on age and environment, with older and lower latitude sites being less suitable for long-term DNA preservation. Furthermore, aDNA research is relatively limited for microvertebrates in part because recovery of aDNA becomes increasingly difficult with smaller amounts of starting material for the extraction process. Here we investigate our ability to obtain aDNA from late Pleistocene to Holocene microvertebrate herpetofauna remains from a low latitude site in Central Texas. We successfully extracted aDNA from 15 out of 20 specimens ranging from 0.9 to 56 mg. We found a positive linear relationship between fossil mass and DNA yield with few fossils below 10 mg producing detectable amounts of DNA. Preliminary shotgun sequence data of aDNA extracted from salamander fossils confirms the presence of endogenous DNA. We performed phylogenetic analyses with extant species and corroborated morphological identification of fossils to a clade of tiger salamanders. aDNA indicates that the fossils are closely related to species that today occur further west and south of the fossil locality. These results reinforce the merit of aDNA for microvertebrate material from low-latitude sites and provide insights into past diversity and biogeography of Central Texas salamanders.
Evaluating genetic dating under different molecular clock calibration scenarios to better understand the evolutionary trajectory Basin of Mexico Columbian mammoths

Eduardo Arrieta Donato

Eduardo Arrieta Donato, Angeles Tavares Guzman, Alejandra Castillo Carbajal, Miriam Bravo Lopez, Ernesto Garfias Morales, Viridiana Villa Islas, Vanssy Li, Alejandro Lopez Jimenez, Mashaal Sohail, Camilo Chacon Duque, Joaquin Arroyo Cabrales, Maria Avila Arcos, Federico Sanchez Quinto

International Laboratory for Human Genome Research - UNAM (México), Centro de Ciencias Genómicas - UNAM (México), Centre for Palaeogenetics (Sweden), Department of Zoology, Stockholm University (Sweden), Laboratorio de Arqueozooología - INAH (México), Subdirección de Laboratorios y Apoyo Académico - INAH (México)

The Columbian mammoth was the only mammoth species endemic to the Americas. To date, its evolutionary history has only been characterized using genetic data from samples across the US. Recently, the discovery of Columbian mammoth remains at the Mexico Basin provides an opportunity to further investigate its evolutionary trajectory across the American continent. Preliminary phylogenetic analyses from 28 capture-enriched mitogenomes and mammoths worldwide, suggest that three different lineages were present in the Mexico Basin. Time-scaled phylogenetic inference is enhanced by including either isotopic or genetic individual dating. Currently only five Mexican samples have been 14C-dated. In order to maximize the number of samples included in phylogenetic analyses, we perform individually genetic dating of non-dated Mexico Basin Columbian mammoths under different molecular clock calibration scenarios using C14-dated data as controls. This approach allowed us to generate a reliable set of genetic dates, as well as to better understand the factors affecting molecular clock calibration. Our findings reveal that sample divergence is the main variable influencing genetic age estimation, and impacted by how well population structure is represented in molecular clock calibration references. This study provides a better understanding of the factors affecting time-scaled phylogenetic inference, as well as a better interpretation of the evolutionary history of the Columbian mammoth in the continent, particularly its demographic dynamics in the Mexico Basin.
Recent paleogenomic studies have suggested a hybrid origin for the Columbian mammoth deriving from the admixture of an ancestral woolly mammoth and a steppe mammoth lineage. While its geographical range once extended from North to Central America, only mitochondrial data is currently available from across the USA. Moreover, only one Columbian mammoth autosomal genome has been generated to date, obtained from a sample close to the putative woolly-Columbian hybrid zone. Such genetic sampling hinders broad conclusions about its evolutionary trajectory. Recently, multiple Columbian mammoth remains have been discovered in the Basin of Mexico, providing a unique opportunity to better understand their evolutionary origin. We screened 72 libraries from M. columbi, authenticating their endogenous origin and molecularly sexing individuals. We performed mitochondrial DNA (mtDNA) capture-enrichment experiments recovering 28 genomes with >10x coverage, and generated a >0.5x nuclear genome. We compared this data to mitochondrial and nuclear reference sequences, respectively, from Columbian and woolly mammoths across North America and Eurasia, and performed phylogenetic and demographic analyses. Inferences stemming from both datasets provide a more detailed picture of the evolutionary trajectory of the Columbian mammoth compared to what has been previously reported. Thus, these results highlight the importance of recovering ancient genomic data from wider geographical ranges in order to fully understand the evolutionary history of extinct species. Our study provides a better understanding of the evolutionary origin of the Columbian mammoth in Mexico and America, as well as valuable insights into the evolutionary trajectory of the mammothus genus.
The Eurasian lynx (Lynx lynx), once widespread across Eurasia, was recently extirpated from western Europe, mainly due to anthropogenic pressures and habitat fragmentation. Although the past distribution of the species encompassed the Iberian Peninsula from the Late Pleistocene up until the 17th century, little is known about the relationship of these Iberian populations with other Eurasian lynxes. Here we present the paleomolecular study of the last recorded specimen of Eurasian lynx from the Iberian Peninsula (dated to ~200 yBP, from Sima Topinoria, Cantabria), which has been analysed at genetic level using cutting edge technologies for the sequencing of ancient DNA. The specimen showed an excellent state of biomolecular conservation, preserving up to 60% of endogenous DNA, which allowed the reconstruction of its complete mitochondrial genome and the analysis of genome-wide nuclear markers. The phylogenetic analysis of the mitogenome revealed that the lynx from Sima Topinoria (ST) belonged to an extinct lineage that is basal to all modern populations of Eurasian lynx, distributed from Northwestern Europe to Northeast Asia. This lineage identifies ST as belonging to a relict population of Pleistocene lynxes that survives in the Cantabrian mountains until recent times. Indeed, the molecular identification of this specimen revises the date of the species extirpation from the Iberian Peninsula to only 200 years ago.
Molecular data from South African species dating to 2 million years ago

Ioannis Patramanis

Ioannis Patramanis, Palesa P Madupe, Claire Koenig, Clement Zanolli, Lauren Schroeder, Fernando Racimo, Jesper V Olsen, Rebecca R Ackermann, Enrico Cappellini

Center for Protein Research, Globe Institute, University of Bordeaux (France), University of Cape Town (South Africa), University of Copenhagen (Denmark), University of Toronto Mississauga (Canada)

Although the field of ancient DNA has made significant advancements in deep time recovery, enabling scientists to recover genetic material from fossils up to a 1 million years ago, such successes are largely confined to specimens from regions of the world where DNA tends to be well preserved, such as Siberia and northern Canada. In contrast, the tropical and subtropical climates of Africa and southern Asia, where most ancient hominin diversity is found, present substantial challenges for DNA preservation. Here, by employing tandem-mass spectrometry, we sequence the enamel proteomes of multiple 2-million-year-old hominin and mammalian specimens from South Africa, offering an alternative for extracting valuable genetic information from areas where ancient DNA recovery is not currently viable. Using the recovered protein sequences, we are able to determine the biological sex of the samples as well as identify within-species variation, unlocking the potential to study sexual dimorphism, population diversity and possible substructure in populations of the deep past. We show that the recovered proteins can also be utilized to generate phylogenetic trees, but demonstrate that phylogenetic resolution depends substantially on the amount of protein sequence available to us. Our approach demonstrates the feasibility of recovering informative genetic sequences from the early Pleistocene and from some of the most challenging preservation conditions in the globe, but also highlights how these molecular data can contribute valuable insights into our understanding of the past.
Insights into the evolutionary history of Late Pleistocene Mexican camels through mitochondrial DNA phylogenetic analyses

María José Rodríguez Barrera

Camelops genus first appears in the fossil record between 4 to 3 million years ago and lived across North America until its extinction ~12,000 years ago at the end of the Pleistocene. Recent paleogenomic studies have shown that mitochondrial genomes from Late Pleistocene camels (Camelops hesternus) from Yukon, Canada, related as a sister clade to the genetic variation of all present-day camel species. This result confirms previous reports of an American origin for the Camelidae family. However, prehistoric camels once lived across all North America, therefore it is not clear to what extent Yukon camel genetic data represents the genetic variation of camels across the continent. The recent discovery of camel paleontological remains in the Basin of Mexico during the construction of the Mexico City International Airport, provides an opportunity to further study evolutionary history of Camelops genus in America with a broader geographical sampling. In this study we extracted DNA, sequenced and analyzed paleogenomic data from seven camel samples from the Basin of Mexico. We identified the molecular sex of these individuals and performed mitochondrial capture enrichment experiments. We recovered three whole mitochondrial genomes >1x coverage, and carried out phylogenetic analysis to test the relationship of Mexican prehistoric camels to both present and ancient Camelops sequences. This study is pivotal in unraveling the evolutionary history of camels in America by characterizing the genetic diversity and demographic trajectory of prehistoric camels in Mexico.
Thinking about the Roman Empire one more time: Roman sheep genomes and the demographic history of domestication in the western Mediterranean

Marianne Dehasque

Marianne Dehasque, Viviana Floridia, Pedro Morell Miranda, Paolo Mereu, Mario Pirastru, Leoni Giovanni Giuseppe, Salvatore Naitana, Nicolò Columbano, Marco Zeddi, Mario Barbato, Torsten Günther

Department of Organismal Biology, Dipartimento di Medicina Veterinaria, Dipartimento di Scienze Biomediche, Dipartimento di Scienze Veterinarie, Università degli Studi di Messina (Italy), Università degli Studi di Sassari (Italy), Uppsala University (Italy), Uppsala University (Sweden)

Sheep (Ovis aries) were domesticated in the Fertile Crescent over 10,000 years ago and spread to the rest of the world through multiple domestication and expansion waves. The Roman period, a time characterised by high mobilisation, has been hypothesised to have had a big impact on the genomic make-up of present-day sheep populations. However, to date no sheep genomes from this period have been published. In this study, we generate genome-wide data for two Roman sheep samples from Sardinia and one from Spain and compare these to an extensive dataset consisting of 12 ancient Mediterranean samples and modern genomes from over 30 wild and domestic sheep lineages. We investigate demographic history and the signatures of selection in our western Mediterranean dataset. Preliminary analysis of the oldest Sardinian sample reveals a unique genetic signature when compared to present-day sheep, suggesting that the history of sheep breeds was even more complex than previously thought. Further analysis will provide important insights into the impact of the Roman period on sheep diversity.
Taxonomic identification of the prehistoric horses from the Basin of Mexico using ancient mitochondrial DNA
Pablo Esteban Uribe-Herrera

Pablo Esteban Uribe-Herrera, Viridiana Villa-Islas, Alejandra Castillo-Carbajal, Ernesto Garfias-Morales, Eduardo Arrieta-Donato, Miriam
Bravo-Lopez, Alejandro Lopez-Jimenez, Joaquin Arroyo-Cabrales, Maria Avila-Arcos, Federico Sanches-Quinto
International Laboratory for Human Genome Research - UNAM (México), Centro de Ciencias Genómicas - UNAM (México), Facultad de
Estudios Superiores Iztacala - UNAM (México), Laboratorio de Arqueozoología - INAH (México)

North America was the birthplace of the Equidae family and witnessed most of its evolution. Mexico has an extensive fossil record of
prehistoric horses up to the Late Pleistocene. Despite this, the exact number of pleistocene equine species in North America, and their
phylogenetic position is problematic due to the limitations of morphological-morphometric characterization methods and the complexity
of the Equus genus. Cheek-teeth morphology characterization has proven useful for species identification. However, it has shown limited
resolution when individuals display traits with similar morphological features. An example of this are the prehistoric horse species
(characterized to date) that inhabited Mexico: E. mexicanus, E. conversidens and Harintonhippus francisci. Ancient DNA provides an
opportunity to perform high resolution species identification. Recently, 15 individualized remains of prehistoric horses were found at the
Santa Lucia site dated to 12-30kya in the Basin of Mexico. In this study, we extracted and sequenced ancient DNA from seven prehistoric
horses, and captured-enriched three mitochondrial genomes to >1x coverage. We performed mitochondrial phylogenetic analyses
including our samples with other ancient and present-day North American-Beringian horses as references and identified their taxon
affiliation either on the Equus or Harintonhippus lineages. This study emphasizes the importance of species identification through ancient
DNA, as well as provides a dataset to further understand the diversity and evolutionary trajectory of prehistoric horses in Mexico and
North America.
Assessing pre-exploitation baseline numbers and population dynamics of the European sardine Sardina pilchardus Walbaum, 1792 using palaeogenomics

Paula F. Campos

Gonçalo E. Themudo, Adolfo F. Fernández, Carlos Fernández-rodriguez, Eduardo GG de Aguero, Maria C del Arco Aguilar, Sónia Gabriel, Rute R da Fonseca, Paula F. Campos

CIIMAR, CIIMAR/UP (Portugal), FAUP (Portugal), The Globe, Universidad La Laguna (Spain), Universidad Leon (Spain), University of Copenhagen (Denmark), University of Leon (Spain), University of Porto (Portugal), University of Vigo (Spain)

Overfishing has been a major problem for several marine species, reducing their effective population sizes to remnants of their pristine levels. However due to the lack of accurate fishery catch data we still have a poor understanding of how these impacted their evolutionary path. One of the most impacted species was the European sardine (Sardina pilchardus), an important pelagic fish resource in Atlantic waters, with enormous economic value especially in Southern Europe and Morocco, where it is the main target of purse-seine fleets and represents a major source of income for local populations. The species is distributed from the southern Celtic Sea and the North Sea to Mauritania and Senegal, including the Azores, Madeira and the Canary Archipelagos, being also abundant in the Mediterranean. Extant populations show three genetic clusters: one including individuals from Azores and Madeira, the second encompassing Iberian populations (the centre of the sampling distribution), and the third gathering the Mediterranean and Canary Islands, with individuals from Iberia showing some degree of admixture. Here we use state of the art ancient DNA techniques to look at samples that pre-date periods of intensive fishing and compare those results to the ones obtained from current populations (post exploitation). Doing this will enable us to get a more accurate perspective of the effective population size and genetic diversity levels of the species prior to overfishing, invaluable information for stock delimitation and management and definition of fishing effort in the different Food and Agricultural Organization (FAO) fishing areas.
Comparison and optimization of protocols and whole genome capture conditions for ancient DNA samples
Reyhan Yaka

Reyhan Yaka, Vendela Kempe Lagerholm, Maja Krzewinska, Anna Linderholm, Füsun Özer, Mehmet Somel, Anders Gotherstrom
Centre for Palaeogenetics, Department of Anthropology, Department of Archaeology and Classical Studies, Department of Biological Sciences, Hacettepe University (Turkey), Middle East Technical University (Turkey), Stockholm University (Sweden)

Ancient DNA (aDNA) data obtained from archaeological remains is an expanding and well-established field that has already helped resolve various questions in evolutionary biology, archaeology and anthropology. However, aDNA also exhibits significant technical challenges. Over the last decade, new approaches have mitigated these issues. But obtaining sufficient authentic aDNA still remains a challenge for samples from temperate, warm and humid areas. Here, we studied a set of methods for obtaining aDNA data from archaeological samples spanning different sites and periods, and with various levels of preservation. We compared DNA extraction and library preparation protocols, and tested the efficiency of whole-genome enrichment (WGC) on ancient samples by modifying a number of parameter combinations. We find that the Dabney extraction protocol performs significantly better than alternatives, and observe a positive trend with the BEST library protocol indicating lower clonality. Our results further show that WGC is effective at retrieving endogenous DNA, particularly from poorly preserved samples. This is an important finding suggesting that WGC can work exceptionally well on aDNA samples with very low endogenous content and contrasts with previous studies. We also suggest WGC as an alternative technique to the targeted SNP-capture method. This is because WGC has the potential to generate unbiased data, and thereby, could present a viable option in future aDNA research targeting low quality libraries. Overall, our results suggest that WGC is an efficient and cost-effective method for retrieving endogenous DNA in ancient samples, and holds promise to improve new techniques in the field.
An autosomal genome of a Mexican Columbian mammoth informs about its hybridization dynamics with the wooly mammoth in the Americas

Rigoberto Padilla Bustos

Rigoberto Padilla Bustos, Alejandra Castillo Carbajal, Eduardo A Arrieta-Donato, Miriam Bravo López, Ernesto Garfias Morales, Ángeles Tavares Guzmán, Viridiana Villa Islas, Alejandro López Jimenez, Mashaal Sohail, Joaquín Arroyo-Cabrales, María C Ávila Arcos, Federico Sánchez Quinto

International Laboratory for Human Genome Research - UNAM (México), Centro de Ciencias Genómicas - UNAM (México), Laboratorio de Arqueozoológía - INAH (México)

The hybridization dynamics between different species have been a long-studied research subject addressed by multiple disciplines, most recently by paleogenomics. Recent studies suggest that the Columbian mammoth resulted from hybridization between an ancestral lineage of Siberian woolly mammoths and possibly the steppe mammoth (50-50%), followed by subsequent gene flow (~10%) from the woolly mammoth into the Columbian mammoth in North America. However, more genetic data on the Columbian mammoth is needed to better understand the hybridization dynamics between both species in the Americas, particularly since only one Columbian mammoth genome is available to date. Additionally, this genome stems from an individual from Wyoming, at their putative hybridization zone between USA and Canada, further complicating the interpretation. The discovery of several Columbian mammoth remains at the Basin of Mexico provides a unique opportunity to further understand the demographic dynamics between both species in the continent. We screened 72 libraries and generated a >0.5x nuclear genome for one individual. We used this nuclear genome to investigate the hybridization dynamics between both species in the Americas, by comparing it to the Columbian mammoth individual from Wyoming, and other woolly mammoth genomes from North America and Eurasia. Our study provides a deeper understanding of the relationship between both mammoth species in the American continent and paves the way for future projects to continue to tackle mammoth evolutionary history.
Museomics and the Tropics: occasionally vexing, but a rewarding partnership
Selina Brace

Presented by self
London (UK), Natural History Museum

Natural history collections can be traced back to the cabinets of curiosities from the 16th century, with early museum collections often privately owned for personal study or explicitly for the use of scientific societies. Whilst public accessibility has changed, natural history museums have remained hugely important scientifically as repositories for type specimens, global biodiversity and population diversity. In recent years however they have entered a new era, that of museomics, and are often now referred to as storehouses of genomic information. Whilst it is widely appreciated that DNA from museum collections will be fragmented and degraded and that age be taken into consideration, environmental factors are also crucial, with warmer environments particularly associated with decreasing DNA preservation. Non-temperate, tropical and subtropical museum samples could therefore be considered as poorly chosen subjects for the new era of museomics, except that tropical regions form the cradle of biodiversity, associated with a wealth of extinct and extant species. This presentation will look at the use of non-temperate museum samples for DNA analyses.Highlighting current museomic work on endemic Indonesian species babirusa (Babyrousaspp) and anoa (Bubalus spp.), New Guinean echidna (Zaglossus spp) and extinct Chilean giant ground sloth (Mylodon darwinii) to illustrate some of the questions that can be addressed, challenges faced and potential solutions and future directions.
What\'s (read) size got to do with it?: An assessment of reference bias and spurious mapping
Stephanie Dolenz

Stephanie Dolenz, Tom van der Valk, Chenyu Jin, Jonas Oppenheimer, Muhammad Bilal Sharif, Ludovic Orlando, Beth Shapiro, Love Dalén, Peter D. Heintzman
Centre for Anthropobiology and Genomics of Toulouse, Centre for Palaeogenetics (Sweden), Department of Bioinformatics and Genetics, Department of Biomolecular Engineering, Department of Ecology and Evolutionary Biology, Department of Geological Sciences, Department of Zoology, Howard Hughes Medical Institute, Science for Life Laboratory (Sweden), Stockholm University (Sweden), Swedish Museum of Natural History (Sweden), University Paul Sabatier (France), University of California Santa Cruz (USA)

The alignment of sequencing reads to a reference genome is a critical step in the characterization of ancient genomes. However, reference bias and spurious mappings pose a significant challenge, particularly as cutting-edge wet lab methods generate datasets that push the boundaries of alignment tools designed for less diverged reads. Reference bias occurs when reference alleles are favored over alternative alleles whilst mapping to the same or a closely related species, whereas spurious mappings stem from either contamination or when endogenous reads fail to align to their correct position of origin. Previous work has shown that the extent of these phenomena is correlated with read length but a more thorough investigation of the nature of reference bias and spurious mappings with ancient DNA mapping parameters has been lacking. Here, we use a range of empirical and simulated palaeogenomic datasets to investigate the impacts of mapping tools, mapping quality thresholds, and reference genome choice, on mismatch rates across all read lengths. We find that mapping algorithms and quality threshold choices dictate reference bias and rates of spurious alignment at different read lengths in a predictable and often conflicting manner, suggesting that optimized mapping parameters for each read length will be a key step in alleviating both reference bias and spurious mappings. This will allow for more robust analyses of faunal palaeogenomic datasets and increased nuance of evolutionary inferences than is currently possible for many extinct taxa.
Sedimentary ancient DNA studies in challenging preservation contexts: Insights from the southern Levant

Viviane Slon

Presented by self
Tel Aviv University (Israel)

The analysis of DNA extracted from ancient biological samples has significantly advanced our understanding of prehistoric ecosystems and biodiversity, revealing intricate evolutionary patterns of ancient organisms. A promising frontier in this field lies in the recovery of DNA fragments from ancient sediments, which are ubiquitous in archaeological sites, and can offer an unparalleled window into past human-environment interactions. Nevertheless, obtaining sufficient sedimentary ancient DNA for meaningful analyses remains particularly challenging in environments with difficult preservation conditions. Here, I will discuss our workflow to recover ancient DNA from archaeological sediments, addressing its advantages and limitations. I will outline our efforts to improve field sampling techniques and data generation methods, which could play a significant role in expanding the scope and accuracy of our investigations. Finally, I will touch upon our ongoing endeavors to recover DNA from prehistoric sites in the southern Levant, aiming to offer new insights into the genetic landscape of this region during ancient times.
S14 - Unveiling the evolutionary history of pathogens through paleogenomics.
Metagenomic Profiling of Ancient Pathogens in Britain

Anastasia Brativnyk

Anastasia Brativnyk, Kyriaki Anastasiadou, Christopher Barrington, Thomas Booth, Alexandre Gilardet, Isabelle Glocke, Sarah Johnston, Monica Kelly, Jesse McCabe, Marina Silva, Pooja Swali, Frankie Tait, Mia Williams, Pontus Skoglund
The Francis Crick Institute (United Kingdom)

In the last decade, the characterisation of ancient pathogens has significantly expanded our understanding of paleoepidemiology and the evolution of infectious diseases, such as plague, smallpox, and tuberculosis. Our laboratory has sequenced ancient DNA samples from across Britain, spanning an 11,000-year period. This time transect provides a basis for a comprehensive study of pathogens within a single region. Investigating the prevalence and dynamics of infectious diseases, clustered by time and space, may help us to uncover previously unknown outbreaks. Furthermore, constructing a spatiotemporal map of pathogens in Britain, allows us to study phylogenetic patterns to date divergences between lineages and taxa and to understand the timing of evolutionary transitions. Additionally, we plan to explore the evolutionary trajectory of specific pathogens, focusing on alterations in functional genomic elements, such as point mutations, insertions and deletions, that signify gains or losses of function over time. Mapping the landscape of ancient pathogens in Britain and examining key hypotheses into the evolutionary forces driving infectious diseases, has the potential not only to illuminate the history of diseases but also to contribute to modern public health strategies and improve our preparedness for future outbreaks.
Streptococcus evolutionary diversity in ancient Great Britain and its associations with oral health outcomes

Ava Gabrys

Ava Gabrys, Abigail Gancz, Laura Weyrich
Australian Centre for Ancient DNA, Department of Anthropology, Department of Biology, Huck Institutes of Life Science, The Pennsylvania State University (USA), University of Adelaide (Australia)

Streptococcus species play key roles in tooth decay, opportunistic infections, and structuring the oral microbiome through adhesion and interspecies interactions. Retracing the evolutionary history of these bacteria is critical to understanding the origins of modern diseases, including the costly development of dental caries. We used ancient DNA recovered from the dental calculus of 162 British individuals dating 2100BCE-1853CE to explore changes in streptococci communities over time. Employing a competitive mapping approach to finely characterize species abundances, we found major shifts in dominant Streptococcus species and their genomes occurred after the post-medieval period. Streptococcus sinensis—a species poorly described in populations today except in its involvement in infective endocarditis—was the dominant Streptococcus species in 54% of ancient samples, and phylogenomics suggest it had a stable population closely related to Streptococcus cristatus before significantly decreasing in abundance in the modern day. Streptococcus sanguinis is consistently the most abundant species in healthy modern individuals but was dominant in only 22% of ancient samples, where it was missing loci for membrane transport, stress response, and sugar usage that are variably present in modern individuals with periodontitis. Surprisingly, although the species thrives in modern carbohydrate-rich oral environments, Streptococcus mutans was nearly absent. We did not find significant associations between the presence of any Streptococcus species and oral pathologies, including caries and periodontitis. These compositional, genomic, and functional shifts suggest changing roles and competition of Streptococcus species post-industrialization, increasing our understanding of their relationships to human health and disease.
A comprehensive investigation of woolly mammoth remains associated microbes
Benjamin Guinet

Benjamin Guinet, Nikolay Oskolkov, Love Dalén, Tom van der Valk
Centre for Palaeogenetics (Sweden), Department of Bioinformatics and Genetics, Department of Biology, Department of Zoology, Lund University (Sweden), National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University (Sweden), Swedish Museum of Natural History (Sweden)

Throughout history, the exploration of ancient DNA (aDNA) has primarily centered on the evolution and demographics of humans and animals, emphasizing the analysis of eukaryotic aDNA. However, the discovery of microbial sequence data has revolutionized this field by uncovering the significance of host-associated microbial aDNA found within eukaryotic remains. Once considered a mere sequencing by-product, this microbial aDNA has emerged as a crucial source of insights into ancient pandemics, lifestyle patterns, and population movements. Despite the broad potential of obtaining evolutionary microbial insights from ancient DNA remains, the vast majority of studies to date have remained focused on humans. In this study we identified the microbial DNA found within 495 woolly mammoth remains, comprising tooth, tusk and skins samples, collected from a broad geographic range and timescale spanning the holocene (~4300 years old) up to samples over a million years in age. This extensive dataset enables us to assess and formulate hypotheses regarding numerous potential interactions between these ancient microorganisms and their hosts. Among the identified microorganisms, several related bacteria are known to interact with contemporary animals in either commensal or pathogenic ways. Taken together, these results will pave the way for a better understanding of the microbial interactions with mammoth populations over time.
Tuberculosis and Sociocultural Dynamics: Critical Insights from Paleogenomics in South America and Beyond

Elizabeth A. Nelson

Elizabeth A. Nelson, Jane E. Buikstra, Nicolas Rascovan
Arizona State University (USA), Institute Pasteur (France)

For millennia, tuberculosis (TB) has exacted a profound toll on human populations, marked by significant mortality rates and societal upheaval. Despite its ancient history, TB remains a leading cause of death worldwide, second only to COVID-19 in recent years. The emergence of antimicrobial-resistant strains underscores the urgent need for an improved comprehensive understanding of TB persistence on a molecular and social level. Paleogenomics is perfectly positioned to contribute contextualized genomic insights into TB evolution, shedding light on large-scale social forces influencing pathogen dynamics and persistence. Paleogenomics has already revolutionized our understanding of TB in the ancient Americas through molecular confirmation and genomic characterization of TB-causing pathogens in pre-Hispanic contexts, providing pivotal insights. Drawing from published literature and ongoing research, we explore the transformative impact of paleogenomic investigations on our understanding of TB histories in the Americas. We examine the TB presence and strain diversity across varied geocultural and ecological landscapes of pre-Hispanic South America. Through an interdisciplinary lens encompassing archaeology, paleopathology, and paleogenomics, we view these results within biocultural and environmental contexts, elucidating the intricate interplay between pathogens, people, and places over time. Considering TB epidemics’ association with socio-political turbulence, climatic disruptions, and subsistence strategies, we argue that contextualizing TB evolution yields critical insights into its prevalence and transmission patterns. Looking forward, we highlight key considerations and future directions to enhance the utility of pathogen paleogenomics in advancing our understanding of infectious disease dynamics in the Americas and beyond.
Tuberculosis and Sociocultural Dynamics: Critical Insights from Paleogenomics in South America and Beyond

Elizabeth A. Nelson

Presented by self
Institut Pasteur (France)

For millennia, tuberculosis (TB) has exacted a profound toll on human populations, marked by significant mortality rates and societal upheaval. Despite its ancient history, TB remains a leading cause of death worldwide, second only to COVID-19 in recent years. The emergence of antimicrobial-resistant strains underscores the urgent need for an improved comprehensive understanding of TB persistence on a molecular and social level. Paleogenomics is perfectly positioned to contribute contextualized genomic insights into TB evolution, shedding light on large-scale social forces influencing pathogen dynamics and persistence. Paleogenomics has already revolutionized our understanding of TB in the ancient Americas through molecular confirmation and genomic characterization of TB-causing pathogens in pre-Hispanic contexts, providing pivotal insights. Drawing from published literature and ongoing research, we explore the transformative impact of paleogenomic investigations on our understanding of TB histories in the Americas. We examine the TB presence and strain diversity across varied geocultural and ecological landscapes of pre-Hispanic South America. Through an interdisciplinary lens encompassing archaeology, paleopathology, and paleogenomics, we view these results within biocultural and environmental contexts, elucidating the intricate interplay between pathogens, people, and places over time. Considering TB epidemics' association with socio-political turbulence, climatic disruptions, and subsistence strategies, we argue that contextualizing TB evolution yields critical insights into its prevalence and transmission patterns. Looking forward, we highlight key considerations and future directions to enhance the utility of pathogen paleogenomics in advancing our understanding of infectious disease dynamics in the Americas and beyond.
Pre-Columbian Treponema pallidum in the Americas and the origin of treponematoses
Fernando González-Candelas

Fernando González-Candelas, Marta Pla-Diaz, Kerttu Majander, Louis du Plessis, Natasha Arora, Jose Filippini, Luis Pezo Lanfranco, Sabine Eggers, Verena J. Schuenemann
ETH Zürich (Switzerland), FISABIO (Spain), Natural History Museum (Austria), University of Basel (Switzerland), University of Sao Paulo (Brazil), University of Valencia (Spain), University of Zürich (Switzerland)

The origins of treponemal diseases have long remained unknown, especially considering the sudden onset of the first syphilis epidemic in the late 15th century in Europe and its hypothesized arrival from the Americas with Columbus’ expeditions. Recently, ancient DNA evidence has revealed various treponemal infections circulating in early modern Europe and colonial-era Mexico. However, there is no genomic evidence of treponematosis recovered from either the Americas or the Old World that can be reliably dated to the time before the first trans-Atlantic contacts. We have obtained a high-quality (33.6X coverage) treponemal genome from nearly 2,000-year-old human remains from Brazil, most closely related to the bejel-causing agent Treponema pallidum endemicum (TEN). Three additional, low-coverage genomes from the same site were also obtained, all clustering in the TEN clade. For their analysis, along with representative ancient and modern genomes of the main T.pallidum lineages, we have developed a new mapping strategy that substantially improves variant calling, thus providing better resolution for phylogenetic inference, including the identification of new recombination events. Contradicting the modern-day distribution of bejel, the results call into question the previous palaeopathological characterization of treponeme subspecies and showcase their adaptive potential. Also, we have obtained improved molecular clock estimations, placing the divergence of modern T.pallidum subspecies firmly in pre-Columbian times. Overall, our study demonstrates the potential of paleogenomics to uncover key events in pathogen evolution and emergence, paving the way to new hypotheses on the origin and spread of treponematoses.
Genomic traces of ancient pathogens in Central Patagonia (6000-100yBP)
Florencia Alvarez Gallego

Florencia Alvarez Gallego, Laura Carrillo Olivas, Miriam Jetzabel Bravo López, Kelly Blevins, Viridiana Villa Islas, Ernesto Garfias Morales, Alejandra Castillo, Jorge Alejandro Suby, María del Carmen Ávila-Arcos, María Laura Parolin
CCT-Conicet-Cenpat (Argentina), CCT-Conicet-Cenpat. (Argentina), Center for Bioarchaeological Research, Department of Archaeology, Durham University (United Kingdom), Instituto de Diversidad y Evolución Austral, Instituto de Investigaciones Arqueológicas y Paleontológicas del Cuaternario Pampeano (INCUAPA), International Laboratory for Human Genome Research. UNAM (Mexico), School of Human Evolution and Social Change (USA), UNCPBA-CONICET (Argentina), Universidad Nacional de la Patagonia San Juan Bosco (Argentina)

Understanding the prevalence and evolution of ancient pathogens in Patagonia, the last peopled part of the Americas, is crucial for revealing health dynamics and evolutionary history of the hunter-gatherer populations that inhabited this region. This study provides molecular insights into endemic infections among hunter-gatherers in Central Patagonia during a 6000 yBP transect. We examined a total of 36 human remains, including dental and bones samples with and without pathological changes. Unmapped reads to the human genome were taxonomically classified using Kraken2 and MALT. As an addition parameter was assessed through coincidences in the virulence factor database VFDB2022. We identified Tannerella forsythia, Treponema denticola, and other oral bacteria linked with various stages of periodontal disease in an individual from the Atlantic coastal region dating back to 950 yBP. In addition, systemic pathogens, including Proteus mirabilis and Corynebacterium urealyticum, associated with ulcers and urinary infections, were detected from the valley and coastal regions, dating back to 6000-2400 yBP, respectively. A noteworthy finding was the identification of Mycobacterium lepramatosis from Patagonia dating back to 2,000 yBP (n=1) and 4,000 yBP (n=2). This discovery has the potential to unveil previously unknown aspects of the health of ancient Patagonians and contribute to the poorly characterized evolutionary history of this pathogen, which was recently identified as a causative agent of leprosy. The next stage involves a capture-enrichment strategy of the pathogens of interest to increase the genomic coverage and facilitate phylogenetic analyses to reveal potential connections with other ancient and modern strains.
Ancient Salmonella enterica genomes and the evolutionary path of the Para C lineage

Gunnar U. Neumann

Salmonellosis, caused by Salmonella enterica, ranks among the most prevalent food-borne illnesses worldwide. Most of the approximately 2600 S. enterica distinct serovars have the capacity to infect a broad spectrum of host species, typically resulting in self-limiting gastrointestinal illness in humans. However, a few serovars are specifically adapted to humans, inducing enteric fever. Within this group, Paratyphi C belongs to the so-called Para C lineage, alongside Typhisuis (primarily infecting swine), and Choleraesuis (affecting both humans and swine). Previous aDNA studies suggested a link between host adaptation in this lineage and the Neolithization process involving pseudogenization and gene gain/loss. Here, we present a temporal and geographical transect of 35 new ancient S. enterica genomes spanning from 5000-500 years BP, collected from across Eurasia. The majority of genomes older than 3000 years form a distinct phylogenetic subclade lacking any modern representatives, suggesting probable extinction. The remaining genomes, dating to the last 2700 years, position along the ParatyphiC and Choleraesuis branches. Their sequential distribution enables a comprehensive examination of host adaptation, involving both the acquisition and loss of genomic loci and gene pseudogenization, and allows the reconstruction of the temporal sequence of these events. For instance, our findings indicate that the acquisition of the Salmonella pathogenicity island SPI-7 preceded the loss of the tcfABCD operon of SPI-6 in ParatyphiC. These findings provide valuable insights into the evolutionary trajectory and host adaptation process within the S. enterica Para C lineage, revealing how changes in genetic composition influenced its virulence.
Exploring the Health of Colonial Enslaved Communities at Hacienda La Quebrada: A Preliminary Study using Metagenomic Methods

Jaime Zolik

Jaime Zolik, Laura Pott, José Luis Santa Cruz Alcalá, Claire Maass, Maria A Nieves-Colón
APACHETA Consultores (Peru), MUAFRO (Afro-Peruvian Museum of San Luis) (Peru), SNA International (USA), Universidad Nacional Mayor de San Marcos (Peru), University of Minnesota-Twin Cities (USA)

The La Quebrada Archaeology Project, an ongoing community-led initiative, investigating an 18th century unmarked historical cemetery located in coastal Peru. The cemetery holds the remains of enslaved African and African-descendant laborers from the former plantation Hacienda La Quebrada. To contribute to the overarching goal of unraveling the history of these individuals and the broader context of African enslavement in Latin America, our laboratory conducted the extraction and analysis of ancient DNA (aDNA) from the dental pulp of 30 cemetery individuals. The recovered endogenous human aDNA had degradation patterns consistent with authentic aDNA. In conjunction with human aDNA analysis, our aim was to detect the presence of ancient microbes within our shotgun sequences to provide insights into the health, life histories, and potential disease dynamics of enslaved communities during the colonial period of Peru. Using established metagenomic tools, such as KrakenUniq and MALT, preliminary results suggest the presence of authentic Streptococcus gordonii, an opportunistic oral pathogen, in one individual displaying osteological indicators of dental disease (e.g., gum resorption, dental plaque). Aligned reads covered 54.85% (average 2.5x depth) of the S.gordonii reference genome. Although S. gordonii can live in environments outside of the human body, the aligned reads exhibited the same aDNA degradation pattern observed in the recovered human aDNA, indicating its authenticity. Future work includes examining phylogenetic context and conducting comparative analysis to modern local S.gordonii genomes. This study shows how combining paleogenomics and bioarch can lead to more holistic reconstructions of health in the past.
Exploring Pathogens and Demographic Dynamics in Colonial Mexico City through Paleogenomics
Laura Carrillo-Olivas

After European colonization, new pathogens caused epidemics that decimated the Indigenous population in the Americas. To understand post-colonization pathogen circulation, we analyzed the paleogenomics of individuals from El Templo de la Inmaculada Concepción ‘La Conchita’ (n=15) and the Hospital Real San José de los Naturales ‘HSJN’ (n=82), dated to 19th 16th-17th centuries respectively. We generated shotgun sequencing for both human and nonhuman genomes. For the human fraction, we determined the mitochondrial (mtDNA) haplogroups and the molecular sex. Both sites reveal a majority of Indigenous mtDNA lineages (n=75) with high confidence, and one HSJN individual has an African mtDNA haplogroup. This aligns with La Conchita’s reported Indigenous demographics and documentary evidence that HSJN was designated for mainly Natives (“Naturales”), but had a more diverse patient population. Sex distribution analysis indicates a high proportion of males in La Conchita, contrasting with a more balanced sex ratio in HSJN. The non-human fraction was taxonomically classified using Kraken2, and pathogenic bacterial species were identified through a thorough comparison with the virulence factors database VFDB2022. In HSJN, we identified the oral pathogen Tannerella forsythia (n=12), the agent of paratyphoid fever Salmonella enterica Paratyphi C (n=3), and tuberculosis Mycobacterium tuberculosis (n=1). Streptococcus pneumoniae, associated with respiratory infections, was identified in both sites (n=2). These results emphasize the multiplicity of pathogens circulating in colonial Mexico. Further sequencing of the ancient pathogen genomes of interest will allow phylogenetic analyses to identify potential epidemiological relationships between these ancient strains and other ancient and present-day pathogens.
Exploring the History of Malaria in the Americas Using Ancient DNA

Megan Michel


(Antigua and Barbuda), Adelaide Data Science Centre, Anatomy Institute, Anthropological Center, Archaeo- and Palaeogenetics, Arqueologia i Història Antiga, Austrian Academy of Sciences (Austria), Austrian Archaeological Institute, BioArch South (New Zealand), BioForge Canada Limited (Canada), Brooklyn College (USA), CNRS, CRAHAM - Centre Michel de Boüard - Centre de recherches archéologiques et historiques anciennes et médiévales, Center for Archaeological Sciences, Centre of Archaeological and Ethnographical Investigations (Russian Federation), Centro Mallqui (Peru), Centro Studi sulla Civiltà del Mare (Italy), Charles University (Czech Republic), Croatian Academy of Sciences and Arts (Croatia), Dalhousie University (Canada), Departamento de Prehistoria, Departamento de Biotecnología, Departamento de Humanidades: Historia, Department of Anthropology, Department of Anthropology and Heritage Studies, Department of Archaeogenetics, Department of Biology, Department of Biomedical Sciences, Department of Environmental Science, Department of Evolutionary Anthropology, Department of Forensic Medicine, Department of Genomes and Genetics, Department of Geological Sciences, Department of Geosciences, Department of History, Department of Human Evolutionary Biology, Department of Legal Medicine, Department of Medical Engineering and Biotechnology, Department of Microbiology and Immunology, Department of Prehistoric Archaeology, Department of Social and Cultural Anthropology, Dienst Archeologie - Stad Mechelen (Belgium), Diversification and Evolution Group, Division of Ancient Pathogens, Economics and Political Science, English Harbour (Antigua and Barbuda), Faculty of Arts, Faculty of Biological and Environmental Sciences, Faculty of History, Farmingdale State College (USA), Finnish Institute for Health and Welfare (Finland), Forensic Medicine Unit, Freelance (Italy), Freie Universität Berlin (Germany), Geografía y Arte, Griffith Centre for Social and Cultural Studies, Griffith University, Harvard University (USA), Heidelberg University (Germany), Helsinki Collegium of Advanced Studies, Heritage Department, HistoGenes (Germany), Hospital Virgen de los Lirios (Spain), Human Evolution and Archaeological Sciences, IUCA-Aragosaurus, Infection, Initiative for the Science of the Human Past at Harvard and Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean (MHAAM), Inrap - Institut national de recherches archéologiques préventives (France), Institut Pasteur (France), Institut für Prähistorische Archäologie, Institut für Ur- und Frühgeschichte, Institut für Urgeschichte und Historische Archäologie, Institute for Archaeological Sciences, Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces (Germany), Institute of Anthropology, Institute of Archaeology of the Czech Academy of Sciences (Czech Republic), Institute of Classical Archaeology, Institute of Heritage Sciences (INCIPIT), Instituto Internacional de Investigaciones Prehistóricas (Spain), Instituto Universitario de Investigación en Arqueología y Patrimonio Histórico, Instituto Universitario de Investigación en Ciencias Ambientales de Aragón, Integrative Prehistory and Archaeological Science, Journal of Infection in Developing Countries (Italy), Lakehead University (Canada), Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt (Germany), Ludwig Maximilian University, MacEwan
Malaria is a vector-borne disease caused by parasitic protozoa of the genus Plasmodium. It exerts one of the strongest selective pressures acting on the human genome and causes over 600,000 deaths each year. Nevertheless, significant questions remain regarding when and how particular malaria-causing parasites emerged as human pathogens and spread around the globe. In particular, debate persists over whether the two most widespread and clinically relevant malaria agents, Plasmodium falciparum and Plasmodium vivax, were present in the Americas prior to European contact, or whether they spread from Europe and/or Africa during the colonial period. To investigate this question, we present genome-wide nuclear and mitochondrial data from 36 ancient P. vivax, P. falciparum, and P. malariae strains, spanning 16 countries and 5,500 years of human history, including a high-coverage P. vivax genome recovered from a Chachapoya individual from the peri-contact Peruvian site of Laguna de los Cóndores. Genomic analysis of this dataset reveals close links between now-eliminated European P. vivax and ancient and modern Latin American lineages, suggesting a European source for strains circulating in the Americas today. Moreover, the Laguna de los Cóndores strain shows greater affinity to present-day Peruvian P. vivax compared to other Latin American populations, providing evidence for persistence of an endemic malaria focus from the early colonial period until today. Overall, our results provide a first glimpse into the processes which allowed P. vivax to gain a foothold in the Americas during the period of the Columbian Exchange.
Non-invasive sampling of Inca mummies yields undocumented pathogens and human ancestry
Michelle Hämmerle

Michelle Hämmerle, Meriam Guellil, Dagmara Socha, Susanna Sawyer, Olivia Cheronet, Mario Bernaski, Gabriela Recagno, Martin Kuhlwilm, Pere Gelabert, Ron Pinhasi
Center for Andean Studies, Human Evolution and Archaeological Sciences (HEAS) (Austria), Museo de Arqueología de Alta Montaña (MAAM) (Argentina), University of Vienna (Austria), University of Warsaw (Poland)

In 1999, the mummies of the "Children of Llullaillaco" were discovered at the top of the Llullaillaco volcano in Argentina at 6739 m. These mummies are among the best-preserved mummies ever found, which can be attributed to a combination of freezing temperatures, mild humidity, and an anaerobic environment. The Inca sacrificed them as part of the ritual of "Capacocha", likely for reasons such as enforcing social control and gaining socioeconomic benefits. In collaboration with the Museo de Arqueología de Alta Montaña in Salta, Argentina, we performed non-invasive and non-destructive sampling on the mummies under strict supervision from the curators. Through extensive sequencing of 27 samples of skin, anal and buccal tissue obtained from three individuals, we have recovered 5.4-fold, 14.3-fold, and 25.3-fold nuclear genomes from each of the children, respectively. We have used these to A) infer their affinities to native American populations, which gives insights into their geographical origins, and B) refine the demographic models of pre-contact South American populations. Our non-invasive non-destructive sampling approach has also enabled us to detect both commensal and pathogenic microbial and fungal species. Through our metagenomic analysis, we have recovered a high coverage Bartonella quintana genome from one of the children. To the best of our knowledge, this is the first ancient genome of this bacterium. We could also identify the presence of Human gammaherpesvirus 4 (HHV4) sequences. The study of the commensal microbiota promises to shed light on the living conditions of the individuals.
A Glimpse into Antiquity: Paleoproteomics Approach to Investigate Ancient Pathogens

Miguel Alejandro Navarro

Miguel A. Navarro, Miriam B. Bravo-Lopez, Glendon J. Parker, Marcela Sandoval-Velasco, Julie K. Wesp, Maria L. Moreno-Cabrera, Juan G. Cervantes-Rosado, Alejandro Meraz-Moreno, Lourdes Marquez-Morfin, Jorge A. Gomez-Valdes, Alejandra Castillo, Ernesto Garfias, Daniel Blanco-Melo, Maria C Avila-Arcos

Center for GeoGenetics, Fred Hutchinson Cancer Research Center (USA), International Laboratory for Human Genome Research, National Institute of Anthropology and History (Mexico), Natural History Museum of Denmark (Denmark), North Carolina State University (USA), UNAM (Mexico), Undergraduate program on Genomic Sciences, University of California (USA)

Proteins are a rich source of evolutionary information, capable of withstanding degradation better than nucleic acids. However, the use of paleoproteomics to identify and study ancient pathogens remains limited. Previous research in our lab identified the presence of ancient pathogen DNA in several samples from individuals buried in Mexico City dating to the 16th century, a period characterized by epidemics spread following the aftermath of the Spanish colonization. For this project, tryptic peptides were extracted from six of these samples using protocols optimized for ancient proteins. This included the implementation of a novel enzymatic digestion method for collagen. Subsequently, the tryptic peptides underwent LC-MS/MS analysis, and the resulting raw data files were processed using protein identification software, including MaxQuant, Protein Discoverer, and PEAKS. Besides the various peptides from environmental bacteria, we detected numerous peptides of pathogenic bacteria from the genus Salmonella, as well as several peptides from the oral pathogen Tanerella forsythia. Importantly, Salmonella and T. forsythia were identified in the same individuals who had previously tested positive for aDNA from these pathogens, further supporting their presence and revealing the possibility of recovering ancient pathogen peptides from archaeological remains. This exploratory study highlights the potential of paleoproteomics in ancient pathogen research and the synergistic benefits of a multidisciplinary approach to unravel the mysteries of ancient epidemics. Keywords: Paleoproteomics, Paleogenomics, pathogens, epidemics.
Reconstruction of one Salmonella enterica Paratyphi C genome from 19th-century Mexico City

Miriam Bravo-Lopez


Brigham and Women’s Hospital in the Pulmonary and Critical Care Medicine Division (USA), Brown University (USA), CNRS UMR 9015, Center for Anthropological Studies, Ecology, Evolution and Organismal Biology, Faculty of Political and Social Sciences, GLOBE Institute, INRIA, Laboratoire Interdisciplinaire des Sciences du Numérique (France), National Autonomous University of Mexico (CEA-FCPyS-UNAM) (Mexico), National School of Anthropology and History, Section for Evolutionary Genomics, UNAM (Mexico), Universidad Nacional Autónoma de México (UNAM) (Mexico), University of Copenhagen (Denmark), Université Paris-Saclay, Université Paris-Saclay (France), International Laboratory for Human Genome Research

The European Imperial expansion during the 19th century in Mexico, resulted in new infectious diseases due to ecological disruptions. To gain insights into the infectious agents introduced during this period, we generated paleogenomic data from seven individuals dated to the 19th century from Mexico City. Taxonomic classification of the obtained sequences enabled us to identify ancient DNA recovered from one individual as Salmonella enterica Paratyphi C. This pathogen causes paratyphoid fever in humans. Remarkably, S. Paratyphi C has been proposed as one of the causative agents of the 1545-1550 cocoliztli epidemic, resulting in an estimated 15 million deaths. To better understand the evolutionary history of S. Paratyphi C in ancient Mexico, we reconstructed its genome at a 10x depth using an in-house capture-enrichment strategy. The reconstructed genome was phylogenetically analyzed alongside ancient and modern genomes of S. Paratyphi C. Our genome falls within a branch closely related to the ancient S. Paratyphi C genomes from south Mexico. The divergence time between them was estimated to 500 years BP, which coincides with the Spanish conquest. Altogether, this study demonstrates that a replacement occurred within S. enterica Paratyphi C after the arrival of the Spanish conquistadors in New Spain.
The geographic origin, evolution and spread of treponemal diseases remains one of the most debated topics of infectious disease history. Treponema pallidum subspecies, the pathogens known to cause modern treponemal diseases that have been genomically characterized, are closely related yet cause various disease syndromes. Each subspecies was traditionally thought to be linked to its own distinct clinical manifestation and environmental context. However, recent genomics research on both modern and ancient Treponema pallidum subspecies has cast doubt on this assumption. Genomic reconstructions of Treponema pallidum subsp. endemicum from pre-European contexts in Brazil and Treponema pallidum subspecies from 15th-century Europe have provided new perspectives on the evolutionary history and phylogeographic patterns of these subspecies. In this study, we unveil a 5,500-year-old Treponema pallidum-like pathogen discovered in Sabana de Bogotá, Colombia. This work provides the oldest reconstructed Treponema genome to date and offers the first pre-Hispanic representation of Treponema in this region. Our contribution expands the existing genomic dataset of treponemal pathogens by millennia, significantly advancing our understanding of treponematosis in the past. Phylogenomic analysis of the 1.7-fold coverage genome indicates this pathogen is basal to all Treponema pallidum subspecies for which genomic data are available, including all published ancient strains. In conjunction with archaeological and paleopathological evidence, our findings contribute to discussions on the disease landscape of mid-Holocene populations of the Americas, treponemal evolution, and the geographic presence of treponemal disease in the past.
Deciphering the emergence and evolutionary history of a bacterial crop pathogen: insights from historical herbarium specimens

Paola Elvira Campos

Paola E. Campos, Olivier Pruvost, Karine Boyer, Myriam Gaudeul, Cláudia Baider, Timothy M.A. Utteridge, Nathalie Becker, Adrien Rieux, Lionel Gagnevin
CIRAD (France), CIRAD (Réunion), MNHN (France), MNHN (Réunion), Mauritius Herbarium (Mauritius), Royal Botanic Gardens (United Kingdom), UCD (Ireland)

Crop pathogens have been a threat to mankind since the beginnings of agriculture. In order to better understand current crop diseases and prevent future epidemics, it is essential to appreciate the factors underlying the emergence, adaptation and spread of pathogens. Recent methodological developments in molecular epidemiology now allow reconstructing disease dynamics in space and time. While the majority of studies previously carried out are entirely based on the sampling of contemporary specimens dating from the last decades, the advent of paleogenomics today makes possible the reconstruction of historical genomes and the study of pathogens evolutionary history with greater precision. Here, we reconstruct 13 historical genomes of the bacterial crop pathogen Xanthomonas citri pv. citri (Xci) from infected Citrus herbarium specimens. Following ancient DNA authentication, we compare them to modern genomes representative of the worldwide genetic diversity to estimate their phylogenetic relationships, pathogenicity-associated gene content and several evolutionary parameters on a global scale. Our results indicate that Xci originated in Southern Asia ~11,500 years ago (perhaps in relation to Neolithic climate change and the development of agriculture) and diversified during the beginning of the 13th century, after Citrus diversification and before spreading to the rest of the world (probably via human-driven expansion of citriculture through early East-West trade and colonisation). Our work emphasises the importance of historical data in the reconstruction of crop pathogens evolutionary history, valourising naturalist collections and generating knowledge bearing the potential of improving disease monitoring and sustainable control of current and future epidemics.
Depletion of the plasminogen activator (pla) virulence gene across pandemics of plague

Ravneet Kaur Sidhu

Yersinia pestis killed ~30-50% of Europe's population during its Black Death epidemic (1345-1353). This was followed by approximately 400 years of post-Black Death outbreaks throughout Afro-Eurasia. These persistent post-Black Death waves are believed to have had lower mortality rates, potentially due to virulence attenuation, host immunity, cultural adaptations, or synergistic interactions between them. Genomic studies of post-Black Death Y. pestis genomes have identified a depletion of the key virulence factor pla, on the high copy number PCP1 plasmid, attempting to link this genomic event with the disappearance of plague from Europe. In this work, we challenge the current perspectives of pla depletion and take steps towards characterizing its genomic and functional characteristics. We start by computationally screening ancient and modern Yersinia samples (n = 2,914), finding previously undescribed instances of this convergent depletion. We then perform targeted capture of pPCP1 on ancient Y. pestis samples from geographically and temporally distinct sites throughout Medieval Denmark, allowing de novo assembly of pPCP1 (pla -). Taking a well characterized pla depleted sample we sequence it to a depth of 400M to evaluate the genomic movement of pla in second pandemic strains. Finally, our work integrates functional data providing the first insights into the in-vitro and in-vivo impacts of pla depletion on Y. pestis virulence.
Characterizing the cariogenic bacterium Streptococcus mutans in ancient and modern Chile

Shreya L Ramachandran, Ricardo A Verdugo, Paola Gonzalez, Mauricio Moraga, Sergio Gonzalez Providell, Josefa Nuñez Belmar, Jaime Andrés Rivas Pardo, Constanza de la Fuente, Maanasa Raghavan

NA (USA), Universidad Mayor (Chile), Universidad de Chile (Chile), Universidad de Talca (Chile), University of Chicago (USA)

Here, I present preliminary data from tooth samples from eight individuals from an archaeological site in the Near North region in Chile, dating between 1200 and 1400 BCE. The individuals, representing different times and a variety of archaeological contexts, all show genomic evidence of ancient Streptococcus mutans, a bacterium responsible for causing cavities. I designed a custom capture panel to gain higher-coverage data of a set of evolutionarily and medically relevant S. mutans genes. Using the combined data from shotgun sequencing and the capture panel, I conducted phylogenetic analyses that use these ancient time-stamped samples to refine estimates of the evolutionary trajectory of S. mutans and its relationship to human agricultural transitions. I will also discuss the adaptation of S. mutans to its local environment, by analyzing the presence and absence of virulence factors in an archaeological context compared with contemporary S. mutans strains. In collaboration with archaeologists and anthropologists, I will study the samples in the context of the diet and lifestyle of these individuals. From these analyses, we can gain a greater understanding of determinants of oral health in ancient South America, a region that has been historically understudied from an ancient metagenomic perspective. At the same time, by studying the evolutionary history of pathogens that continue to affect us today, we can gain insight into potential targets for improving human health.
RNA virus genomes from historical specimens of great apes

Sojung Han, Christian Urban, Olivia Cheronet, Ron Pinhasi, Verena Schuenemann, Sébastien Calvignac-Spencer, Martin Kühlwilm
Department of Environmental Sciences, Department of Evolutionary Anthropology, Helmholtz Centre for Infection Research (Germany), Human Evolution and Archaeological Sciences (HEAS), Institute of Evolutionary Medicine, University of Basel (Switzerland), University of Vienna (Austria), University of Zurich (Switzerland)

Infectious diseases are one of the main drivers of evolutionary change, with RNA viruses such as measles, influenza and coronaviruses being a major threat. With their potential for spillover between species, understanding pathogen genomes across species is important. Pathogens also appear to have quite varying mutation rates, which highlight the necessity of sampling their genomes at different historical time points. Compared to ancient DNA virus genomes, ancient RNA virus genomes are much less studied, due to technical difficulties: RNA is less stable than DNA, and RNases are abundant in our environment, easily degrading RNA molecules. However, it has been shown that RNA virus genomes could be sequenced from lung tissue kept in formalin, and that RNA fragments could be retrieved from museum muscle and skin specimens older than a century. It is necessary to obtain knowledge on the types of historical specimens that may preserve sufficient amounts of RNA for sequencing, in particular from viruses. Here, we extract RNA from different types of historical great ape museum specimens, such as formalin-preserved tissue, teeth, dental calculus, skin and fixative. Using protocols for ancient nucleotides adapted for RNA, we determine differences in host RNA content and complexity across these sample types for a total of 100 specimens. Since RNA viruses are expected to be present, if at all, at low abundance, we perform enrichment capture for a range of lineages. This allows us to obtain insights into the historical RNA virome of great apes, and delineate limitations of specimen types.
Exploring oral microbial evolution in the context of European colonization of the Americas
Tanvi P. Honap

Presented by self
University of Oklahoma (USA)

The field of microbial paleogenomics has traditionally focused on pathogens that occupy public consciousness due to their historical significance and documented devastation, such as those causing plague, leprosy, and tuberculosis. Pathogens causing non-communicable diseases, such as oral diseases, have received less attention, despite these afflicting approximately 3.5 billion people today. Oral diseases are typically caused by dysbiosis in the oral microbiome resulting in the proliferation of opportunistically pathogenic species. Paleogenomic studies utilizing archaeological dental calculus have provided insights into the evolution of our oral microbiomes throughout human history. However, few studies have focused on reconstructing the evolutionary histories of individual oral microbes. Here, through an analysis of shotgun metagenomic data recovered from archaeological dental calculus samples, we present an investigation into oral microbial evolution in the context of European colonization of the Americas. We reconstructed the genomes of opportunistic pathogens such as Tannerella forsythia, commensal species such as Anaerolineaceae bacterium oral taxon 439, as well as several understudied oral microbes. Our phylogenomic analyses reveal diverse evolutionary trajectories for different microbial species within the same oral ecology and demonstrate that European colonization impacted the evolutionary histories of not only humans but also of their associated microbes. Overall, our findings provide insights into the complex interplay between human migrations and oral microbial evolution and have implications for understanding the dynamics of oral diseases across different historical contexts.
Metagenomic Analysis of Oral Pathogens in Mammuthus columbi Remains from Mexico

Victoria Pastor

Victoria Pastor Martinez Ulloa, Miriam J Bravo Lopez, Alejandra Castillo Carbajal, Ernesto Garfias Morales, Angeles Tavares Guzman, Viridiana Villa Islas, Alejandro Lopez Jimenez, Mashaal Sohail, Joaquin Arroyo-Cabales, Federico Sanchez Quinto, Maria C. Avila Arcos. International Laboratory for Human Genome Research, Laboratorio de Arqueozooología, 62220, Avenida Universidad s/n, Cuernavaca, Instituto Nacional de Antropología e Historia, International Laboratory for Human Genome Research, Laboratorio de Arqueozooología, Mexico City, Mexico. 3 Centro de Ciencias Genómicas, Morelos, México (Mexico), México. (Mexico), Querétaro, Subdirección de Laboratorios y Apoyo Académico, Universidad Nacional Autónoma de México, Universidad Nacional Autónoma de México (UNAM)

Mammuthus is an extinct genus of elephants that appeared in Africa about five million years ago. Mammuthus columbi, a species within this genus, was the only mammoth endemic to the Americas. Excavations near the Santa Lucia Military Air Base in Mexico have unearthed at least 80 Mammuthus columbi remains. These findings present an ideal opportunity to gain insight into the evolutionary and ecological dynamics of this extinct species. The study of the oral microbiome through paleogenomics offers a window into the past oral health of individuals, and it has not yet been fully explored in Mammoths. In this study, we characterized the oral metagenomic profiles of 72 mammoths found in Santa Lucia Military Air Base and Tultepec Mexico. Sequencing libraries were built from dentin samples; the resulting shotgun sequence data were used to profile ancient microorganisms\' DNA. The taxonomic classification revealed a significant presence of reads associated with Propionibacterium acidifaciens and Streptococcus mutans, oral pathogens implicated in dental caries lesions in humans and other animals, in three samples. These findings provide insights into mammoth health, contributing to our understanding of ancient mammalian biology. In our future research, we plan to perform a phylogenetic placement analysis to understand the relationship between these pathogen sequences and modern strains.
S15 - Paleogenomics and human evolutionary history: new insights and novel methods.
The impact of selection on human sex-specific genetic diversity: an ancient genomes perspective
Adamandia Kapopoulou

Stefan Strütt, Adamandia Kapopoulou, Stephan Peischl, Laurent Excoffier
CMPG, Institute for Ecology and Evolution, Interfaculty Bioinformatics Unit, Swiss Institute of Bioinformatics (Switzerland), University of Bern (Switzerland)

Differences in inferred demographic histories derived from human uniparental markers have frequently been linked to sex-specific transitions from migratory to sedimentary cultures during the Neolithic revolution. However, purifying selection against deleterious mutations can significantly distort genealogies at linked neutral sites. This poses a major challenge to accurately inferring demographic histories, particularly in regions with low or absent recombination, such as in human uniparental markers. To address whether the presence and extent of putatively strong selection in mitochondrial DNA and weak selection in Y-chromosomes could account for the observed differences in sex-specific inferred human demography, we pursued two primary objectives. Firstly, we extended a structured coalescent model for strong purifying selection in regions without recombination to allow arbitrarily weak or strong selection. Secondly, we curated a high-quality dataset of newly sequenced ancient Mesolithic and Neolithic genomes in Europe. To achieve this, we developed a customized pipeline for variant calling tailored to the unique challenges presented by ancient haploid chromosomes. Including our recently acquired ancient dataset and using our extended structured coalescent, we compare patterns of diversity in ancient and modern Y and mtDNA chromosomes to see when differences emerged and whether these differences can be attributed to the sedentarisation of male early farmers, as previously postulated.
It has been demonstrated that viruses act as strong selective agents in host populations, although much of this research is based on identifying adaptation in proteins (e.g., specialized antiviral proteins). However, it is thought that most adaptation in mammalian genomes occurs in sequences that regulate gene expression. In this study, we seek to address the disconnect between research in virus-driven adaptation and adaptation that occurs at the regulatory level. Bats are of particular interest in characterizing virus-induced selection, given their capacity to carry several families of zoonotic pathogens stemming from an extensive history of co-evolution alongside those pathogens. Here, we quantify virus-driven adaptation in gene expression regulatory sequences from both human genomes (as a proof-of-concept) and Myotis bat genomes. Because we do not have experimental annotations of regulatory sequences in the bat genomes, we use regulatory sequences identified by the neural network Enformer. We modify the Approximate Bayesian Computation extension of the McDonald-Kreitman test (ABC M-K) to apply it to regulatory regions; this test quantifies possible adaptive evolution. The ABC M-K test demonstrates possible signals of adaptive evolution in the DNA sequences that regulate the expression of virus-interacting proteins. These signals of regulatory adaptation allow us to use a greater portion of the genome to identify the pathogens that caused ancient epidemics than would be possible in just relying on coding sequences.
Evaluating allele frequency trajectory and selection coefficient estimates from genealogies including ancient DNA
Aina Colomer i Vilaplana

Aina Colomer-Vilaplana, Antonio Barbadilla, Sònia Casillas, Leo Speidel
Department of Genetics and Microbiology, Francis Crick Institute (United Kingdom), Genetics Institute, Institute of Biotechnology and Biomedicine, UAB (Spain), University College de Londres (United Kingdom)

During the dispersal across continents, humans have faced vastly different environments, pathogenic exposure, and technological innovations. Yet, a question still unsolved is the extent to which selection has played a role in shaping our genomes across different time periods of human evolution. The sequencing of increasingly large ancient DNA cohorts from single populations is now starting to make selection acting on our genomes over time directly observable, with several recent techniques enabling the inference of allele frequency trajectories and the associated selection coefficients. However ancient DNA preservation is typically best in colder climates and becomes progressively more difficult to obtain in the deeper past. These dense ancient DNA time series are therefore still only available in limited settings. Recently, new methods for inferring genealogies jointly for modern and ancient genomes - such as Relate, tsinfer, and ARGneedle - have made it possible to fully leverage haplotype information for the inference of selection, including to infer allele frequency trajectories and selection coefficients. This presentation will explore the benefits of combining genealogical and ancient DNA data to observe genomic adaptation through time. We present a framework to simulate genomes undergoing positive selection that allows the sampling of ancient DNA. We benchmark several new techniques for inferring selection on this data, with and without genealogies and with varying densities of ancient genomes, under different selection regimes, showing that genealogy inference using ancient DNA allows us to extrapolate beyond time periods where aDNA is readily available and improves selection estimates.
The Neolithic and Bronze Age (BA) periods in the Aegean are characterized by genetic transitions, extensive population interactions, and migrations. Questions pertaining to the Neolithisation process, the contribution of local hunter-gatherer (HG) populations and the extent and timing of a Steppe-like migration during the BA, are still unanswered. To characterize the population structure across time, we generated whole-genome sequence data for 14 Neolithic to Late BA individuals from the Aegean and Anatolia, complemented them with published ancient genomes, and conducted population genetic analyses including admixture modeling, f-statistics and IBD analyses using both pseudo-haploid and imputed data. By examining the effect of imputation on allele frequencies and Linkage Disequilibrium (LD), we observe that imputation distorts the LD decay and we discuss the potential impact on downstream analyses. Our results show that the Neolithic Aegean population was predominantly derived from Western Anatolian farmers, confirming previous studies and suggesting a genetic turnover from the Mesolithic. By the Late Neolithic and onwards there was an influx of a Caucasus HG/Iran Neolithic-like component, possibly introduced by migrants from the east. During the BA, population heterogeneity in the Aegean increased. An Eastern HG-related component already appeared during the Early BA, suggesting that a Steppe-like ancestry may have reached the Aegean earlier than previously described. The EHG-related component is maximized in Middle BA Aegean populations, but compared to other European BA populations its proportion remains low, suggesting that Steppe-like ancestry reached Greece later than mainland Europe.
Elucidating the history of the European crow hybrid zone with paleogenomics
Chyi Yin Gwee

Climatic oscillations influence the population dynamics of closely-related lineages. During the glacial period populations are forced into disparate refugia from which they emerge following the retreat of ice sheets. This results in secondary contact of gene pools that have experienced a period of reproductive isolation. All-black carrion and grey-coated hooded crows meeting along a narrow hybrid zone in central Europe are a textbook example. The hybrid zone is shared by distinct vertebrate groups, thus this contact zone is hypothesized to be shaped by the repeated events of isolation and recolonization during the glacial-interglacial cycles. This biogeographic model is inferred based on genetic signatures of remnant population samples, but accurate interpretation of the demography and historical admixture events are hindered by the presence of ghost population(s) and/or anthropogenic activities, which may directly or indirectly influence migration patterns. Using the carrion and hooded crows C. corone as a model, our study aims to directly reconstruct the migration route of a widespread European avian lineage after the Last Glacial Maximum with ancient genomic samples dating as far back as 20,000 years ago. The present study uses genomic data from 20,000 to 100 cal BP to reconstruct the formation of the hybrid zone and the movement of phenotypic genes through time and space. The data lends itself to study the coexistence of crows and humans by comparing their migration pathways as crows are synanthropes and have been culturally important to different civilization through time.
Overcoming bias and postmortem damage to improve the accuracy of ancient genome analysis

Dilek Koptekin

Bilkent University (Turkey), Department of Biological Sciences, Department of Biology, Department of Computational Biology, Department of Computer Engineering, Department of Health Informatics, Graduate School of Informatics, Lund University (Sweden), Middle East Technical University (Turkey), Swiss Institute of Bioinformatics, University of Lausanne (Switzerland)

Ancient DNA (aDNA) contains invaluable information about population history and evolution, but its analysis is plagued by challenges such as reference bias and postmortem damage (PMD). Here, we propose new strategies to address these challenges. First, we show that masking variable genomic positions before alignment or using a graph genome both effectively reduce reference bias in ancient human genome data. However, since quality-filtered BAM files and 1240K capture data retain bias, this solution depends on data availability in FASTQ or unfiltered BAM formats. Second, we evaluate various strategies for removing PMD, including trimming, rescaling base quality, and a new algorithm, bamRefine, that masks polymorphic regions vulnerable to PMD. Compared to trimming and rescaling, bamRefine significantly improves genotyped loci by up to 20% while increasing accuracy. Overall, we advocate for the adoption of graph alignment in conjunction with bamRefine to minimize data loss and bias in aDNA research. Furthermore, our study highlights the importance of publishing FASTQ files for comprehensive analysis. By addressing these technical challenges, our work increases the reliability and depth of information obtained from ancient DNA, which in turn will improve our inference of population history and evolution.
The extreme scarcity of human remains from the Palaeolithic period means the demographic processes which shaped the genomic makeup of Europe after the end of the Last Glacial Maximum (LGM) are still largely unknown. Several distinct genetic ancestries have been reported; the first associated with individuals from Solutrean and Magdalenian archaeological contexts, and the second, named for the Villabruna individual, represents the genetic ancestry which formed the majority ancestry fraction of later European Mesolithic individuals. To investigate the processes by which these major genetic lineages of post-glacial Europe migrated and admixed, we generated six new shotgun sequenced genomes spanning 10,000 years, from the end of the LGM to the beginning of the Holocene. These genomes originate from the south-west of France, a key region in decoding LGM population demography as it may have been continuously occupied since the Aurignacian period. Through the co-analysis of these genomes with existing genetic data, using imputation and cutting-edge haplotype-based methods, we present a framework for population movements and contact during the late Upper Palaeolithic.
Deep Reconstruction of the Migration History of the World
Eran Elhaik

Eran Elhaik
Lund University (Sweden)

Reconstructing the history of the world is one of the ultimate goals of paleogenomics, yet thus far, only a dozen major migration routes have been identified or speculated. Geographical inference based on anatomical or morphological information is highly complex and error-prone, particularly when the remains are physically damaged or fragmented. Utilizing ancient DNA for localization presents additional challenges due to the lack of intermediate samples over space or time, the small number of SNPs, and their spurious nature. Poorly dated samples pose an even greater challenge to the accuracy of the reconstruction. We developed ancient GPS (aGPS), which implements a dual machine-learning approach. Firstly, to predict the age of the ancient samples, and secondly, to predict their geographical site of origin. We demonstrate that aGPS age predictions align with known dates. We also show that aGPS accurately predicts the burial site of the ancient samples, comparable to when applied to modern samples. When set to identify ancient origins, aGPS maps historical migration routes, thereby unraveling the journey of humankind as far as aDNA data permit. aGPS is not limited to humans and enables the resolution of questions regarding origins and population movements. Addressing long-standing historical questions concerning the identity of Old World residents, aGPS stands as a powerful tool in the field of paleogenomics.
Grave Matters: Discerning Ancient DNA Profiles from Grave Dirt vs. Skeletal Remains

Gözde Ata?

Gözde Ata?, Benjamin Vernot
Max Planck Institute for Evolutionary Anthropology (Germany)

Sedimentary ancient DNA (sedaDNA) has become a valuable tool for studying past populations in the absence of human remains. Here, we explore the use of sedaDNA as an alternative to ancient DNA from skeletal elements. We collected skeletal material, along with grave dirt from the skeletal surfaces, at the Late Bronze Age site of Didnauri (Georgia) and the Medieval site Alt-Inden (Germany). We generated mitochondrial and nuclear genome data using target enrichment, and analyzed the differences in sequence profiles from sediments versus bones. First, we compare the qualities, coverage distributions, variant properties and haplogroup inferences. Then, we test for the presence of non-human DNA in grave dirt, and evaluate the identified human DNA for sources other than the individual to which the skeletal material belongs. Further, we explore the differences between the two types of data in population genetics analyses, particularly assessing the relative affinities to the adjacent populations, and the consequent clustering patterns. Overall, by investigating ancient DNA sequences derived from the two sources, our study seeks to elucidate whether a comparable accuracy can be achieved in genetic studies by sampling sediments rather than the skeletal material itself.
Reconstructing the genetic relationship between ancient and present-day Siberian populations
Haechan Gill

Haechan Gill, Juhyeon Lee, Choongwon Jeong
Republic of), Seoul National University (Korea

Human populations across a vast area in northern Eurasia, from Fennoscandia to Chukotka, share a distinct genetic component often referred to as the Siberian ancestry. Most enriched in present-day Samoyedic-speaking populations such as Nganasans, its origins and history still remain elusive despite the growing list of ancient and present-day genomes from Siberia. Here we reanalyze published ancient and present-day Siberian genomes focusing on the Baikal and Yakutia, resolving key questions regarding their genetic history. First, we show a long-term presence of a unique genetic profile in southern Siberia, up to 6,000 years ago, which distinctly shares a deep ancestral connection with Native Americans. Second, we provide plausible historical models tracing genetic changes in West Baikal and Yakutia in fine resolution. Third, the Middle Neolithic individual from Yakutia, belonging to the Belkachi culture, serves as the best source so far available for the spread of the Siberian ancestry into Fennoscandia and Greenland. These findings shed light on the genetic legacy of the Siberian ancestry and provide insights into the complex interplay between different populations in northern Eurasia throughout history.
Steppe Ancestry in western Eurasia and the spread of the Germanic Languages
Hugh McColl

According to linguistic consensus, the common ancestor of the Germanic languages, including German, English, Frisian, Dutch and the Nordic languages, arrived with the migrations of Bronze Age Steppe Pastoralists into Europe around 5,000 years ago. In much of Europe, there has also been an assumption of genetic continuity since this last major migration, with limited genetic impacts from the surrounding regions. However, the ability to detect later migrations within Europe is hindered by the close relation between populations and limited ancient samples. We sequenced 710 ancient human genomes and analysed them together with 3,940 published genomes suitable for imputing diploid genotypes. By analysing recent coancestry through shared IBD segments we find a series of new clines associated with migrations within Europe. From the Bronze Age onwards, we find two broad clusters within northern and western Europe corresponding with Corded Ware and Bell Beaker cultures. Within the former, we find evidence of a cross-Baltic migration into Eastern Scandinavia that later spread west and south, arriving 800 years later but consistent with the assumed paleo-Germanic range. Later, during the migration period, we identify a region around Southern Jutland / Northern Germany as the source for migrations west to the British Isles and north into the Danish Isles and Southern Sweden, corresponding with the spread of West and North Germanic languages respectively. This work demonstrates the feasibility of detecting migrations within closely-related populations, and reveals a series of archaeologically invisible migrations of archaeological, historic and linguistic significance.
Unlocking the Past: Ancient Protein Analysis Sheds Light on Early Human Evolution in Southern Africa

Imke Lankheet

Imke Lankheet, Carina Schlebusch
University of Johannesburg (South Africa), Uppsala University (Sweden)

Although aDNA has significantly improved our understanding of human evolution and human history, it also has its limitations. One of the biggest limitations is the survival and preservation of DNA. Proteins generally preserve longer than DNA and harbor crucial biological information, particularly relevant for early human evolution in Africa. Through the analysis of in-silico translated genomic data and ancient protein sequences, our study identifies several amino acid differences in bone-related proteins between southern African Khoe-San groups and the broader human population. In conclusion, this study marks an important step towards advancing the field of population genetics through the utilisation of ancient proteins. We demonstrate the potential of ancient protein analysis in unraveling the complexities of early human evolution. These findings not only underscore the importance of considering proteins alongside DNA in ancient genomics but also pave the way for future studies examining older samples. Ultimately, this research lays the foundation for a new era in population genetics, where ancient proteins offer a unique perspective on population continuity and ancient population structure in sub-Saharan Africa.
Detecting introgressed archaic haplotypes in ancient human genomes

Jiaqi Yang

Jiaqi Yang, Stéphane Peyrégne, Janet Kelso, Benjamin Peter
Max Planck Institute for Evolutionary Anthropology (Germany)

Interbreeding with archaic humans introduced Neanderthal DNA to all non-African populations and Denisovan DNA to Asian, American, and Oceanian populations. Investigating archaic ancestry in ancient genomes makes it possible to detect selection on introgressed haplotypes and study the dispersal and adaptation of both archaic and modern humans. However, applying haplotype-based methods to ancient genomes has been challenging, due to the poor data quality of these genomes. Here we compare two methods: 1. Admixfrog, a reference-based method that works with genotype likelihoods and requires high-quality sequenced archaic genomes as sources of introgression; and 2. Hmmix, a reference-free method, which takes genotypes from the target genome as input and needs an unadmixed outgroup to detect haplotypes introgressed from divergent lineages. We first show how we can adapt the reference-free method to medium-to-low-coverage datasets without imputation. Based on both simulations and empirical data, we find that the reference-free method detects more introgressed haplotypes than the reference-based method when the sequenced archaic genomes are divergent from the introgressing populations, as observed for Denisovan haplotypes in present-day human genomes. However, the reference-based method remains the only solution when the coverage of the target genome is ultra-low. We then investigate Denisovan haplotypes in ancient genomes and show how Denisovan ancestry is distributed in modern humans through space and time.
Exploring the genetic diversity in the Americas through ancient whole genomes

Judith Ballesteros Villascán

Judith Ballesteros Villascán, Kay Prüfer, Kathrin Nägele
Max Planck Institute for Evolutionary Anthropology (Germany)

Whereas considerable strides have been made to characterize the present-day genetic diversity of the Americas, much less data is available for individuals who lived in the past. Fortunately, whole ancient genomes have been produced for approximately 150 individuals to investigate some rudimentary aspects of the human genetic history of the Americas. But whole genomes are rare since DNA is often not well preserved. To overcome this limitation, researchers have employed targeted enrichment methods. These allow for poorly preserved specimens to be analyzed. However, the design of some of the current oligonucleotide probes for enrichment was based on data available from two present-day American populations; therefore, it is possible that this does not represent the complete genetic variation of past American populations well. In this project, we analyzed the published whole genomes from 30 ancient individuals from the Americas to test how well the enrichment reflects ancient diversity by current tools. We examined how much of the past genetic diversity would be retained after enrichment and found that it can introduce biases that, in some cases, can lead to wrong conclusions. Based on these insights, we propose an alternative strategy that will allow us to gain a more complete understanding of the Americas’ genetic history, particularly in the tropical regions where DNA is poorly preserved.
Medieval genomes from eastern Mongolia share a stable genetic profile over a millennium

Juhyeon Lee

Juhyeon Lee, Takehiro Sato, Atsushi Tajima, Tsend Amgalantugs, Batmunkh Batmunkh, Shigeki Nakagome, Toshihiko Miyake, Noriyuki Shiraishi, Choongwon Jeong, Takashi Gakuhari
College of Human and Social Sciences, College of Humanities, Faculty of Humanities, Faculty of Medicine, Graduate School of Medicine, Institute for the study of Ancient Civilization and Cultural Resources, Institute of Archaeology, Kanazawa University (Japan), Mongolian Academy of Sciences (Mongolia), Niigata University (Japan), Republic of, School of Biological Sciences, School of History, School of Medicine, Seoul National University (Korea, Shukutoku University (Japan), Trinity College Dublin (Ireland), University of the Ryukyus (Japan)

Recent archaeogenomic studies in Mongolia have elucidated the genetic origins of people from the Xiongnu and Mongol eras, but left the Medieval period between them only tangentially explored. Due to this dearth of ancient genomes, the dynamic history of Medieval Mongolia with the rise and fall of numerous polities still lacks a genomic perspective. To fill in this knowledge gap, here we report whole-genome sequences of nine ancient individuals from eastern Mongolia, who were excavated from two nearby cemeteries, Gurvan Dov and Tavan Khailaast. They are distributed from the Xiongnu-Xianbei period (ca. 200 CE) to the Mongol era (ca. 1,400 CE), forming a local time transect encompassing nearly 1,200 years. Remarkably, despite the long-time span, all nine individuals derive most of their ancestry (85–100%) from the eastern Eurasian lineages and show low heterogeneity in their genetic composition. This is in contrast to the general pattern observed in previously published Medieval genomes from central Mongolia, who showed higher heterogeneity and overall less eastern Eurasian ancestry, thus calling for a comprehensive archaeogenetic survey of Medieval Mongolia to fully capture the dynamic genetic history in this period.
Ancient Genomes of Neandertals and Modern Humans from Europe
Kay Prüfer

Presented by self
Max Planck Institute for the Science of Human History (Germany)

Neandertals lived for hundreds of thousands of years in Europe until they disappeared from the fossil record, around 40,000 years ago. The analysis of Neandertal genomes revealed that they had a comparatively small population size and that modern humans met and interbred with Neandertals after the out-of-Africa event. This admixture event resulted in typically 2-3% Neandertal ancestry in the genomes of the descendants of this out-of-African population. Few modern human genomes are available from before 40,000 years ago when both Neandertals and modern humans inhabited Europe. In my talk, I'll present the first high-coverage modern human genomes from Europe that date to around 45,000 years ago. The analysis of these genomes yields further insights into the timing of the Neandertal admixture and the relationship of these first modern humans in Europe to contemporaneous and present-day modern human groups.
Reconstructing the history of archaic introgression in modern humans: Insights from whole genome sequences of worldwide populations

Laurits Skov

Laurits Skov, Priya Moorjani
UC Berkeley (USA)

Analysis of archaic and modern human genomes has revealed evidence of gene flow from archaic hominins—Neanderthal and Denisovans—into modern humans and highlighted its critical role in shaping the genetic and phenotypic variation in modern humans. However, most studies have focused on Europeans and East Asians, with very few genomes from other parts of the world, leaving our understanding of human evolutionary history incomplete. In this study, we integrate data from four different studies including ~31,000 previously published genomes and ~2700 newly sequenced genomes from South Asia. Using present-day genomes, we recover 1.7 Gb of Neanderthal genome and 1.1 Gb of Denisovan genome hidden in modern human genomes. This is the largest fraction of archaic ancestry recovered. By comparing the distribution of archaic segments across worldwide populations, we infer that the largest variation in Neanderthal ancestry is observed in South Asia. For Denisovan ancestry, we infer a large fraction of population-specific variation in Oceanians and South Asians. Finally, we develop a new framework for inferring the timing of archaic admixture using the mutation clock that leverages the new mutations that arose on the introgressed segments since the admixture. This new approach allows us to date Neanderthal admixture to around 50,000 years ago, consistent with the dates from previous studies. Together, these analyses provide a comprehensive picture of the archaic admixture in modern humans and the role of archaic ancestry in our evolution.
Genetic affinities and social structure analysis of individuals found at the Sa Galera Sanctuary inferred from paleogenomic data

Leonardo Yair Correa Mendoza

Leonardo Yair Correa Mendoza, Jose Jorge Arguello Menendez, Ramon Martin Gordon, Anders Gotherstrom, Cristina Valdiosera, Federico Andres Sanchez Quinto

International Human Genome Research Laboratory - UNAM (Mexico), Asociación Amics de na Galera (Spain), Center of Genomic Sciences (Mexico), Department of Archaeology and Classical Studies (Sweden), Department of History - University of Burgos (Spain)

Recent paleogenomic studies have investigated the demographic history and social structure dynamics across time and space in Europe. The Bronze Age stands out as a time period with significant cultural shifts and human mobility. In example, interesting findings have been obtained investigating this epoch across the Mediterranean. However, during this time period some regions remain unexplored. The sanctuary of Sa Galera is an archeological site on an islet of the Bay of Palma de Mallorca (Spain), archaeologically dated between 200 BCE and 100 CE. At this site, eight individuals have been found, as well as ceramic artifacts associated with a Phoenician archeological context. In this project, we extracted, sequenced and analyzed DNA from the Sa Galera individuals to investigate their social structure and demographic origin. We observed deamination patterns and contamination estimates consistent with an endogenous origin of the data. Regarding their social structure, we characterize the uniparental marker lineages and investigate kinship relations among Sa Galera individuals using autosomic data. Aiming to assess their genetic affinities to other ancient and present-day reference Mediterranean populations, we performed exploratory and hypothesis-driven demographic analyses. Results suggest kin connections among Sa Galera individuals and supported by archeological data. Moreover, uniparental and autosomic demographic analyses suggest at least three different demographic origins for the individuals at this site. Our study provides further evidence of the high mobility of individuals during the Bronze Age, where individuals which are almost contemporary to each other and buried on the same islet, had a different demographic history.
Optimising imputation and IBD segment retrieval in ancient genomes

Linda Ongaro

Linda Ongaro, Emily M Breslin, Marco Rosario Capodiferro, Emilia Huerta-Sanchez, Lara M Cassidy
Brown University (USA), Ecology and Evolutionary Biology and Center for Computational and Molecular Biology, Smurfit Institute of Genetics, Trinity College Dublin (Ireland)

Ancient DNA has revolutionised the study of genetic variation and population movements over time. However, significant gaps in ancient genetic data, alongside geographic and temporal variations, hinder fine-scale characterisation of genomic structure and relatedness between individuals and/or archaeological sites. In this study, we implemented the imputation pipeline proposed by Hui et al. (2020), involving a two-step approach: (1) Conducting genotype imputation on individual ancient samples using the GLIMPSE software with the 1000 Genomes Project reference panel. (2) Constructing a multi-individual dataset of confident genotype calls from GLIMPSE for input into Beagle5, enabling a second round of imputation and phasing to improve genotype accuracy and minimise missing data. To benchmark our imputation pipeline, we downsampled high-coverage ancient genomes, including shotgun sequences and 1240k target enrichment data. We investigated each step of the pipeline testing different filter thresholds, including genotype posterior probability after the imputation steps. Besides the missingness, as heterozygote sites pose challenges for imputation, we prioritise heterozygote sensitivity as our primary metric for assessing imputation quality. Our findings highlight that a second round of imputation enhances the quality of imputed genotypes for both shotgun sequences and SNP capture samples. Moreover, we evaluate the performance of the second imputation with the identification of Identity-By-Descent (IBD) shared fragments between individuals using refinedIBD software, including long fragments, by examining known parent-offspring pairs. In conclusion, this study sheds light on improving ancient DNA analysis methodologies offering insights into population genetics and evolutionary dynamics.
A method for reconstructing ancient DNA methylation of additional archaic samples
Liran Carmel

Krystal White, Peleg Berkowicz, Eran Meshorer, Liran Carmel
The Hebrew University of Jerusalem (Israel)

Reconstructed DNA methylation maps of ancient samples serve as a proxy for gene activity patterns and provide a valuable tool for studying recent human evolution. Reconstruction of the premortem methylomes of a single Neanderthal, a single Denisovan, and a number of anatomically modern humans identified thousands of Differentially Methylated Regions (DMRs) separating these human groups. In addition to the limited availability of archaic genomes, the small number of archaic genomes was a result of the fact that the current DNA methylation reconstruction method is constrained to high-coverage genomes that have been treated with uracil DNA glycosylase and endonuclease VII (USER) to remove uracils prior to sequencing, a treatment that artificially creates large differences in deamination rates between methylated and unmethylated cytosines. However, this procedure was followed for only one of the three publicly available high-coverage Neanderthal genomes. Here, we develop methods to reconstruct DNA methylation from nonUSER-treated genomes by leveraging a difference in postmortem deamination rates between methylated and unmethylated cytosines. We show that the resulting DNA methylation maps of two additional Neanderthals capture meaningful biological information and can be used to identify lineage-specific DMRs. This triples the number of available Neanderthal methylomes, allowing for discrimination between DMRs that are specific to a single individual and DMRs that are likely to be common to all Neanderthals. We identify and discuss several interesting DMRs.
Evaluating the bias affecting population genetics analyses when co-analyzing Whole Genome Shotgun sequencing and 1240K capture genomes

Lucas Anchieri

Lucas Anchieri, Samuel Neuenschwander, Anna-Sapfo Malaspinas, Jazmín Ramos-Madrigal
Faculty of Health and Medical Sciences, Globe Institute, Swiss Institute of Bioinformatics (Switzerland), University of Copenhagen (Denmark), University of Lausanne (Switzerland), Vital-IT

Over the past decade, the amount of publicly available ancient human genome-wide data has increased exponentially. Most of this data has been generated either by whole genome shotgun sequencing (WGS) or in-solution enrichment targeting ~1,24 million SNPs (1240K capture). Previous work has suggested that different sequencing methods may introduce bias into downstream population genetic analyses. Yet, the extent and the nature of this bias remain to be thoroughly investigated and WGS and 1240K are routinely analyzed together in ancient genomic studies. Here, we compiled a dataset consisting of publicly available genome-wide data including WGS and 1240K capture data (with some individuals sequenced by both strategies) resembling a reference panel used in an ancient DNA study. We carry out standard population genetics analyses such as MDS, admixture clustering, and f-statistics, showing the potential biases that can arise from having a mix of WGS and 1240K data in state-of-the-art population genetic analysis. Preliminary results show clustering analyses such as MDS and Admixture are substantially affected by having a mixture of data from different sources, with individuals sequenced similarly appearing more related, and clustering together on early dimensions of an MDS. Our results suggest that this bias could have important implications for the interpretation of ancient DNA data. We also show how some precautions, such as restricting the analyses to a set of SNPs as suggested by Rohland et al. (2022), can reduce the impact of this bias.
Human genetic histories at the Himalayan frontiers
Maanasa Raghavan

Presented by self
University of Chicago

The Himalayas have played a critical role in shaping human biocultural diversity at the crossroads of Central, South, and East Asia. Studies of present-day Himalayan populations have revealed diverse genetic ancestries in the region, suggesting complex and heterogenous patterns of past human migrations and admixture. To characterize the evolving dynamics of these human interactions, we generated and analyzed genomic data spanning the last ~1,300 and ~3,300 years from locations close to the northern and southeastern frontiers of the Indian Himalayas, respectively. At sampled high-elevation ancient and present-day sites near the northern end of the mountain range in the Trans-Himalayas, we observed contributions from Tibetan as well as Central and South Asian-related genetic sources. While a Tibetan component was seen consistently over time, the occurrence of the latter two sources was more dynamic. At least some of these admixture events may have been facilitated by ancient trade and cultural networks across the Trans-Himalayan region. In contrast, at the southeastern foothills of the Himalayas, we predominantly found genetic ancestries related to ancient lowland East Asian sources. The exact sources and their proportions varied over time and may have been associated with agricultural expansions. While rough terrains have limited the scope of archaeological investigations in both study locations, paleogenomic insights from this collaborative research between archaeologists, geneticists, and local communities offer new perspectives on human movements and contacts near the northern and southeastern Himalayan peripheries.
A little over a decade ago, studies showed that when Anatomically Modern Humans expanded into Eurasia, they interbred with Archaic Humans (Neanderthal and Denisovan). This discovery redefined the origin of humans and opened new research directions to study the impact of archaic introgression in human evolution. Up to now, the majority of archaic introgression studies have focused on contemporary individual genomic data, revealing that non-African populations harbour varying levels of Neanderthal and Denisovan ancestry. Despite recent advances in identifying and quantifying archaic introgression in humans, little is known about the evolution of archaic variants within modern human populations after the introgression event (40,000 years ago). Archaic variation continued to evolve within humans and was likely shaped by a population’s unique demographic history as well as natural selection. While ancient DNA offers the potential to address this scientific gap, its poor quality has hindered the exploration of ancient genomes. Here, we investigate the feasibility of using imputation to enhance global and local archaic ancestry inference in ancient genomes. We downsampled 20 high-coverage genomes, representing individuals from diverse temporal and geographical contexts, to 0.0625X, 0.5X, 1X and 2X. Following the imputation of downsampled genomes, we compared them with high-coverage genomes to assess global and local archaic ancestry inference, using also archaic reference genomes to ascertain the inferred introgressed segments origin. We unveil the significant capacity to infer archaic ancestry in imputed ancient genomes. Notably, local ancestry inference in imputed genomes outperforms that of the original high-coverage genomes.
Quantifying differential host adaptation during the viral life cycle

Mary Reed-Weston

Mary R Reed-Weston, David Enard
University of Arizona (USA)

Due to their notorious ability to asymptptomatically host highly virulent pathogens, bats are an intriguing study system to examine adaptations that contribute to viral tolerance. Bats have unique life history traits, such as their longevity, high metabolism, and ability to fly, which have been suspected to contribute to their high viral tolerance. While some recent studies have shown that bats lack certain genes related to immunity and inflammation, molecular mechanisms underlying bat viral tolerance are still unknown. Previous research has also made it clear that there is significant adaptation in host genomes in response to viral infection, but little is known about what host proteins related to specific steps of viral infection are under the strongest selection. Therefore, we intend to characterize adaptation by looking at viral-interacting proteins (VIPs) at different stages of viral infection (e.g. entry into the cell, intracellular transport, viral release from the cell). Using human data as proof of concept, we quantified adaptation using an extension of the Mcdonald-Kreitman test. We identified the viral release category specifically as a hotspot of adaptation in humans, as well as data indicating a broad range of adaptation patterns across the different categories of host proteins. We will conduct the same analysis in Myotis bats, where we expect higher power to quantify adaptation due to significantly higher levels of genetic variation than in humans. This will ultimately reveal what viruses bats can and cannot tolerate, and the ways in which they adapt in response to viral infection.
Preservation of ancient DNA in archaeological sediment from Stone Age Mongolia

Michael James Boyle

Michael James Boyle, Arina M Khatsenovich, Byambaa Gunchinsuren, Evgeny P Rybin, Dashzeveg Bazargur, Tserendagva Yadmaa, Bolorbat Tsedendorj, John W Olsen, Diyendo Massilani

Department of Evolutionary Genetics (Germany), Department of Genetics (USA), Institute of Archaeology, Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences (Russian Federation), Max Planck Institute for Evolutionary Anthropology, Mongolian academy of Sciences (Mongolia), School of Anthropology (USA), University of Arizona, Yale University School of Medicine

Reconstructing the history of human occupation in Eastern Eurasia is hindered by gaps in the fossil record and unreliable dates. How early modern humans colonized this region, and to what extent they interacted with local archaic human groups such as the Denisovans, are unresolved questions that can be investigated through the retrieval of ancient DNA from archaic and early modern humans in Eastern Eurasia. However, the scarcity of human fossils from this region hinders ancient DNA analyses. To overcome this challenge, we developed an approach utilizing sediment from archeological sites with evidence of human occupation as a source of ancient DNA. We collected sediment samples from four archaeological sites in Mongolia: Tolbor-4, Tolbor-21, Tsagaan Agui cave, and Khargany Gol 5. We investigated ancient DNA preservation in these samples with shotgun sequencing and by capturing ancient mitochondrial DNA from the most abundant mammalian families found in Eurasia during the Stone Age. We found evidence of ancient DNA preservation in the sediment of Tolbor-4, Tolbor-21 and Tsagaan Agui cave. These results suggest that ancient environmental DNA approaches and higher density sampling of sediment in archaeological sites in Mongolia has the potential to provide insights into the genomics of past human populations in Eastern Eurasia.
Evaluating the applicability of imputation and kinship analyses for ancient sedimentary DNA datasets

Pnina Cohen

Pnina Cohen, Sarah Johnson, Elena I. Zavala, Priya Moorjany, Viviane Slon
Dan David Tel Aviv University (Israel), Tel Aviv University (Israel), University of California (USA)

Ancient DNA (aDNA) analyses— the study of genetic material from individuals that died hundreds or thousands of years ago— have revolutionized the research in human evolutionary genetics. At most archaeological sites dated to the Middle or Late Pleistocene (780,000-12,000 years ago), no human remains have been found. However, recent studies have shown that aDNA can be recovered from archaeological sediments, even in the absence of any skeletal remains, providing an exciting new avenue to learn about our evolutionary past. Despite that, so far only a limited number of studies have successfully recovered, identified and authenticated ancient human DNA from prehistoric sediments. This is due to the complexity of ancient sedimentary DNA datasets, which tend to be composed of small quantities of short and degraded fragments, often representing multiple species and multiple individuals, within an overwhelming background of environmental DNA. We focus on advancing and developing analytical frameworks to support the study of ancient human DNA from sediments. Here, we evaluate current methods for imputation, phasing and kinship analyses of genetic variants such as Glimpse, ancIBD, Beagle and LinkImpute - testing them for accuracy on both empirical data and on simulated datasets mimicking the characteristics of ancient sedimentary DNA, including very low genetic coverage, restricted relevant reference panels, mixture of individuals within a single sample, and the presence of non-human ancient DNA. Based on this, we propose appropriate filters and parameters for the tested software, when applied to the study of ancient human DNA from sedimentary samples.
The study of ancient DNA (aDNA) has revolutionized our understanding of past civilizations, providing unprecedented insights into the genetic makeup of ancient populations. Within this realm, the exploration of viral communities preserved in coprolites and dental calculus presents a unique opportunity to delve into the microbial ecosystems of previous eras. In these ecosystems viruses have emerged as crucial players, yet their study remains challenging due to degradation and contamination risks. These ancient viromes offer valuable insights into the diversity, evolution, and interactions of viruses in past populations. Moreover, they provide a unique lens through which to explore the coevolutionary relationships between viruses and their hosts. In this study we try to assess the importance of the choice of methods used for identification of viruses within aDNA metagenomes. How each methods can bring information at different level (reads, contigs, vMAGs), but also can miss some other informations.
Tracing admixture in European early farmers using local ancestry inference
Sandra Oliveira

Sandra Oliveira, Nina Marchi, Adamandia Kapopoulou, Laurent Excoffier
CNRS-MNHN-Université Paris Cité (France), Institute for Ecology and Evolution, Swiss Institute of Bioinformatics (Switzerland), UMR7206
Eco-anthropologie, University of Bern (Switzerland)

In Europe, the transition from a hunting and gathering lifestyle to food-production and sedentism unfolded with the expansion of people and culture from Anatolia. The ancestry of the first European farmers, who spread along the Mediterranean or Danubian routes, can be traced back to the Aegean basin, but recent demographic reconstructions show that these early farmers already carried substantial amounts of Western hunter-gatherer (WHG) ancestry. Genomic evidence thus suggests connectivity between Western Europe and Southwest Asia, and potentially early admixture in Anatolia or the Near East. In addition, Neolithic farmers further admixed with local WHG-related groups when expanding to Europe. However, the frequency, timing, and outcomes of these admixture events are still unclear.

Here, we use local ancestry inference to investigate these complex admixture processes and their genomic signatures in early Neolithic farmers spreading from Anatolia up to the British Isles, whose genomes were sequenced to high depth. Using different ancient individuals as references for early farmer and WHG admixture sources, we identified and compared the distribution, overlap, and similarity of WHG ancestry tracts across early farmer genomes. Our work provides insights into ancient individual relationships, their admixture history, as well as the role of admixture in shaping our genomes during range expansions.
Evaluating the reliability of f-statistics for making population genetics inferences using archaeological sediment samples

Sarah A Johnson

Sarah A Johnson, Elena Zavala, Pnina Cohen, Viviane Slon, Priya Moorjani
Tel Aviv University (Israel), University of California (USA)

Sediment DNA extracted from archaeological sediment samples is the next frontier in ancient DNA studies, as it provides the possibility of comprehensively sampling individuals present at a given archaeological site, even in the absence of skeletal remains. Sediment DNA samples tend to be composed of small quantities of degraded genomic fragments, often representing multiple species and individuals, within an overwhelming background of environmental DNA and present-day DNA contamination. The computational methods to analyze these data are still in their infancy and many widely used population genetics inference tools have not been tested for use with sediment DNA. In this study, we develop a novel simulator and assess the reliability of widely used f-statistics for application to sediment DNA. We systematically test the feasibility of answering two key questions: 1) Did admixture occur between a given set of populations? 2) Which known population is most closely related to a population of interest? Using a demographic model of human evolution, we assess the false discovery rate of f-statistics across various timescales. We further examine power as a function of SNP number and quality, and sampling age. We find that false discovery rates are relatively low in most situations, but the power of all tested statistics significantly decreases with fewer than 10,000 SNPs and in the presence of extreme ancient DNA damage or contamination. Based on our analyses, we present best practices for conducting population genetic analysis using archaeological sediment DNA, and revisit analyses of ancient human DNA in published sediment samples.
Genomic signatures of archaic introgression in human populations across the Pacific
Sofía Vieyra-Sánchez

Sofía Vieyra-Sánchez, Consuelo D Quinto-Cortés, Stephen J Oppenheimer, Carmina Barberena-Jonas, Alexander Mentzer, Alexander G Ioannidis, Andrés Moreno-Estrada
CINVESTAV (Mexico), Department of Biomedical Data Science, Department of Biomolecular Engineering, National Laboratory of Genomics for Biodiversity (UGA-LANGEBIO), School of Anthropology, Stanford Medical School (USA), The Centre for Human Genetics, University of California Santa Cruz (USA), University of Oxford (United Kingdom)

Numerous lines of evidence indicate that humans interbred with Neanderthals and Denisovans. While Neanderthal introgression appears to be evenly distributed among human populations, Denisovan introgression has been found by previous studies to be predominantly found in Pacific Islanders. Historically, most genetic studies in the Pacific have focused on Near Oceania, the western part of the region. The Oceanian Genome Variation Project addresses this gap by generating SNP array data from 981 individuals from East Asia and both Near and Remote Oceania. We used this novel dataset to explore the distribution of archaic introgression in a broader geographical range.

Our results confirm that the highest global proportions of Denisovan introgression are found in Pauans and Negrito populations from the Philippines. We also found evidence that Polynesian populations have detectable Denisovan introgression, although at lower global proportions than Pauan populations. Ancestry-specific analyses were performed to assess differential segregation of Denisovan introgression in modern human populations. These results revealed that Denisovan introgression in Papua and Polynesia is exclusively linked to their Pauan component, while Denisovan introgression in Negrito populations from the Philippines is recovered when looking at both the Pauan and East Asian components. This supports the idea that Denisovan introgression involved different human populations at different times, and unveils a wider distribution than previously thought. Finally, the multiple sources of Denisovan ancestry associated with different ancestry components in Negrito populations likely contributed to an unusual assimilation of archaic ancestors, explaining its ranking as the modern population with the highest Denisovan ancestry.
A high-quality genome from a 200,000-year-old Denisovan
Stéphane Peyrégne

Stéphane Peyrégne, Diyendo Massilani, Yaniv Swiel, Alba Bossons Mesa, Arev Pelin Sümer, Michael James Boyle, Cesare de Filippo, Bence Viola, Elena Essel, Sarah Nagel, Julia Richter, Antje Weihmann, Barbara Schellbach, Hugo Zeberg, Johann Visagie, Maxim B Kozlikin, Michael V Shunkov, Anatoly P Derevianko, Benjamin M Peter, Matthias Meyer, Svante Pääbo, Janet Kelso
Department of Anthropology, Department of Physiology and Pharmacology, Institute of Archaeology and Ethnography, Karolinska Institutet (Sweden), Max Planck Institute for Evolutionary Anthropology (Germany), Russian Academy of Sciences (Russian Federation), Siberian Branch, University of Toronto (Canada)

It is a little over a decade since the discovery of Denisovans, a human group that split from a common ancestor with Neandertals around 400 thousand years ago (ka). Nearly all we know about Denisovans is inferred from the single high-quality genome sequence of the Denisova 3 individual, who lived approximately 60ka. However, the presence of multiple Denisovan ancestry components in modern human genomes suggests that there were multiple genetically diverse Denisovan populations. Reconstructing more Denisovan genomes is therefore essential for understanding the population history underlying this genetic diversity. We describe a new high-coverage Denisovan genome reconstructed from DNA extracted from Denisova 25, a molar found at Denisova Cave. We estimated that this male individual lived around 200ka, in agreement with dating of the corresponding sediment layer. By comparing the genomes of Denisova 25 and Denisova 3, we determined that he belonged to an early Denisovan population that split about 220ka from later Denisovans at the cave, highlighting a population turnover. Approximately 12% of his genome is in long runs of homozygosity, suggesting that he lived in a small population and that there was recent inbreeding among his close ancestors. We also found evidence of recurrent mixing with Neandertals among the Denisovans at Denisova Cave; Denisova 25 inherited Neandertal ancestry from contacts with yet unknown divergent Neandertal populations. Furthermore, using both Denisovan genomes, we dissected Denisovan ancestry components in present-day human genomes. Denisova 25 provides a new view on Denisovan history and their interactions with other hominins.
Human history is characterized by intertwined biological, cultural and ecological processes. Discretizing space and time is often necessary for comparing these processes across different periods and/or regions of the world. However, partitioning archaeological and biological data into analytically comparable spatiotemporal groupings is far from straightforward. The general approach in archaeology is to group observations into archaeological cultures based on the similarity of discovery circumstances, such as proximity, material culture and stratigraphy. Depending on the type of discretization used, archaeological cultures could have various definitions which do not lend themselves naturally to a comparative analysis with other types of datasets (e.g. genetics). We propose a novel spatiotemporal partitioning algorithm designed for local time-series analysis that aims to aggregate data points in an automated manner, driven by the distribution of data itself. The algorithm is characterized by an approach that searches for an optimal division of space-time based on user requirements. In our specific case, we sought to compare time-series data in different geographical regions. Thus, our implementation works by minimizing within-group variance in space, while simultaneously maximizing data coverage across time. We applied this to the Big Interdisciplinary Archaeological Database (BIAD) with records spanning the European Neolithic and Bronze Age, enabling us to jointly compare evolutionary and cultural processes throughout Western Eurasia. We expect our method to be widely applicable to archaeo-scientists aiming to model different types of datasets in a statistically explicit and reproducible manner, allowing for more rigorous interdisciplinary studies of the past.
Benchmarking imputation accuracy in Ancient DNA datasets of pre-Columbian individuals from Mexico

VIRIDIANA VILLA ISLAS

VIRIDIANA VILLA ISLAS, Emily M Breslin, Flora Jay, Emilia Huerta-Sanchez, Maria C Avila-Arcos
Brown University (USA), CNRS, Department of Ecology and Evolutionary Biology, INRIA, International Laboratory for Human Genome Research, Laboratoire Interdisciplinaire des Sciences du Numérique (France), Smurfit Institute of Genetics. Trinity College Dublin (Ireland), Universidad Nacional Autónoma de México (UNAM) (Mexico), Université Paris-Saclay

Imputation of ancient genomes from the Americas has been assessed using genomes with relatively high coverage (>10x) and downsampling to lower depths, with results suggesting that imputation is reliable for aDNA datasets with > 0.5x genome coverage, for variants with MAF >5%. Here, we tested the reliability of imputation on lower depth genomes (<4.7X) from pre-Columbian individuals from Mexico. We evaluated the influence of biases introduced during imputation of these low coverage pre-Columbian genomes for which there is not a good representation of Native American individuals in the reference panel. First, we used one genomic dataset of a pre-Columbian (671-867 CE) individual from Mexico (2417Q) with a coverage of 4.7x, and down-sampled it to 2x, 1x, 0.5x, 0.2x, and 0.1x. We then, imputed each down-sampled genome at 77.8 million biallelic SNPs using the 1000 Genomes Project haplotype dataset as the reference panel. The evaluation of concordance between the variants observed in the original dataset and the imputed down-sampled genomes was low (0.67-0.72). Next steps of the project involve testing the effect of different parameters that could be influencing the imputation quality (e.g. usage of genotype likelihoods, improvement of the reference panel by adding local genomes) and evaluation of concordance (e.g. in function of the minor allele frequency or ascertained SNPs). Also, imputed down-sampled genomes from these and other pre-Columbian individuals from Mexico will be assessed to observe whether they conserve the same population structure as the original data set.
Genomic capture and typing of HLA in individuals from the pre-hispanic and colonial period of Mexico

Walter Nicolas Ortega


Brown University (USA), CNRS, Centro de Ciencias Genómicas, Department of Ecology, Department of Sociology & Anthropology (USA), Dirección de Salvamento Arqueológico, Escuela Nacional de Antropología e Historia, Evolution, INAH (Mexico), INRIA, Instituto Nacional de Antropología e Historia (Mexico), Laboratoire Interdisciplinaire des Sciences du Numérique (France), Laboratorio Internacional de Investigación sobre el Genoma Humano, Max Planck Institute for Evolutionary Anthropology (Germany), UNAM (Mexico), Université Paris-Saclay, and Organismal Biology

A factor that possibly influenced the decline of the Indigenous population following European colonization could be a poor response of the adaptive immune system against pathogens brought from other regions of the world. In this project, we analyze the genetic diversity of the HLA system, one of the most important region related to the adaptive immune response, in archaeological samples from pre-Hispanic and colonial periods. Since ancient DNA (aDNA) genomic libraries usually have low concentrations of endogenous DNA, we performed in-solution genomic capture of the HLA regions with a myBaits kit for 39 ancient individuals (34 from colonial and 5 from pre-Hispanic period, respectively), followed by NGS. We observed an increase in the coverage and depth in 26 genes in the HLA region, amounting to 100% coverage and depths close to 25x in our best libraries. HLA genotyping is currently being carried out with the tool Optitype, which has been successfully used before in aDNA data. The preliminary findings involve the identification of HLA class I alleles A*02; A*24; B*35; B*40; C*04 and C*03 as the most frequent in the dataset, which coincides with those reported today in Mexico.
Local ancestry inference (LAI) is a computational method crucial for identifying the mosaic of ancestry segments within admixed chromosomes. Its application has recently extended to the rapidly growing field of human ancient DNA (aDNA). However, the distinctive characteristics of aDNA data, such as low sequencing coverage and high error rate, present challenges for existing LAI tools designed for modern genomes. Furthermore, the performance of these tools on a subcontinental level with ancient DNA remains unclear. To explore these caveats, we conduct systematic benchmarking of one of the most widely used LAI software packages (RfMix), which is now also commonly applied to aDNA data. Our approach involves first constructing diploid chromosomes by integrating segments of farmer ancestry into one haploid of a phased Bronze Age Steppe-ancestry genome. Subsequently, we simulate typical aDNA fragments from these constructed ground-truth mosaics. Following the processing with a standard aDNA pipeline, we evaluate the accuracy of RfMix across different sets of reference panels, parameter settings, and aDNA quality. Our findings highlight the critical importance of reference panel size and reference population selection as primary variables affecting the accuracy of LAI. Additionally, parameters such as window size and generation significantly influence LAI performance. Finally, we identify a concerning systematic error manifesting as a high false positive rate at positions of a true positive segment on the alternate haploid.
When and where did our ancestor interbreed with Denisovans?

Yoko Satta

Yoko Satta, Sayaka Chiku

SOKENDAI, The Graduate University for Advanced Studies (Japan)

During the out-of-Africa dispersal of modern humans in the Upper Paleolithic, eastern non-Africans interbred multiple times with Denisovans after the Neanderthal admixture and as a result, currently harbor three distinct Denisovan genomic segments, (D0, D1, and D2) that differ both in the geographic distribution and in the extent of sequence divergences from the Altai Denisovan genome (D3). However, it is not known when and where these admixtures occurred. We developed a simple method that identifies Denisovan-derived segments in unphased genotype data. The method assumes a paraphyletic relationship among four genomes from a tested modern human and the reference D3 in addition to clock-like accumulation of nucleotide substitutions. We applied it to 3500-3800 year-old Jomon (F23), 31000 year-old Yana1, and extant Onge1, and obtained 500-900 segments per individual with each length being 30-100 kb. We first classified those segments into D0 and D2 using a two-component Gaussian mixture (GM) model and found a strong correlation between the number of Denisovan-derived segments and DNA contents. The D0/D2 ratio for F23 and Onge1 is smaller than for East Asians, but much higher than for South Asians. This is consistent with the even higher D0/D2 ratio for Yana1, and strongly suggests that D0 admixture occurred in northern East Asia. The GM model also estimated the D0 and D2 divergence time from D3 as 150-160 kya and 300-340 kya, respectively. We currently date D0 and D2 admixture events.
One of the most fundamental discoveries in evolutionary biology is the molecular clock—the notion that mutations occur steadily over time and thus could serve as an "evolutionary clock" for dating past events. Recent whole genome sequencing studies have challenged the validity of the molecular clock by revealing an almost two-fold difference in mutation rate over the course of human evolution. This discordance has been hypothesized to stem from a decrease in mutation rates towards the present in humans. To investigate this hypothesis, we leverage the history of archaic gene flow into modern humans to learn about mutation patterns at deep evolutionary timescales. Using whole genome sequences of ~900 individuals from the Human Genome Diversity Panel, we identify Neanderthal and Denisovan archaic segments in present-day individuals from diverse populations. We compare the number and types of derived mutations on the archaic and non-archaic syntenic fragments. We find similar overall mutation rates across lineages over the past 500,000 years. Comparison of mutation spectrum—i.e., composition of different types of mutations—reveals significant differences between archaic and modern human segments. We discuss the implications and mechanisms related to these changes. Future studies that use older calibration points can help to recover the mutation rate at deeper timescales and facilitate the recalibration of the molecular clock.
S16 - Evolutionary medical genomics.
Searching for adaptations to malaria parasites in blood group genes across non-human primates

Alejandra Herbert Mainero

Alejandra Herbert Mainero, Ellen Leffler
Department of Human Genetics. The University of Utah (USA)

Malaria parasites have influenced the evolution of multiple blood group genes in humans, presumably by altering the receptors these parasites use to gain access to red blood cells. Despite the presence of closely related parasites across the primate clade little is known of how they have impacted blood group evolution and if similar selective pressures have led to convergent adaptations. Here we applied selection tests to 44 genes that underlie human blood groups and expanded the number of primate species to identify putative adaptations to malaria-causing parasites across the clade. Our results show that 31.8% of the blood group genes show some evidence of positive selection (dN/dS>1) using PAML. We used BEB analysis with P>99% to identify positively selected sites within these genes. Unsurprisingly, in five genes the selected sites coincide with regions that may be accessible to these parasites. In the other nine genes, two possess no sites with P>99%, and seven either have no positively selected sites in extracellular regions or do not encode a surface protein, suggesting Plasmodium may not be the main selective pressure. We also noted a negative relationship between the number of antigens in the human blood group system and dN/dS>1 values, indicating possible ongoing diversifying selection. Therefore, we tested for balancing selection in primates with sufficient polymorphism data using betascan and found evidence in some cases of both selection across timescales. These results indicate new targets of selection across primate blood group evolution, some of which are consistent with a malaria selection.
Epistasis and pleiotropy constrain and expand the evolution of SARS-CoV-2 Omicron lineage

Alief Moulana

Alief Moulana, Thomas Dupic, Michael Desai
Harvard University (USA), Harvard Universityt (USA)

The Omicron lineage of SARS-CoV-2 has undergone extensive evolution, resulting in varied disease manifestations. The receptor binding domain (RBD) of the viral spike protein, has acquired numerous mutations, affecting its stability, ACE2 affinity, and immune evasion capability. These mutations' effects are often modulated by epistatic interactions and pleiotropy among the phenotypes they influence, complicating evolution prediction. Our study addresses this by systematically measuring ACE2 and antibody binding of more than 100,000 relevant variants. Particularly, we created four RBD variant libraries to explore the protein landscapes: (i) all BA.1 mutation combinations on the Wuhan Hu-1 strain; (ii) full spectra of BA.1, BA.2, BA.4, and BA.5 mutations; (iii) common Omicron mutations on divergent backgrounds; and (iv) random mutations on these backgrounds. First, we highlighted a compensatory epistasis in which specific BA.1 mutation combinations on the Wuhan Hu-1 background can mitigate the negative impact on ACE2 binding caused by immune escape mutations. Analysis of the second and third libraries illustrates how subvariant mutations that contribute to antibody escape diversity harbor independent effects on ACE2 binding, albeit with significant pleiotropic tradeoffs between ACE2 affinity and immune escape. Further, multiple mutations deleterious to ACE2 binding on the BA.1 background are well-tolerated on the BA.2 background, possibly contributing to the rise of BA.2 over BA.1. Together with more incoming data on random mutations, our findings further elucidate the tradeoffs between binding affinity and immune escape and the constraints presented by epistasis, providing valuable insights into viral evolution and potential therapeutic targets.
As SARS-CoV-2 evolves, mutations arising in new variants could potentially leverage epistasis to their advantage, increasing receptor-binding affinity or allowing the virus to escape antibodies. While large epistatic shifts occurred to enable the emergence of Omicron due to the N501Y substitution, it is unknown whether this epistatic trend will continue. Here, we aim to understand how epistasis is affecting the evolution of more recent viral variants. Using a yeast display system for deep mutational scanning, we constructed libraries of recent Omicron BQ.1.1 and XBB.1.5 spike receptor-binding domains (RBDs) containing all possible single amino acid mutations and single-codon deletions within the RBD, and used FACS-seq titrations to measure the impact of each mutation on ACE2 receptor-binding affinity. Patterns of mutational effects were compared against those previously measured for earlier SARS-CoV-2 variants to identify changes in mutational effects over time. We observed that many of the amino acid mutations and deletions were well tolerated, some even enhancing binding to ACE2, with many sites consistent in mutational impact across variants. However, we did identify epistatic interactions between the R493Q reversion that separates BQ.1.1 and XBB.1.5 from early Omicron variants and mutations at positions that later emerged in the more recent XBB.1.5-derived EG.5 lineage. Our results demonstrate the influence of epistasis on SARS-CoV-2 variant evolution. Our ongoing studies investigate the more recent BA.2.86, EG.5 and HK.3 SARS-CoV-2 variants, data that we hope will inform viral monitoring and forecasting.
High-throughput methods have enabled exploration of protein genotype-phenotype maps by comparing effects of many mutations on a background, as well as the paths by which proteins traverse these effects as they evolve in function. The former is typically done by deep mutational scanning and the latter by assaying a large combinatorial library of <20 substitutions separating two proteins. Thus, deep mutational scans are broad across the protein but shallow in the number of mutations made, while assays of combinatorial libraries are narrow but deep. Here we investigate genotype-phenotype maps in both breadth and depth, using directed evolution combined with high-throughput phenotyping to explore mutational effects along separate evolutionary paths. The protein we use is the germline form of the antibody CR9114, whose somatic form neutralizes a very broad range of influenza by binding a conserved epitope of the hemagglutinin antigen. Antigenic conditions and molecular pathways leading to the broad neutralization of flu are of interest to vaccine development and more generally to understanding dynamics of protein evolution during affinity maturation. We perform four rounds of directed evolution by yeast-displaying error-prone PCR mutants of germline CR9114, incubating with various hemagglutinin strains and combinations thereof, and selecting better-binding mutants by fluorescence-activated cell sorting. We then measure binding affinity at each round and under each antigenic condition via Tite-Seq. We thereby quantify genotype-phenotype maps and compare molecular patterns of affinity maturation across tens of replicates and under different antigenic conditions.
Radiation-induced stress and signatures of selection on anti-tumor immunity in Chornobyl wolves

Cara Love

Cara Love, Stacey Lance, Brian Arnold, Nicolas Rochette, Thomas Hinton, James Beasley, Bridgett vonHoldt, Michael Byrne, Dmitry Shamovich, Brian Nadel, Sarah Webster, Shane Campbell-Staton
EcoTur Belarus (Belarus), Fukushima University (Japan), Norwegian University of Life Sciences (Norway), Princeton University (USA), UCLA (USA), University of Georgia (USA), University of Missouri (USA)

Investigating the complex interplay between contaminant exposure, physiological responses, and evolutionary changes in wild populations is critical for understanding the impacts of anthropogenic environmental change. Such investigations can shed light on the broader health implications of these changes and may provide novel models for examining adaptive responses to disease causative agents. We explore the consequences of chronic multigenerational radiation exposure on gray wolves (Canis lupus) in the Chornobyl Exclusion Zone (CEZ) as a model for examining adaptive anti-tumor immunity. By analyzing patterns of regulatory and genetic variation, we uncover molecular signatures of immune stress and adaptive evolution. Whole blood transcriptome data reveal significantly altered PBMC populations among Chornobyl wolves and regulatory divergence at both single gene and co-regulatory levels in CEZ wolves. These results highlight radiation-induced immune modulation, cellular apoptosis, and antitumor immune response as potential physiological consequences of radiation stress and targets of selection within the CEZ population. We further examine regulatory decoherence and genomic signatures to identify signatures of selection within the CEZ wolf population. Combining these techniques, we identify several targets of selection overlapping genes influential in cancer physiology. These genes encompass crucial functions such as anti-tumor immunity, cellular invasion, and migration. Functional modeling of selective targets further suggests significant alterations in proteins crucial for regulating immune response and cancer physiology.
Investigating the dynamics of protein constraint across the tree of life with deep Bayesian hierarchical models.
Charlie W J Pugh

Charles WJ Pugh, Jonathan Frazer, Mafalda Dias
Centre for Genomic Regulation (Spain)

Generative probabilistic models of protein sequences are transforming protein design, pathogen forecasting and clinical variant annotation. In the latter case, provided certain criteria are met, these models can now be considered strong evidence for missense variant pathogenicity and are an invaluable diagnostic tool. Crucially to their success, these expressive models are able to capture constraints on protein sequences by learning from variation across the tree of life. However, current variant effect prediction approaches model the distribution of broad protein families or, in the case of protein language models, the entire protein universe in order to make predictions in a specific species. Modelling all sequences as from a single distribution in this way overlooks the changes in constraint on a protein across the tree of life. To address this issue we have developed a novel modelling strategy to capture the distribution of sequences conditioned on the position in the phylogeny. We achieve this with a deep Bayesian hierarchical model based on the phylogenetic structure of large multiple sequence alignments. Our model learns from diverse sequences across the phylogeny while also capturing constraints that have evolved more recently. This has enabled us to obtain state-of-the-art scores on the ProteinGym benchmark. Our approach has allowed us to trace the evolutionary history of constraints in a selection of disease relevant human proteins. With this we have identified differing evolutionary signatures in gain-of-function and loss-of-function variants, giving promise that our approach can help classify disease aetiology.
Deep learning phenotype imputation to measure dominance of rare variant effects on rare disease in UK Biobank

Daniel M Jordan

Daniel M Jordan, Ben Omega Petrazzini, Robert Chen, Joshua K Park, Iain S Forrest, Ha My T Vy, Áine Duffy, Ghislain Rocheleau, Ron Do
Icahn School of Medicine at Mount Sinai (USA)

Recent studies on the genetic architecture of human traits and diseases have suggested that traits thought to be recessive or dominant may actually have a range of intermediate dominance. However, it has been difficult to test this hypothesis on rare variants or rare diseases, as non-additive genetic association studies require very large sample sizes to be powered. We have previously developed a deep learning phenotype imputation approach that allows us to conduct association studies on rare phenotypes (prevalence < 1 in 2,000) for aggregated ultra-rare coding variants (minor allele frequency < 10^{-5}). We report recessive and additive ultra-rare variant burden association tests for 155 rare phenotypes conducted using REGENIE in approximately 500,000 UK Biobank participants. Even with the lower power of recessive model association tests, we find many more significant disease-associated genes with the recessive model than the additive model (143 recessive vs. 39 additive), as well as a substantial bias of sub-significant effects towards recessivity. We find no clear examples of intermediate dominance, in contrast to results in common-variant and common-disease regimes. Overall, our findings suggest that the effects of rare variants on rare disease are largely recessive, with an additional significant role for additivity and minimal intermediate dominance. Testing these hypotheses has not previously been possible, providing a powerful example of how state-of-the-art deep learning methods applied to population-scale biobanks can reveal insights about the genetic architecture of human traits and diseases.
HLA class I escape drives the evolution of SARS-CoV-2 in human population
Ekaterina Riumina

Ekaterina D Riumina, Evgenii I Alekseeva, Galya V Klink, Georgii A Bazykin
A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences (Russian Federation), MSU (Russian Federation), Skolkovo Institute of Science and Technology (Russian Federation)

SARS-CoV-2 evolution is shaped by human adaptive immunity, with mutations that allow escape from the B-cell response conferring selective advantage and spreading in the population. Meanwhile, the role of the escape from T-cellular cytotoxic response remains controversial. Here, we study the origin and spread of SARS-CoV-2 variants that allow escape from presentation by the HLA class I alleles that are common in human populations. We find that 35% of mutations that are characteristic of the variants of concern, and 34% of mutations that have reached high (>5%) frequencies in viral populations, facilitate escape from T-cellular alleles. Mutations associated with escape from common HLA alleles reach higher frequencies than those that allow escape from less common HLA alleles. Moreover, viral escape mutations reach higher frequencies in those countries where the causal HLA alleles are more frequent, indicating that viral escape is driven by the local genetic composition of the human host population. Together, these data indicate that CTL escape is a major driver of SARS-CoV-2 evolution and an epidemiological concern, and reveal a novel facet of selection on this virus. The study was supported by RSF grant 21-74-20160.
Characterisation of haploinsufficient tumour suppressor genes across human cancers

Elle Loughran, Aoife McLysaght
Trinity College Dublin (Ireland)

Tumour suppressor gene (TSG) losses are canonically recessive, following Knudson’s seminal two-hit model of retinoblastoma occurrence. However, it has become clear that some important tumour suppressor genes, including TP53 and PTEN, are haploinsufficient: loss of one copy, or in the case of PTEN even mild decreases in activity, are enough to promote tumorigenesis. These haploinsufficient TSGs are potential targets for activation therapy, and their study may additionally offer insights into the functional impacts of copy number alterations in general. In this project, we aim to systematically identify dosage-responsive tumour suppressor genes across cancer types from the Cancer Genome Atlas and the Cancer Cell Line Encyclopedia and characterise features that distinguish dosage-responsive and non-responsive TSGs. We use a combination of approaches to screen for altered tumorigenicity after partial TSG loss, including survival analysis, inference of cell proliferation rate from gene expression data, and information from DepMap RNAi knockdowns. We then search for associations between TSG dosage-responsiveness and features including germline dosage sensitivity and dosage compensation. Finally, we will compare the loss patterns of dosage-responsive and nonresponsive TSGs in the context of polyploid tumours, where additional ‘backup’ copies of TSGs make full knockout less likely.
Investigating the robustness of response to environmental perturbations across primate cardiomyocytes

Emma M. Pfortmiller

Emma M. Pfortmiller, Michelle C. Ward
The University of Texas Medical Branch at Galveston (USA)

Susceptibility to complex disease, such as cardiovascular disease (CVD), is influenced by a poorly-understood balance between genetics and environment. Given that CVD-associated genetic variants are largely located within non-coding regions of the genome, their effects are likely mediated through the regulation of gene expression (GE). GE regulation is context-dependent and impacted by environmental stimuli. Due to the sequence similarity between humans, it can be challenging to identify inter-individual differences in GE and therefore susceptibility to disease. To magnify GE changes, we take an evolutionary approach comparing humans to our closest evolutionary relative. We developed an in vitro model using induced pluripotent stem cell-derived cardiomyocytes from humans and chimpanzees to study robustness of response to stimuli. We exposed cells for 24 hours to eight stimuli known to induce the unfolded protein response (UPR), DNA damage response (DDR), innate immune response (IMR), as well as manmade chemicals where the response mechanism is unknown (MCR). Across species, UPR, DDR, and MCR treatment classes decrease viability at high concentrations, while IMR does not. We selected sub-lethal concentrations of each treatment to measure the transcriptional response to stimuli. We identified an up-regulation of known response genes including HSPA5 (Tunicamycin-UPR), CDKN1A (Nutlin-3-DDR), IL-6 (TNF?-IMR) and PPARA (PFOA-MCR) in both human and chimpanzee cells. With this model, we are measuring global GE changes to gain insight into species-specific molecular mechanisms induced following perturbation, which may contribute to CVD susceptibility.
Atherosclerosis is a global disease characterized by a hardening of the arteries. This is caused by a plaque accumulation, primarily composed of foam cells derived from macrophages that take up oxidized low-density lipoprotein (oxLDL). To understand the disease in detail, mainly mouse models and in vitro human models are used. To bridge the evolutionary gap between mice and humans, non-human primates (NHPs) are also used as a model system for biomedical research. However, we know very little about the molecular similarities and differences among NHPs and humans in atherogenesis. In particular, it has been suggested that NHPs are less likely to form the typical plaques in response to a high-fat diet, making it especially important to elucidate common and distinct mechanisms in atherosclerosis among NHPs. Here I present an in vitro model of foam cell formation from induced pluripotent stem cell (iPSC)-derived macrophages. Our methodology allows the differentiation of NHP iPSCs into macrophages and their uptake of oxLDL. In turn, this allows to compare the transcriptional profiles between human and NHP macrophages and foam cells to identify conserved and diverged regulatory networks of primate atherogenesis.
The evolutionary origins of autoimmune and infectious disease risks
Evan K. Irving-Pease

Evan K. Irving-Pease, William Barrie, Eske Willerslev, Lars Fugger, Astrid K. N. Iversen
University of Bremen (Germany), University of Cambridge (United Kingdom), University of Copenhagen (Denmark), University of Oxford (United Kingdom)

Autoimmune diseases are one of the leading causes of chronic illness, and their prevalence has been rising for decades. Our understanding of the genetic architecture of autoimmunity has been greatly advanced by large genome-wide association studies (GWAS), which have shown that the majority of risk variants are commonly occurring and many are located within or near immune genes. Many of these risk variants also confer protection against infectious diseases, suggesting that natural selection may have had to balance protection from pathogens against increased risk of autoimmunity—in a process known as ‘antagonistic pleiotropy’. Previously, we have shown that genetic risk for multiple sclerosis was subject to strong polygenic selection in Eurasia at a time when human populations experienced increased exposure to pathogens. Here, we model the evolutionary history of ten autoimmune diseases (including type 1 diabetes, coeliac disease and psoriasis). Using a panel of >1,000 ancient genomes, we inferred the allele frequency trajectories and selection coefficients of genetic risk variants for each autoimmune disease, while accounting for the complex demographic history of prehistoric Eurasia. Our results reveal that many autoimmune diseases show evidence of strong polygenic selection during recent human evolution. Using a joint statistical test designed to measure the impact of selection on genetically correlated pairs of traits, we tested for pleiotropic interactions between each autoimmune disease and more than 30 infectious disease GWAS phenotypes. Our results reveal the complex evolutionary trade-offs which have shaped present-day disease risk for both autoimmune and infectious diseases.
Polygenic disorders are a broad class of diseases caused by multiple genetic mutations, often involving noncoding and regulatory variants. The most common approaches to predicting disease phenotype from genotype are scores computed with linear regression models but their potential as diagnostic tools is still heavily debated. Indeed, these models still exhibit severe overfitting to the training cohort, with performance deteriorating as one moves to more distantly related cohorts. They also provide limited insight into disease aetiology. These limitations are thought to be largely due to current approaches being restricted to considering only common variants and the difficulty in identifying the causal relations between variants and their pathways. To overcome these issues we propose a generative modelling approach which leverages evolutionary constraints. Indeed, growing evidence shows that genomic models trained on complementary evolutionary scales offer a very accurate guide for disease-causing variants in severe developmental disorders (Orenbuch et al, 2024; Fiziev et al 2023). Still, the potentiality of evolutionary information in complex polygenic disorders remains mostly unexplored. Our approach provides a unified framework for predicting the joint impact of both common and rare variants and revealing novel disease sub-classes. We apply this modelling approach to type 2 diabetes, a disease with large genetic heterogeneity, for which a clear diagnostic criteria is still lacking.
Quantifying virus-driven adaptation in Myotis genomes using the McDonald Kreitman test

Genavieve Kathryn Gray-Sandoval

Genavieve K Gray-Sandoval, M. Elise Lauterbur, Juan Manuel Vazquez, Peter Sudmant, Lucie Etienne, David Enard
CIRI, Ecole Normale Supérieure de Lyon, University of Arizona (USA), University of California Berkeley (USA), Université Claude Bernard Lyon 1 (France)

Bats are a natural reservoir for a multitude of viruses including coronaviruses, filoviruses, henipaviruses, and retroviruses. Bats are thought to be a major host of past and future zoonoses, so it is important to understand their evolutionary history. We can develop a better understanding of this evolutionary relationship between bats and viruses by evaluating patterns of virus-driven adaptation in bat genomes. Myotis bats provide details of these adaptations due to their recent radiation allowing us to look for differences among and within species. We aim to detect ancient viral epidemics in Myotis bats through genomic adaptation analyses. We created reference genomes for 9 Myotis species in the western United States. This allows us to conduct an unbiased genome-wide approach to identify gene candidates that have undergone strong selection. We have resequenced Myotis bat genomes assembled with PacBio HiFi using short reads and genome-wide variant calls to estimate protein adaptation using the framework of the McDonald Kreitman test. We then focused specifically on adaptation driven by viruses at bat genes that are known to interact with viruses. Viral interacting proteins (VIPs) are known to be substantial drivers of mammalian adaptation and can allow us to identify signals of adaptation. Positive selection at VIPs may indicate the impact of viruses on ancient bat populations. Our results highlight patterns of virus-driven adaptation in previously largely unexplored bat genomes.
Interactions between sickle cell and Plasmodium falciparum genotypes in asymptomatic malaria

Helena Hopson

Helena Hopson, Alejandro Herbert Mainero, Heejung Shim, Sandrine E Nsango, Ellen M Leffler
Centre Pasteur du Cameroun (Cameroon), Department of Biological Sciences, Department of Human Genetics, Faculté de Médecine et des Sciences Pharmaceutiques, Malaria Research Unit, Melbourne Integrative Genomics, University of Melbourne (Australia), University of Utah School of Medicine (USA), Université de Douala (Cameroon)

The sickle cell allele (HbS) is protective against severe malaria caused by Plasmodium falciparum, but the mechanism remains unclear. Intriguingly, multiple nonsynonymous mutations in parasite genes have recently been found to be enriched in HbS carriers who do develop severe malaria, suggesting these changes allow parasites to overcome the protective effect of HbS. However, it is also unclear why this would be beneficial to the parasite since HbS does not protect against asymptomatic infection, and the majority of infections in HbS carriers are asymptomatic. To begin to understand the functional and fitness consequences of these parasite genotypes, we investigate the impact of parasite genotype and host HbS status on gene expression and mosquito transmission in blood samples from 70 children with asymptomatic infections from Cameroon. Using dual RNA-seq, we identify host and parasite genes differentially expressed between HbS and parasite genotypes. We find differential expression of genes that transport parasite virulence proteins to the red blood cell surface, supporting their reduced export as a likely key mechanism for HbS protection. We also find evidence for epistatic interactions on transmission quantified by membrane feeding assays, suggesting the fitness of parasite genotypes is linked to HbS genotype. We hypothesize this may have led to a balanced polymorphism in parasites in which HbS is the selective pressure. Analysis of this ongoing interaction offers a unique approach to reveal key mechanisms of HbS protection and co-evolutionary dynamics between host and parasite genetic variation.
Malaria in African monkeys: recovering the genomes of Plasmodium and Hepatocystis parasites from a co-infected monkey

Hend Abu-Elmakarem

Hend Abu-Elmakarem, Reuben W. Nowell, Weimin Liu, Lindsey J. Plenderleith, Beatrice H. Hahn, Paul M. Sharp

Biological and Environmental Sciences, Department of Medicine, Institute of Ecology and Evolution (IEE), Philadelphia (USA), University of Edinburgh (United Kingdom), University of Pennsylvania, University of Pennsylvania (USA), University of Stirling (United Kingdom)

Plasmodium species of Old World monkeys form a clade with human-infecting parasites for which an African origin has been suggested. However, most members of that lineage infect Asian monkeys, while only two genetically characterized species are known to infect African monkeys, P. gonderi and P. mandrilli (formerly P. sp. DAJ?2004), and these infections are rare. In contrast, African monkeys are often found to be infected with strains of Hepatocystis; relatives of Plasmodium that are transmitted through Culicoides biting midges rather than Anopheles mosquitoes. Genomic sequences of these African monkey parasites are scarce; in addition to P. gonderi, there is one genome available for a strain of Hepatocystis. We used Selective Whole Genome Amplification to obtain genome sequences from a dried blood spot, from an African monkey (Cercopithecus pogonias), known to contain Plasmodium DNA. Analysis revealed a co-infection, with both P. mandrilli and a Hepatocystis parasite. We first present a bioinformatic pipeline to disentangle the genomes of these closely related species. Next, we conduct a comparative analysis of the evolution of Plasmodium species infecting African monkeys, Plasmodium species infecting Asian monkeys, and Hepatocystis species infecting African monkeys. Analyses of the rates of synonymous and non-synonymous substitutions of orthologous genes are used to look for signatures of adaptive evolution in the parasite genomes. We also characterise genomic differences between Hepatocystis and Plasmodium that parasitize African monkeys, particularly in regards to species-specific genes families, their host invasion mechanism and their possible role in the variation in the parasites’ prevalence levels.
Phylogenetic analysis of Mycobacterium tuberculosis sequences from Mexico in the global situation
Ikuri Alvarez-Maya

Ikuri Alvarez-Maya, Manuel Garcia-Ulloa, Jaime Martinez-Urtaza (Mexico), Autonomous University of Barcelona (Spain), Center for Research and Assistance in Technology and Design of the State of Jalisco (CIATEJ), Department of Genetics and Microbiology, Faculty of Biosciences, Medical and Pharmaceutical Biotechnology

The global spread of Mycobacterium Tuberculosis (MTB) remains a major public health issue. Since data on resistance to anti-tuberculosis drugs are collected, the country coverage of such data is much more limited in the country as Mexico. We study the population structure of Mexican populations, through the analysis of the introduction nature of genotypes, by searching for clusters and genomic variation indicating possible transmission events involved in spreading Mycobacterium tuberculosis multidrug resistant (MDR-TB) behavior in Mexico. MTB raw Illumina short reads were collected from projects PRJEB30933, PRJNA824124, and PRJEB44165 on the website www. ebi.ac.uk/ena repository. Additionally, the scaffolds from project PRJNA751891 at ww.ncbi.nlm.nih.gov. were also collected. Multidrug resistance profiles, lineages, and sublineages were assigned to strains in silico with tb-profiler v4.3.0. The alignment of all Mexican and worldwide strains yielded a total of 31,951 variable sites which were used for the phylogenetic reconstruction. Lineage 4 had the highest frequency, with 336 (66.8%) of the global MTB complete genomes belonging to this lineage. This was followed by lineage 2, which accounted for 89 (17.7%) of the global MTB complete genomes. Interestingly, our in-silico analysis of drug resistance worldwide revealed a high percent of strains belonging to lineage 2 with Pre-Extremely Drug Resistant TB phenotype. Overall, Mexico has a percentage greater than 90% of L4 strains, as well a particular sublineage with 90% percent of MDR-TB was found endemic. Thus, our study contributes to clarifying the implications for the evolution of this pathogen in Mexico and worldwide.
Rapid evolution of mutation rate and germline maintenance under strong pathogen selection: new insights from experimental evolution experiments

Imroze Khan

Basabi Bagchi, Dipendra Nath Basu, Joy Bose, Srijan Seal, Imroze Khan
Ashoka University (India)

Pathogenic infections and costly inflammatory responses are major regulators of mutation susceptibility, imposing challenging threats on health, but the evolutionary and demographic causes and consequences of such impacts are unclear. Here, we took a novel integrated approach aimed at expanding our understanding of how infection and immunity play a causative role in modulating mutation rates and their transmission across generations. Using tractable insect models (e.g., Drosophila melanogaster and Tribolium castaneum), we first show that immune activation and pathogenic infections can consistently increase the rate of deleterious (dominant and recessive) mutations in subsequent generations by compromising germline maintenance. Resource limitation or poor-quality diet can further amplify the effects of such mutation transmission. However, a long-term evolution under strong pathogen selection can drastically reduce the accumulation of deleterious mutation load in immune-challenged and infected individuals by improving germline maintenance. We could also identify the possibility of increased activity of multiple DNA repair pathways (e.g., homologous recombination, mismatch repair and non-homologous end joining) in the ovary of evolved females that contributes to their improved germline maintenance, using large-scale transcriptome analyses and RNAi. Together, these results can provide a novel conceptual framework of how variation in immune responses can, in principle, be linked to mutation load and germline maintenance, generating fundamental insights into what drives natural variation in organismal health and resistance to pathogens across generations.
An individual’s chronological age does not always correspond to the health of tissues in their body, especially in cases of disease. Therefore, estimating an individual’s chronological age is a useful tool to diagnose disease and its progression. We present novel metrics based on the phenomenon of clonal hematopoiesis, which is associated with blood-related diseases, to quantify the loss of phylogenetic diversity in hematopoietic stem cells (HSCs), the precursors of differentiated blood cells. These metrics capture the increase in blood cancer incidence with age, establishing a model which estimates the phylogeny-based physiological age of HSCs in an individual, which we call “PhyloAge.” We will present novel approaches for estimating PhyloAge from high resolution data like single cell sequences of HSCs, or from bulk sequencing of blood that produces single nucleotide polymorphisms in HSCs. For both types of datasets, we are able to accurately estimate the PhyloAges of healthy individuals within 10 years. We will also present results for individuals suffering from blood cancers, which we hypothesize to have HSC PhyloAges exceeding their chronological ages. We will also test the hypothesis that the presence of blood cancer drivers is associated with longevity, as there is a reported overabundance of these drivers in individuals over the age of 90. Thus, PhyloAge represents a complement to conventional biomarker-based methods of estimating the physiological age of blood and assessing disease risk. Its effectiveness for bulk sequencing, in addition to single cell sequencing datasets, would enable its adoption in research and clinical settings.
Mutation and fitness landscape of SARS-CoV-2 reveal strong selection on synonymous mutations

Jean-Francois Gout

Jean-Francois Gout, Jori Symons, Claire Chung, Bert M. Verheijen, Sarah Shemtov, Dorien de Jong, Gimano Amatngalim, Monique Nijhuis, Marc Vermulst

Department of Medical Microbiology, Division Child Lung Diseases, Leonard Davis School of Gerontology (USA), Mississippi State University (USA), Regenerative Medicine Center Utrecht, Translational Virology, University Medical Center (Netherlands), University of Southern California

New variants of SARS-CoV-2 are a constant threat to current vaccine and treatment strategies for COVID-19. Because these variants are created by mutations, we obtained precise mutation rate and spectrum of five SARS-CoV-2 variants and quantified the fitness effect of almost 4,000 mutations using ultra-deep circular RNA consensus sequencing (CirSeq). We identified several virus and host-specific features that influence SARS-CoV-2 evolution. Most notable, we discovered that the secondary structure of the viral RNA molecule protects the SARS-CoV-2 genome against RNA damage-derived mutations. We also show that the preservation of these secondary structures is a major driver of fitness. This study improves our understanding of SARS-CoV-2 genome evolution and identifies key weakness of the virus which can help optimize the generation of treatments and vaccines.
The genetic architecture and the evolutionary consequences of the human pelvic form
Liaoyi Xu

Liaoyi Xu, Eucharist Kun, Marianne Brasil, Tarinder Singh, Vagheesh Narasimhan
Columbia University (USA), The University of Texas at Austin (USA), Western Washington University (USA)

Since diverging from chimpanzees, pelvic shape has evolved significantly in the human lineage and is thought to be affected by the obstetrical dilemma—an evolutionary mismatch between the increasing infant brain size and a narrowing birth canal. To elucidate the genetic underpinnings of pelvic morphology, we applied a deep learning model to 31,115 DXA scans from the UK Biobank, extracting a set of seven pelvic proportion (PP) phenotypes. These phenotypes are highly heritable (25-40%) and linked to 179 independent loci, indicating significant genetic determinants of pelvic shape. Unlike other skeletal proportions, the subpubic angle, associated with birth canal morphology, exhibits a genetic correlation between sexes significantly less than 1, aligning with its reproductive significance. Phenotypic and genetic association analyses revealed that larger birth canal phenotypes were associated with reduced walking pace, decreased risk of back pain, and increased risk of hip osteoarthritis, all phenotypes linked to locomotor efficiency. We also observed that narrower birth canals are associated with a reduced risk of pelvic floor disorders but an increased risk of emergency cesareans delivery and obstructed labor due to insufficient dilation, reinforcing the obstetrical dilemma’s impact. Moreover, the genetic correlation between birth canal width and infant brain size (proxied by birth weight) supports the co-evolution of these traits, but no link was found between pelvic phenotypes and gestational duration. This comprehensive study provides significant insights into a longstanding debate on human evolution, emphasizing the obstetrical dilemma’s role in the co-evolution of the human brain and pelvis.
The evolution of virulence is a classical example of life-history tradeoffs with major implications for public health. For HIV, the tradeoff has been quantitatively described from epidemiological evidence. We show that phylogenetic inference coupled with clinical measurements of set-point viral load reveals a single-peaked viral fitness landscape, providing a direct demonstration of the trade-off. However, evolution of the virus is not restricted to this landscape. We discovered a highly virulent variant of subtype-B HIV-1 spreading in the Netherlands in 1998-2014. Infected individuals exhibit a 3.5-5.5 fold increase in viral load, as well as CD4 cell decline twice as fast as individuals with other subtype-B strains. Epidemiological features suggest that the increased virulence is attributable to the viral strain. Genetic sequence analysis suggests that this variant arose in the 1990s from de novo mutation, not recombination, with increased transmissibility and an unfamiliar molecular mechanism of virulence. This surprising finding illustrates the difficulties in predicting the evolution of virulence and the need for genetic monitoring of viral variants.
Hidden in Plain Sight; Staphylococcus argenteus as an obscured public-health risk.

Maitiú Marmion

Maitiu Marmion, Cheryl P Andam
Department of Biological Sciences, University at Albany State University of New York (USA)

Staphylococcus aureus is one of the best characterized bacterial species, due to its prevalence in the human holobiont and its role in antibiotic-resistant infection. However, it has become evident in recent years that Staphylococcus argenteus, which had previously been considered as part of Staphylococcus aureus Clonal Complex-75, poses a novel risk to human health. This study aimed to assess the global prevalence, phylogenetic relationships, and genomic features of S. argenteus. Ecological sources from which publicly available sequences of S. argenteus has been isolated include sick and healthy humans, animals, food, and environment, with most samples found in Asia and Europe. We identified four major sequence types (ST2250,ST2198,ST2793,ST1223) which vary in the number and types of antibiotic resistance and virulence genes present per genome. 37 virulence genes and 6 antimicrobial resistance genes were detected in all genomes. Type 7 secretion systems were significantly more present in non-pathogenic isolates, while distinct cap8 capsular and set exotoxin profiles were found among different STs and ecological sources. The mecA gene, which confers broad resistance to beta-lactams, was detected in 27.86% of S. argenteus genomes. Time-calibrated phylogenetic analysis revealed the time of the most recent common ancestor was 281 C.E., followed by multiple diversification events. This study provides interesting insights into the spread and diversity of a lesser-known pathogen which has often been eclipsed by S. aureus. Our results highlight the need for broader and systematic surveillance efforts to identify reservoirs, resistance determinants, and clinical pathologies of S. argenteus.
On the origins of the BA.2.86 (\'Pirola\') SARS-CoV-2 lineage

Marina Escalera-Zamudio

Marina Escalera-Zamudio, Cedric CS Tan, Lucy van Dorp, François Balloux
Department of Genetics, END-VOC consortium (United Kingdom), Evolution & Environment, UCL Genetics Institute, University College London (United Kingdom)

Global SARS-CoV-2 surveillance enabled fine-scale monitoring of divergent variants that display unusually high number of mutations presumably conferring enhanced transmissibility and/or immune evasion, and that may lead to epidemic waves. One example is the Omicron-descending BA.2.86 sublineage (\'Pirola\'), first detected in mid-2023. With 34 \'unique\' non-synonymous mutations in the Spike protein, and a predominant circulation in the USA/Europe, it was classified as a variant of interest (VOI) in November 2023. By February 2024, the BA.2.86 diverged into multiple sublineages, and dominated worldwide (BA.2.86*/JN.1*). Using multiple fine-scale evolutionary analyses applied to selectively subsampled large-scale datasets, we explore the evolutionary process underlying the emergence of the BA.2.86*/JN.1* lineages. Our results reveal a basal cluster of hybrid BA.2/BA.2.86-like genomes displaying some (but not all) BA.2.86-specific lineage defining mutations. Hybrid genomes displayed synchronous circulation in multiple countries prior to the emergence of the BA.2.86. Additionally, we find evidence for diversifying selection, and for putative BA.2 X BA.2.75 recombination in the branch leading to the BA.2.86*/JN.1* clade. The dominant hypothesis explaining the emergence of divergent SARS-CoV-2 variants postulates that these may arise through accelerated evolution in chronic infections. Our findings point out to cryptic co-circulation and gradual evolution, incompatible with a single emergence associated with prolonged shedding. Our results contribute to understanding the evolutionary mechanisms behind the emergence of (some) divergent SARS-CoV-2 variants, and underscore the importance of adequate subsampling strategies when interpreting initial phylogenetic placements of emerging lineages.
Transcriptional, epigenetic and evolutionary mechanisms shaped the acute inflammatory response of endothelial cells across mammals.

Mohamed Bayoumi Fahmy


Cincinnati Children's (USA), Genetic and Genomi biology department, Paris sud (France), Princess Margaret Cancer Centre, The hospital for sick children (Sickkids) (Canada), University Health Network (Canada), University of Toronto (Canada), University of Wisconsin-Madison (USA), Wisconsin institute of discovery

Endothelial cells line all blood vessels and play a key role in maintaining physiological homeostasis. Augmented inflammation of ECs is a known risk factor for cardiovascular disease (CVD) and bleeding disorders. However, regulatory circuits that govern the inflammatory response of ECs remain poorly understood. Here, we used comparative and functional genomics to explore regulatory landscape that controls the acute inflammatory response. We characterized the inflammatory response of aortic EC from five mammals namely, human, mouse, rat, dog, rat and pig before and after simulation with TNF after 45 minutes using RNA-seq, ChIP-seq and ATAC-seq. We found number of inflammatory effectors to be conserved across all species such as CCL2 and CSF1. We also observed a markedly stronger response in mouse compared to other mammals. Transcription factor (TF) footprinting identified NF-κB related motifs (e.g. REL) to be dominant TF observed after acute TNF treatment in all species except in mouse where we found AP1-related TF (e.g. c-JUN) footprints at similar level to the NF-κB footprints. Consistently, REL and c-JUN ChIP-seq data for human and mouse EC treated with TNF for 45 minutes identified both TFs to be differentially accessible at similar levels after treatment in mouse, unlike human. Moreover, we found accelerated substitutions in the ape lineage enriched in the regulatory elements closest to genes of the highest response in human. In conclusion, we found various mechanisms that have shaped the inflammatory response of EC in mammals.
An introgressed KIR haplotype enabled positive selection on HLA-A, enhancing Natural Killer cell function in Oceania

Neus Font-Porterias

Neus Font-Porterias, Liyen Loh, Philippa M Saunders, Camilla Faoro, Neda Nemat-Gorgani, Genelle F Harrison, Suraju Sadeeq, Luca Hensen, Shu Cheng Wong, Jacqueline Widjaja, Shiyiing Zhu, Gonzalo Montero-Martin, Katherine M Kichula, E Birdie Clemens, Marcelo Fernandez-Vina, Lisbeth A Guethlein, Julian P Vivian, Jane Davies, Adrian Miller, Alexander J Mentzer, Stephen J Oppenheimer, William Pomat, Peter Parham, Jamie Rossjohn, Steven YC Tong, Katherine Kedzierska, Andrew G Brooks, Paul J Norman (Australia), Big Data Institute, Biomedicine Discovery Institute, Cardiff University School of Medicine (United Kingdom), Department of Biomedical Informatics, Department of Immunology and Microbiology, Department of Infectious Diseases, Department of Microbiology and Immunology, Department of Pathology, Department of Structural Biology and Department of Microbiology and Immunology, Griffith University (Australia), Indigenous Research Unit, Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Institute of Infection and Immunity, Institute of Social and Cultural Anthropology, Li Ka Shing Centre for Health Information and Discovery, Menzies School of Health Research, Monash University (Australia), Papua New Guinea Institute of Medical Research (Papua New Guinea), Peter Doherty Institute for Infection and Immunity, Royal Darwin Hospital (Australia), School of Anthropology and Museum Ethnography, Stanford Blood Centre, Stanford University (USA), The Royal Melbourne Hospital, University of Colorado School of Medicine (USA), University of Melbourne (Australia), University of Oxford (United Kingdom), Victorian Infectious Diseases Service, Wellcome Centre for Human Genetics

Pathogen exposure and demography diversify immunity-related genes across human populations. In particular, admixture and local adaptation enhance diversity of the killer cell immunoglobulin-like receptors (KIR) and their ligands, human leukocyte antigens (HLA) class I. HLA class I presents protein fragments derived from infecting pathogens at the cell surface, which are recognized by KIR to modulate immune responses. Enabling the unique analysis of the combinatorial diversity of KIR and HLA class I, Oceanians show complex admixture patterns and evidence of archaic introgression. We analyzed genome-wide SNPs together with HLA and KIR sequencing data from 1200 Oceanians including 80 First Nations Australians. The most prevalent KIR ligand in Oceania is HLA-A*24:02, previously associated with poor prognosis following influenza virus infection. The high frequency of HLA-A*24:02 is coupled with evidence for positive selection and low sequence diversity. We also identify KIR3DL1*114, an allele of the inhibitory KIR3DL1, uniquely characterized by positively-selected Phenylalanine at HLA-contacting residue 166. KIR3DL1*114 is restricted to Oceanians, having highest frequency in Papuans, and D-statistics and haplotype divergence indicate likely Denisovan introgression. Functional analyses of KIR3DL1*114+ donor immune cells, molecular and crystallography data all show that KIR3DL1*114 binds more strongly than other allotypes to HLA-A*24:02. KIR3DL1*114 also has differential specificity for the peptides presented by HLA class I, including those derived from influenza A virus. Together, our results suggest co-evolution of KIR with HLA maximizes innate immune cell function in Oceanians.
Selection drives the evolution of genes involved in the response of the human vaginal epithelium to Lactobacillus crispatus

Omar E Cornejo

Omar Eduardo Cornejo
University of California Santa Cruz (USA)

Differences in the vaginal microbiome are associated with differences in risk to sexually transmitted diseases, pre-term birth and bacterial vaginosis. Five main types of microbial communities are widely recognized in the literature, four of which are enriched with different species of Lactobacillus, and one of them is comprised of a diverse assemblage of species found at intermediate abundances. Microbial communities enriched by Lactobacillus crispatus are commonly associated with positive health outcomes, while diverse communities are commonly associated with increased risk to disease. We have performed experiments co-culturing immortalized vaginal epithelial cells with Lactobacillus crispatus bacteria. Our experiments have shown that the presence of L. crispatus drives the upregulation of genes involved in the keratinization and cornification of the stratum corneum as well as genes involved in the innate immune response. Because of the association between Lactobacillus enriched communities and good health outcomes, we further explored if genes involved in the response to Lactobacillus crispatus present signatures of positive selection. We find that a small fraction of genes differentially regulated by Lactobacillus are under selection in human populations, narrowing the set of genes that we anticipate might have been driven by selection in a microbe-mediated manner. Our work shows that selection shapes the evolution of genes involved in microbe mediated risk to disease.
Why aren’t all E. coli resistant to antibiotics?

Pleuni Pennings

Presented by self
San Francisco State University

Drug resistance is a problem in many pathogens, including viruses, bacteria, fungi and parasites (Murray et al. 2022). While overall, levels of resistance have risen in recent decades, there are many examples where after an initial rise, levels of resistance have stabilized (Krieger et al. 2020; Colijn et al. 2009; Diekema et al. 2019; Rhee et al. 2019; Rocheleau et al. 2018). The stable coexistence of resistance and susceptibility has proven hard to explain (Krieger et al. 2020; Colijn et al. 2009; Cobey et al. 2017; Kouyos, Klein, and Grenfell 2013; Blanquart et al. 2018). Here, I show that a simple stochastic model, mathematically akin to mutation-selection balance theory, can explain several key observations about drug resistance: (1) the stable coexistence of resistant and susceptible strains (2) at levels that depend on population-level drug usage and (3) with resistance often due to many different strains (resistance is present on many different genetic backgrounds). The model works for resistance due to mutations or horizontal gene transfer (HGT). It predicts that new resistant strains should continuously appear (through mutation or HGT and positive selection within treated hosts) and disappear (due to the fitness cost of resistance). The result is that while resistance is stable, which strains carry resistance is constantly changing. I used data from a longitudinal genomic study on E. coli in Norway to test this prediction for resistance to five different drugs and found that, consistent with the model, most resistant strains indeed disappear quickly after they appear in the dataset.
Human complex traits, including common diseases, are highly polygenic. Almost universally, the genetic component of the traits is governed by myriads of common non-coding variants of individually small effects with the minor contribution of rare coding variants of larger effect sizes. It is unclear how the balance of evolutionary forces maintains stable variation in complex traits in the human population. The mechanistic biology underlying the phenotypic variation is equally mostly unknown. We tested existing theoretical models of population genetics using massive GWAS datasets. We relied on two statistics: first, on the joint distribution of risk allele frequency and effect sizes; second, on effect size correlations between proximal variants. Surprisingly, our analysis supports the models of stabilizing selection favoring intermediate values of genetic liability even for disease phenotypes. This analysis also generates specific hypotheses regarding the density and directions of phenotypically important mutations.
Dissecting the Spatial Dynamics of Pseudomonas aeruginosa Persistence in Cystic Fibrosis under CFTR Modulator Therapy

Tongqiu Jia, Sam Durfey, Pradeep Singh, Alison Feder
University of Washington (USA)

During chronic infections in people with cystic fibrosis (CF), Pseudomonas aeruginosa (Pa) diversifies into genotypically and phenotypically distinct subpopulations across lung regions. While CFTR modulator therapies reshape lung environments and reduce bacterial loads, they often fail to eliminate Pa. Through whole genome, spatially-resolved, longitudinal sequencing of nine people with CF harboring Pa lung infections, we sought to resolve the basis of Pa persistence through CFTR modulators spatially and genotypically. We found that the number of surviving bacterial lineages through therapy varied among participants (2-36 lineages), suggesting widely varying bottleneck sizes. To disentangle environmental factors associated with persistence, we applied Bayesian phylogeographic approaches to infer time-calibrated phylogenetic trees jointly with bacterial intra-lung migration histories. We fit logistic regression models to explain how characteristics of the pre-treatment lung environment (bacterial density, lung damage, inflammation, and spatial position) associate with lineage persistence and found that regions with higher damage pre-treatment were more likely to possess persistent lineages. Notably, persistent lineages were not spatially confined and frequently recolonized other putatively cleared regions. For regions with multiple persistent lineages, surviving bacteria were genotypically diverse, suggesting that environment might drive lineage survival above genotype. Our findings highlight the role of pre-treatment lung environment for predicting Pa persistence, and suggest migration between lung regions shapes the identity and structure of Pa populations post-treatment. Overcoming these spatial dynamics are essential for developing targeted therapeutics to better manage infection and underscore the importance of spatial factors in pathogen responses to therapy.
Cancer involves a number of different changes that lead to abnormal and uncontrolled cell growth. Analogous to Darwinian evolution in the origin of species, cancer development is based on two constitutive processes: the continuous acquisition of heritable genetic variation in individual cells by more or less random mutation, and natural selection acting on the resulting phenotypic diversity. This results in a high mutational, biochemical, and histological tumor heterogeneity that makes driver mutation identification very challenging. The arrival of second-generation sequencing technologies has allowed us to produce a huge amount of data, including the whole genomes from cancer patients. Together with this trend, we are seeing the emergence of personalized medicine, which aims to maximize the value of data from individual patients. Historically, patients have been stratified according to their mutated driver genes or driver mutations but in many cases, this clustering does say little about their treatment response. Our approach exploits the power of evolutionary and big tumor data to cluster patients by cancer aetiology. We use whole genome information combined with the functional impact of the mutations learned from genetic variation seen on different evolutionary timescales to underpin the different commonly affected biological processes. Furthermore, this has revealed clusters with different drug responses. In summary, this method enables us to identify the characteristics that account for the differences between tumors, suggest which drug or treatment is more likely to be effective for a given patient, and to explain why.
Molecular evolutionary mechanisms of acquisition of anticancer drug-resistance in lung cancer.

Yosuke Seto

Yosuke Seto, Naoya Fujita, Ryohei Katayama
Japanese Foundation for Cancer Research (Japan)

Lung cancer is the leading cause of cancer death in the world. Activating mutation of epidermal growth factor receptor (EGFR) which induces constitutive activation of EGFR signaling pathway is a major oncogenic driver mutation of lung cancer. Inhibitors against EGFR tyrosine kinase (EGFR-TKIs) dramatically improved the prognosis of EGFR mutation-positive lung cancer. However, acquired resistance against EGFR-TKI after the initial clinical response is widely observed. To date, multiple mechanisms of EGFR-TKI resistance including bypass pathways (e.g., MET and HER2 signaling pathway) and secondary resistance mutation (e.g., T790M) have been reported. Especially because the de novo T790M resistance mutation emerges in initially T790M-negative drug-tolerant cancer cells during EGFR-TKI treatment, it is important to elucidate the process of the emergence of resistance mutations in the drug-tolerant cancer cells during the EGFR-TKI therapy. To elucidate how the cancer cells survive under EGFR-TKI treatment, we conducted single-cell RNA-seq and single-cell ATAC-seq analyses using EGFR mutation-positive lung cancer patient-derived cells (PDCs). We observed the epithelial-to-mesenchymal transition (EMT)-like morphological change under EGFR-TKI treatment. Consistent with the observation, our single-cell analyses showed EGFR-TKI treatment-induced dynamic cell differentiation, and differential gene expression and chromatin accessibility analyses revealed induction of anti-apoptosis response and NF-kappa B mediated inflammatory response. Combined with the study of EGFR-TKI selection, we found that EGFR-TKI-induced cytidine deaminase plays a critical role in the emergence of resistance mutations.
S17 - Aging from a multidisciplinary overview: evolution, longevity and biomedicine.
Fungal microorganism, including those that inhabit our bodies and cause infection, often undergo stationary phases characterized by cell-cycle arrest. The capacity to survive in this state is known as the chronological lifespan—a widely used model to understand postmitotic cellular aging—allowing organisms to reenter a proliferative phase and to survive in their natural environments. Despite its relevance, we know little about the diversity of chronological lifespan among species and how it relates to the specific environments in which they live. In this talk, we will present the implementation of novel experimental approaches to portray the chronological lifespan of Candida species, a group of microorganisms best known for causing a variety of human diseases, but mostly composed of non-pathogenic yeasts. Our results show great lifespan diversity among these fungi and changes in longevity by caloric restriction. Lifespan is not directly associated to pathogenicity, as there are short- and long-lived species among the usual inhabitants of the human body. Surprisingly, lifespan does not vary according to the phylogenetic position of these species, suggesting that stationary-phase survival is a plastic trait in evolutionary terms. We will present analysis of specific gene knockout mutants shedding light into the molecular underpinnings of lifespan diversity, with possible implications for the control of these clinically important fungi.
Comparative Analysis of primate genomes to unveil the genetics of maximum lifespan determination.

Fabio Barteri

Fabio Barteri, Noelia Rodriguez Perez, Miguel Ramon Alonso, Alejandro Valenzuela Seba, David De Juan, Arcadi Navarro Cuartiellas
BarcelonaBeta Brain Research Center, Center for Genomic Regulation (CRG), Department of Medicine and Life Sciences, IBE, Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Pompeu Fabra. (Spain), Institute of Evolutionary Biology (UPF-CSIC), Pasqual Maragall Foundation (Spain), Pasqual Maragall Foundation. (Spain), The Barcelona Institute of Science and Technology. (Spain), Universitat Pompeu Fabra . (Spain), Universitat Pompeu Fabra. (Spain)

Maximum life span (MLS) is a trait that varies widely across primates – from 2 to 60 years when excluding humans – but remains consistent within each individual group. The genetic architecture of MLS involves numerous genes and regulatory elements that belong to different networks, many of which related to senescence and/or the onset of age-related diseases. Understanding the genetic background of MLS can provide insights into various aspects of the evolution of aging and may suggest interventions upon lifespan and healthspan in humans. Here, we will discuss our comparative study of MLS within the primate phylogeny leveraging an unprecedented dataset formed by the genomes of primate species that was recently released by Kuderna et al. (2023). We analyzed 16,133 protein-coding regions across the primate phylogeny, identifying genes and even individual Convergent Amino Acid Substitutions (CAAS) associated with the variation of Longevity Quotient (LQ, MLS normalized for body mass). Additionally, we investigated genes whose rates of protein evolution correlate with LQ. Our preliminary results have returned a set of 2,191 genes, the 11.6% of which (255) had been already retrieved amongst the 996 genes associated with longevity in mammals identified by Farré et al. in 2021. These genes are involved in pathways that have been associated with male infertility, chronic respiratory conditions and degenerative disorders, such as Alzheimer’s Disease and Huntington’s Disease. Our work represents an opportunity to exploit the potential of comparative genomics approaches to identify the genetic background of aging.
Accounting for common genetic variation improves chronological age prediction in African populations
Gillian L Meeks

Gillian L Meeks, Brenna M Henn, Hana Asnake, Brooke Scelza, Shyamalika Gopalan
Clemson University (USA), University of California Davis (USA), University of California Los Angeles (USA)

Aging is associated with genome-wide DNA methylation changes in humans, facilitating the development of epigenetic age prediction methods for forensic and health-care applications. However, most prior research studying these changes have focused on European-ancestry individuals, which represent only a fraction of extant human genetic diversity. We test one of the most popular epigenetic age predictors, the Horvath model, on data from central African Baka foragers, southern African ?Khomani San foragers, and southern African Himba pastoralists, and observe substantially higher mean errors in these cohorts compared to European-ancestry samples. Common genetic variants are known to influence baseline methylation levels, introducing noise to the relationship between methylation and age, and potentially affecting prediction performance. However, these interactions are not well characterized for genetically diverse African-ancestry populations. We hypothesize that this unaccounted-for variation partially explains why published age predictors perform worse in our cohorts. Supporting this, we find that 25% of the predictor loci used by the Horvath model are influenced by genetic variants that segregate in at least one of our cohorts. We show that accounting for these types of variants improves age association at Horvath’s predictor loci, and furthermore enables the discovery of 211 additional age associations. Next, we developed a novel age prediction model that corrects for the effect of nearby genetic variants on DNA methylation levels at age-associated loci. Our work shows that accounting for a broad range of human genetic diversity enables more accurate and universally-applicable epigenetic age prediction methods.
Longitudinal multi-omics data analysis of cynomolgus macaque genome throughout their lifespan reveals age-related immune patterns

Hyeon-Mu Cho

Hyeon-Mu Cho, Se-Hee Choe, Ja-Rang Lee, Hye-Ri Park, Min-Gyeong Ko, Yun-Jung Lee, Hwal-Yong Lee, Sung Hyun Park, Sang-Je Park, Young-Hyun Kim, Jae-Won Huh

(Korea, Korea Research Institute of Bioscience and Biotechnology (KRIIB), University of Science & Technology (UST), University of Science & Technology (UST) (Korea)

Despite aging research obtains huge benefits from recent technological advances, its biological mechanism remains unknown. Therefore, we investigated multi-omics data of cynomologus macaques, a laboratory monkey, well protected from environmental confounding factors other than aging. With transcriptomic data, both immune cell composition analysis and gene ontology analysis showed that the genes of innate immune cell such as myeloid cells and NK cells were upregulated while the genes of adaptive immune cell such as Th cells and B cells were downregulated as they get older. We found that the cis-regulatory regions of the upregulated genes were mostly down-methylated in aging according to the whole genome bisulfite sequencing data. Additionally, the correlation analysis indicated that the number of genes with increased expression by down-methylation on cis-regulatory region were twice as many as that with decreased expression by up-methylation. In ATAC sequencing analysis, we observed that chromatin keeps opening with down-methylation on a genome scale. Conversely, we found eight peaks that chromatin was gradually closing while aging. For instance, the peak in the intronic region of TP73 5'UTR were gradually closed with its downregulated gene expression. As further analysis is still ongoing, we will put more efforts into the correlation analysis of these multi omics data.
Evolutionary medicine, aging and longevity
Luis Miguel Francisco Gutierrez Robledo

Presented by self
Instituto Nacional de Geriatria (Mexico)

Theories of aging often focus on internal mechanisms, neglecting the crucial role of the external environment. This presentation argues for a more holistic approach, integrating environmental pressures into our understanding of aging. Secondly, with the rise of epigenetics, the information theory of aging gains traction. This theory posits that aging results from the accumulation of errors in cellular information transfer, potentially influenced by environmental factors. Finally, the "grandmothers' hypothesis" proposes that human longevity evolved in part due to the selective advantage of grandparental care. By investing in grandchildren's survival, grandparents indirectly enhance their own genetic legacy, potentially influencing the evolution of longer lifespans. By considering the interplay between environment, information theory, and human social and family networks, through grandparental investment, we can gain a richer understanding of the complex forces shaping human aging and longevity. Evolutionary medicine provides a critical lens for interpreting the interplay between environment, information, and social behaviors in the context of aging and longevity. It guides medical professionals to consider the evolutionary roots of our health and aging, ultimately leading to better strategies for promoting healthy lifespans.
Signatures of extreme longevity: a perspective from bivalve molecular evolution

Mariangela Iannello

Mariangela Iannello, Giobbe Forni, Giovanni Piccinini, Ran Xu, Jacopo Martelossi, Fabrizio Ghiselli, Liliana Milani
University of Bologna (Italy)

Longevity is a complex and poorly characterized trait: we know that some species can live longer than others, but the molecular players responsible for the extended longevity phenotype are largely unknown. In this context, bivalve molluscs can provide novel perspectives: the class Bivalvia shows the highest lifespan disparity within Metazoa, ranging from 1 to 500+ years, and includes the longest-lived non-colonial animal species known so far, the clam Arctica islandica. Bivalves therefore represent important resources to provide insights into the molecular mechanisms of longevity. In this work, we leveraged transcriptomic resources spanning thirty bivalve species and we investigated genes with signature of convergent molecular evolution in bivalves showing extreme longevity. We found that a majority of genes showing convergent evolution in long-lived bivalves constitute a continuous network of protein-protein interactions. Such network is enriched for factors with experimental support for a role in longevity in other animals. The network highlights that an integration of different genes and pathways is required for the extended longevity phenotype, and genes involved in cell proliferation control, translational machinery, and response to hypoxia seem to have a central role in lifespan extension. Our results suggest that the mechanisms underlying extended longevity are, at least partially, similar across metazoans: while some genes in the network have experimental support of a role in longevity in model species, other genes in the network may represent new possible candidates with a role in extending lifespan, both in bivalves and other animals.
Contribution of transposable elements in the sex gap longevity of different Drosophila species
Miriam Merenciano

Miriam Merenciano, Sonia Janillon, Camille Mermet-Bouvier, Nelly Burlet, Matthieu Boulesteix, Marie Fablet, Cristina Vieira, (France), Laboratory of Biometry and Evolutionary Biology (France)

In Drosophila, like in many other animal species, females tend to live longer than males, a phenomenon known as sex gap in longevity (SGL). One of the possible causes underlying this phenomenon could be related to the high number of transposable elements (TE) in the Y chromosome (toxic Y effect). TE activity is normally repressed by epigenetic mechanisms. However, it is known that this regulation is disrupted with age. Since the Y chromosome is rich in TEs, more TEs may become active in old males than in old females, generating more somatic mutations, and reducing longevity in males. In this work, we studied the natural variation in SGL in several natural populations of three different Drosophila species that vary in their TE content: Drosophila melanogaster, Drosophila simulans, and Drosophila suzukii. Furthermore, we found that the replacement of the Y chromosome between strains with different SGL reduces male lifespan over generations and thus increases SGL, suggesting an important role of the Y chromosome in male longevity. Finally, RNA-seq analysis from old and young flies suggested that there is an increased number of upregulated TE families in old samples, and more specifically in old males compared to old females, and that the total fraction of transcripts derived from repeats increase during aging depending on the species and the population tested. Overall, this work tries to better understand the genomic differences that lead to variation in longevity patterns between sexes, and emphasizes the importance of TEs in male longevity.
Do Methylation Clocks Generalize Across Admixed Populations?
Sebastián Cruz-González


Biological and Medical Informatics Graduate Program, Case Western Reserve University (USA), Case Western Reserve University School of Medicine (USA), Department of Nutrition, San Francisco (USA), System Biology and Bioinformatics, UCSF (USA), University of California, University of Miami Miller School of Medicine (USA)

Biological aging is the progressive accumulation of cellular damage leading to degeneration and eventual organismal death. Unlike chronological age, biological age varies among individuals due to factors including genetics, smoking, and diet. Methylation clocks, which link methylation levels at CpG sites across the genome to chronological age, have emerged as potential biomarkers for biological age. Accelerated biological aging, as measured by methylation clocks, is significantly associated with an increased risk of age-related ailments, such as Alzheimer’s disease (AD) and coronary heart disease. Despite this promise, existing clocks have been trained and evaluated in predominantly European ancestry cohorts and lack validation in genetically diverse populations, hindering their broader applicability. We evaluate the accuracy of several common methylation clocks and their ability to serve as biomarkers of Alzheimer’s disease risk in admixed populations from the Americas. We leverage the MAGENTA study, a large cohort of 313 diverse admixed AD patients and 308 control individuals with methylation data and genotyping data. We generate biological age predictions from the methylome for all individuals and compare their biological age to their chronological age and test the association of accelerated aging with AD status. Furthermore, we evaluate the accuracy of the methylation clocks on admixed individuals with varying levels of global European, Amerindigenous, and African ancestry. This study addresses the lack of diversity in the development and application of genomic tools for precision medicine and assesses the utility of methylation clocks as biomarkers for age-related diseases across diverse populations.
Identifying de novo structural variation in the aging male germline using long reads

Stacy Li

Stacy Li, Peter Sudmant
Center for Computational Biology, Department of Integrative Biology, UC Berkeley (USA)

Aging is an emergent phenomenon hallmarked by the deterioration of physiological processes over time. Genomic instability occurs throughout normal and pathological aging, typically driven by unresolved or erroneously remediated DNA damage. De novo germline mutations are of critical importance to reproductive success in aging: Previous work using short-read sequencing in pedigree-based studies identified a parental age-associated increase in small de novo variants, with a strong bias from the paternal germline. De novo structural variants (dnSVs) have profound impacts on health, disease, and genome evolution, but remain challenging to characterize due to the technical limitations of short-read sequencing. To address this, we conducted a study using PacBio HiFi long-read sequencing to identify dnSVs in bulk sperm samples from eight donors of varying ages. Under this framework, we anticipate recall of clonal dnSVs (shared amongst sperm descended from a common progenitor), with the potential to recover singleton/doublet dnSVs arising in meiosis. To this end, we developed a method to identify de novo retrotransposition events, which can be characterized within the span of a single read. We identified multiple high-confidence AluYb8, AluYa5, and L1HS retrotransposition events. We estimate between 2.1 to 10.2 events per 100 individual cells, with the oldest donor sample yielding the greatest number of events. Furthermore, several L1 events originally raised as insertions were revealed to be complex clonal dnSVs following manual examination. These results suggest that long-read sequencing is a promising method to survey dnSVs in the aging germline.
Why do males and females have different lifespan? I will focus on ecology and life histories of vertebrates, and evaluate the different causes of male vs female longevity. Sex differences in life-styles, cost of reproduction and sexual selection have been proposed to explain sex difference in longevity, but do data support these hypotheses? Recently the sex determining system is also emerging as one of the predictors of life-span differences, although the processes that generate this association will need to be investigated. I will also overview the implications of sex different longevity, especially from the perspectives of social behaviour and breeding systems. The relative frequencies of males and females in a population is one of the key demographic variables, and we are beginning to understand the wide-spread implications of skewed population sex ratios. Importantly, with recent global changes in climate and the impacts of humans on the planet, the life-styles, ecology and behaviour of many organisms will change so that we’ll need to focus on understanding the implications of these anthropogenic changes on longevity and sex different reproductive strategies.
Clonal Hematopoiesis of Indeterminate Potential and HIV infection synergistically affect all-cause mortality
Valeria Timonina

Valeria Timonina, Konstantin Popadin, Jacques Fellay
Immanuel Kant Baltic Federal University (Russian Federation), Lausanne University Hospital and University of Lausanne (Switzerland), Swiss Institute of Bioinformatics (Switzerland), attendancesmbe2024@gmail.com (Switzerland), École polytechnique fédérale de Lausanne (Switzerland)

Understanding the mechanisms of normal, healthy aging can be enhanced by studying the cases of accelerated aging. It is well established that both Clonal Hematopoiesis of Indeterminate Potential (CHIP) and human immunodeficiency virus (HIV) are linked to accelerated aging, as evidenced by increased all-cause mortality risks. Previous research has identified a higher prevalence of CHIP in People Living with HIV (PLWH). In this study, we investigated whether CHIP and HIV interact synergistically to exacerbate aging outcomes. Utilizing data from the Swiss HIV Cohort Study of people living with HIV (N = 1889), we found that CHIP carriers exhibit a hazard ratio for all-cause mortality of 1.35, which is in the range of the ratios of 1.18-1.4 in the general population reported in the literature. To corroborate these findings, we analyzed the effect of both HIV and CHIP on all-cause mortality in age- and sex-matched PLWH (N=481) and control individuals (N=2405) from the UK Biobank. Our results revealed a significant interaction between HIV and CHIP (Hazard ratio for HIV = 1.98, p<0.001; Hazard ratio for CHIP = 1.16, p=0.5; Hazard ration for HIV and CHIP interaction = 2.99, p=0.05), further elevating the risk of all-cause mortality. Investigating the underlying mechanisms of an interaction (potentially through increased inflammation) and exploring similar interactions of CHIP with other infections represents a promising avenue for future research on aging. This research was conducted using the UK Biobank Resource under Application #84415.
S18 - One Health and microbial evolution: New ideas and perspectives.
Functional consequences of reductive protein evolution in a minimal eukaryotic genome

Aaron W Reinke

Jason Jiang, Rui Qu, Maria Grigorescu, Winnie Zhao, Aaron W Reinke

University of Toronto (Canada)

Microsporidia are parasites with the smallest known eukaryotic genomes. The extent of protein loss in these organisms has been well documented, but much less is known about how compaction of microsporidia proteins affects their function. Taking a comparative genomic approach, we identified microsporidia orthologs of budding yeast proteins and show that these orthologs are enriched for essential yeast genes. We show that the median microsporidia protein is 21% shorter than its yeast counterpart and although extensive protein loss occurred after the divergence of microsporidia, reduced protein sizes were already present in microsporidian relatives. Microsporidia proteins are shorter through reduced domain lengths, diminished linker lengths, and domain loss, with 21% of microsporidia orthologs having lost domains present in yeast. On average, 34% of microsporidia orthologs have lost C-terminal residues essential for function in yeast, including 13 essential domains lost per genome. We also found that microsporidia display distinct phylogenetic patterns of domain loss, with losses occurring in a clade-specific manner. To investigate conservation of function, we used yeast complementation assays to test orthologs from several microsporidia species and their relative Rozella allomycis. These experiments reveal that most microsporidia proteins cannot complement their yeast orthologs, the ability to complement is about three-fold less than observed for R. allomycis orthologs, and proteins that do not complement are more reduced in length than their yeast orthologs. Altogether, our results demonstrate the drastic reduction of microsporidia proteins and show that these reductions have resulted in functional divergence from their fungal ancestors.
Multi-omics reveals host-microbiome interaction in the oral cavity underlying obesity

Aashish R Jha

Ahmed A Shibil, Tsedenia W Denekew, Anique R Ahmad, Salah Abdelrazig, Christopher E Leonor, Lina Utenova, Guihao Zhang, Mamon AbdelBaqi, Yashaswi Malla, Mohammed Arshad, Marc Arnoux, Nizar Drou, Raghib Ali, Shady A Amin, The UAE Healthy Future Study Investigators Group Idaghdour, Youssef Idaghdour, Aashish R Jha

1Genetic Heritage Group, 5Center for Genomics and Systems Biology, Bioinformatics Core, Center for Genomics and Systems Biology, Core Technology Platform, Genetic Heritage Group, Marine Microbiomics Lab, New York University Abu Dhabi (UAE), Program in Biology, Public Health Research Center

Gut microbiome alterations have been causally linked to several human diseases, but whether perturbations of the oral microbiome – the second largest microbial ecosystem in the human body – contributes to human health remains markedly underexplored. Here, we first applied 16S rRNA sequencing to find associations oral bacteria and 14 cardiometabolic traits in a cohort of ~700 Emirati nationals from the United Arab Emirates. We found several features of the oral microbiota was strongly linked with obesity in this population. To elucidate the mechanism by which oral microbiota contributes to obesity, we implemented a multi-omics analysis framework in a case-control subset of ~200 individuals. Integrated analysis of metagenomics and metabolomics revealed functional alterations in the oral microbiome in obesity contributes to several hallmarks of obesity. Specifically, oral microbes in obese individuals lead to increased uridine production, which is associated with feeding behavior. They also metabolize dietary carbohydrates into lactate, which leads to increase in visceral fat deposit. Finally, they produce obesogenic amino acids, such as methionine, cystine, and, threonine. Conversely, several pathways associated with Vitamin-B and heme production is depleted in obesity. Furthermore, several of the bacterial elevated salivary metabolites were positively associated with several obesity associated clinical markers in blood, indicating these metabolites can transcend the well vascularized oral epithelia and interfere with host physiological process. This study robustly links oral microbiome alterations with host metabolic changes, indicating that mechanistic understanding of host-microbiome interactions in the oral cavity may provide a promising avenue for obesity intervention.
The evolutionary correlates of viral host jumps

Cedric C.S. Tan

Cedric CS Tan, Lucy van Dorp, Francois Balloux
UCL Genetics Institute, University College London (United Kingdom)

Most emerging and re-emerging infectious diseases stem from viruses that naturally circulate in non-human vertebrates. When these viruses cross over into humans, they can cause disease outbreaks, epidemics and pandemics. While zoonotic host jumps have been extensively studied from an ecological perspective, little attention has gone into characterizing the evolutionary drivers and correlates underlying these events. To address this gap, we harnessed the entirety of publicly available viral genomic data, employing a comprehensive suite of network and phylogenetic analyses to investigate the evolutionary mechanisms underpinning recent viral host jumps. Surprisingly, we find that humans are as much a source as a sink for viral spillover events, insofar as we infer more viral host jumps from humans to other animals than from animals to humans. Moreover, we demonstrate heightened evolution in viral lineages that involve putative host jumps. We further observe that the extent of adaptation associated with a host jump is lower for viruses with broader host ranges. Finally, we show that the genomic targets of natural selection associated with host jumps vary across different viral families, with either structural or auxiliary genes being the prime targets of selection. Collectively, our results illuminate some of the evolutionary correlates of viral host jumps that may contribute to mitigating viral threats across species boundaries.
Host ACE2 Sequence Evolution Identifies SARS-Related Sarbecovirus Reservoirs

Eric X Gan, John A Frank, Viraj Jansari, Akiko Iwasaki
Stanford University (USA), Yale University (USA)

SARS-CoV-2 (SCV2) is a zoonotic virus and the etiologic agent of the ongoing COVID-19 pandemic, which has severely impacted public health on a global scale. The documented spillover of SCV2 from humans into animals including felines, mink, and white-tailed deer highlights the critical need for improved surveillance and identification of virus reservoir species. Viruses that have persisted within host lineages for significant periods of evolutionary time are likely to have coevolved with host receptors, leaving behind a signature of selection-dependent sequence variation. We hypothesized that evolutionary sequence analysis of host virus receptors could be leveraged to identify reservoir animal lineages. Using SCV2 as a model, we performed a dN/dS analysis on 572 ACE2 sequences across all vertebrate orders to search for positive selection at residues known to be targeted by SCV2. Consistent with the literature, our ACE2 analyses identified the order Chiroptera as exhibiting positive selection consistent with persistent circulation of SARS-related sarbecoviruses. Because Chiroptera are the second most species-rich mammalian order, we subsequently performed bat lineage-specific dN/dS analyses of ACE2. We determined that the horseshoe/roundleaf (Rhinolophus spp./Hipposideros spp.) bat lineage displays a distinct pattern of positive selection at clusters of residues with known sarbecovirus spike interactions. These data suggest that SARS-related sarbecoviruses have circulated in and applied selective pressure to this lineage for approximately 64 million years. The application of selection-based evolutionary analyses to pinpoint viral reservoir species could be vital in informing future viral surveillance efforts.
Campylobacter jejuni is a commensal bacterium that colonises livestock animals, and is responsible for more than 80% of campylobacteriosis cases in humans. Its high genome plasticity and recombination rate allow it to adapt to multiple hosts and leads to the rapid emergence of lineages, some of which can be antibiotic resistant. To curb an increase in antimicrobial infections in humans, the State of California enacted Senate Bill 27, which prohibits the preventative use of antibiotics in livestock. Our aim in this study was to examine the population structure of Campylobacter in California, assess its transmission between animal hosts and humans, and to consider possible effects of the policy implemented by SB27. We sequenced 69 human clinical isolates of C. jejuni from California and placed them in a concentric context of human and animal isolates. By comparing groups of isolates, we detected significant levels of differentiation between the human isolates post-SB27 and animal isolates from California. Through phylogenetic reconstruction we demonstrated that the human C. jejuni clinical isolates are largely derived from food animal sources. By identifying the genomic regions that were contributing to population differentiation among the different host groups, we were able to identify protein variants potentially responsible for host adaptation and propensity to cause infection in humans. By describing the population structure of Campylobacter in animal food sources, and the relationship of these populations to human-infecting isolates, this study attempts to identify what genes drive host adaptation and examine how policy-making shapes bacterial populations.
Evolutionary genomics and the one health approach to AMR
Fernando González Candelas

Presented by self
Unidad Mixta Infección y Salud Pública FISABIO/Univ. Valencia (Spain)

Antimicrobial resistance (AMR) is currently one of the most serious threats for public health. AMR is the foreseeable outcome of a simple evolutionary process, natural selection. As such, it has been evolving for billions of years. Yet, in the last decades we have interfered at such an extent with this natural process, through an inadequate and abusive use of antimicrobials, that the dynamic equilibrium between producers and defenses is being displaced towards the later. Furthermore, the combination of hot-spots for AMR and pervasive presence of horizontal gene transfer mechanisms in most bacteria is driving the increasing incidence of multidrug (MDR) and extremely-drug resistant (XDR) pathogens, pointing to a back-to-the-past era of pre-antibiotic medicine. Evolutionary genomics provides the ideal conceptual and methodological framework to address some of the pressing questions posed by AMR namely which are the genes driving the highest levels of resistance, where they originated, where are they carried, what can be done to reduce their incidence, and so on. My research group has been working on antimicrobial resistance at several levels, from the analysis of nosocomial outbreaks of MDR and XDR bacteria to the study of ARGs and their distribution in time, space, and ecosystems. I will present results from the analysis of carbapenemases, enzymes that block the action of some of the most useful last resort antibiotics for many infections, in hospitals, wastewaters, and agricultural products in the Comunitat Valenciana region (Spain).
Characterizing Pareto fronts: Trade-offs in the yeast growth cycle constrain adaptation
Jason Tarkington

Jason Tarkington, Gavin Sherlock
Stanford (USA)

Mapping trait space accessible by single mutations reveals the constraints imposed on organismal fitness. The fitness of an organism in an environment is often dependent on more than one phenotype, for yeast these include fermentation, respiration, and stationary phase performance. Such fitness-related phenotypes can be constrained due to the pleiotropic effect of other fitness related phenotypes, resulting in a trade-off. Previous work has shown that following evolution in a glucose containing carbon limited media a pareto front, which is indicative of an underlying trade-offs, emerges between stationary phase and respiration and between respiration and fermentation but, not between stationary phase and fermentation. I aim to understand why such trade-offs emerge among some fitness-related traits but not others by allowing barcoded yeast to evolve in a non-fermentable carbon source with varying amounts of time spent in stationary phase. This eliminates selection for fermentation entirely and creates a gradient of selection for respiration and stationary phase between the two-day transfer regime and the 10-day transfer. Under these conditions stationary phase performance may be free to increase unconstrained by fermentation performance resulting in the emergence of a Pareto front between these components of fitness. In addition to the phenotypic analysis, I have characterized the molecular basis of 480 underlying adaptive mutations that emerge to understand the mechanistic basis of trade-offs in the growth cycle and have identified distinct genomic signatures adaptation that depend on time spent in stationary phase.
Creating Online Phylogenetic Resources for the Mycobacterium Tuberculosis Complex
Lily M Karim

Lily M Karim, Francisco J Martinez-Martinez, Ash O'Farrell, Angie Hinrichs, Iñaki Comas, Russell Corbett-Detig
CSIC (Spain), Instituto de Biomedicina de Valencia (IBV), Tuberculosis Genomics Unit, UC Santa Cruz Department of Biomolecular Engineering and Bioinformatics (USA), UC Santa Cruz Genomics Institute (USA)

During the SARS-CoV-2 pandemic, online phylogenomics played a crucial role in genomic surveillance, contact tracing, and identifying new variants. This method involves continuously updating a phylogeny of all available samples and integrating new samples into the existing tree on demand, enabling rapid research and response despite the large volume of genomic data generated. Extending online phylogenomics to other pathogens holds the potential to significantly enhance research speed and capabilities for various illnesses. The Mycobacterium tuberculosis complex (MTBC) is a group of closely related bacteria responsible for 10 million tuberculosis infections worldwide. Current tuberculosis research typically involves subsampling from available genomes, which limits insights into the deep-time evolutionary histories and nearest neighbors of samples. Previous limitations in storage and computational capacity hindered the creation and maintenance of a global phylogenetic tree for the MTBC. With the emergence of pandemic-ready phylogenetic software tools like USHER, we are now hosting and maintaining an up-to-date global phylogeny of the MTBC. This comprehensive phylogeny facilitates tracing new outbreaks, identifying pathogen lineages and antibiotic resistance, and analyzing genomic divergence within the complex. Incorporating all available sequences and associated metadata enables a detailed examination of mutation and antibiotic resistance variation among lineages. Rather than relying on sub-sampled datasets with approximate neighboring sequences, the closest related samples for each new sequence can be used for transmission information and outbreak analysis within local communities. Thus, the up-to-date, comprehensive MTBC phylogeny enables the most effective response to new tuberculosis infections worldwide.
Developing Robust Machine Learning Models for AMR Prediction in Pseudomonas aeruginosa
Lucia Grana Miraglia

Lucia Grana-Miraglia, Pauline W Wang, Nadia Morales-Lizcano, Allison McGeer, David S Guttman
Center for the Analysis of Genome Evolution and Function (Canada), Center for the analysis of Genome Evolution and Function (Canada),
University of Toronto (Canada)

The progression of antimicrobial resistance (AMR) continues to be a significant challenge for worldwide public health. Increasing AMR means fewer treatment options, which highlights the need for improved diagnostic methods to allow rapid detection of resistance patterns. The availability of large-scale databases containing sequenced resistant microbial isolates represents a significant milestone in our study of AMR, underscoring the importance of advanced computational tools to thoroughly analyze the genetic basis of antimicrobial resistance. Machine learning and associated statistical genetic approaches are widely applied in microbiology for predicting traits of interest, such as AMR, and unveiling the underlying genetic basis. Although both supervised and unsupervised machine learning methods have proven valuable in these areas, there are significant inherent challenges in deploying machine learning within the field of genomics. Among the most important, high dimensional data, sampling biases and the non-independent evolutionary backgrounds among samples (population structure). Pseudomonas aeruginosa (PA), responsible for causing both severe acute and chronic persistent infections, presents significant challenges. This opportunistic pathogen possesses a high level of both intrinsic as well as acquired resistance. We used a large and representative dataset of 2311 PA isolates, collected from seven different facilities in the Great Toronto Area, to predict AMR from genomic data. Using different learning algorithms, feature extraction and selection techniques, we were able to handle high dimensionality and reduce the effect of population structure to obtain high performing and generalizable models. Obtaining accurate predictions is essential for forecasting treatment outcomes, designing effective interventions, and allocating societal resources appropriately.
Phenology and environmental characteristics are important factors driving wildlife microbiomes. However, studies addressing the effect of both phenology and environmental variation in wildlife microbiomes are scarce. Here, we sequenced the hypervariable region V3-V4 of the 16S rRNA gene to profile the nest soil and feather microbiota of Magellanic penguins across three phenological times (i.e., courtship, egg incubation, and chick rearing) and five islands with contrasting environmental characteristics in the Magellan Strait, Chile. In addition, we sequenced the global metatranscriptome of penguin feathers (one metatranscriptome per island/time). Preliminary analysis indicates that Moraxellaceae bacteria are the most abundant taxa in the feather microbiota across all conditions. Moreover, we detected seven core bacterial taxa associated with penguin bodies across all islands. Psychrobacter was the core taxa with the highest abundance across all conditions. Furthermore, we also detected island-specific core bacterial taxa in four islands. Among these core taxa, Tomitella, Paeniglutamicibacter, Gelidibacter, Chryseobacterium, and Flavobacterium were shared between these. In turn, feather microbiota compositional patterns differed across islands. It consistently differed from the nest soil and across phenological stages in Contramaestre and Rupert Islands, whereas it had more variable patterns in the other islands. Ongoing analyses are being carried out to test the processes influencing feather microbiota composition, and the global function of penguin feather microbiota through metatranscriptomics.
Within-patient evolution of ciprofloxacin resistance in Pseudomonas aeruginosa across a large-scale clinical trial
Matthew J Shepherd

Matthew J Shepherd, Niamh E Harrington, Anastasia Kottara, Kendall Cagney, Taoran Fu, Steve Paterson, Joanne L Fothergill, James D Chalmers, Dylan Z Childs, Michael A Brockhurst
University of Dundee (United Kingdom), University of Liverpool (United Kingdom), University of Manchester (United Kingdom), University of Sheffield (United Kingdom)

The antimicrobial resistance (AMR) crisis threatens to endanger modern medicine within the next 20-30 years. A key part of the problem is the emergence of AMR within patients themselves, during courses of antibiotic therapy. Whilst within-patient AMR emergence is increasingly receiving attention in the literature, investigations commonly involve a small number of patients and study single-patient cases which may poorly represent the more general patient population. Here we report finding on the evolution of ciprofloxacin resistance by Pseudomonas aeruginosa across a large-scale clinical trial. This trial utilised inhaled liposomal ciprofloxacin as therapy for patients with the lung condition non-cystic fibrosis bronchiectasis (NCFB), who suffered with P. aeruginosa infection. We isolated a total of ~25,000 independent bacterial colonies, prior to treatment and during the year-long duration of the trial, and measured the ciprofloxacin MIC and growth rates in KB broth for each of these isolates. Across the trial, ciprofloxacin MICs increased during periods of treatment and decreased during treatment withdrawal, in a pattern suggestive of resistance fitness trade-offs operating within patients. We also find that distinct phenotypic adaptive trajectories are followed in different patient cases, indicating diverse evolutionary dynamics driving resistance emergence. These findings allow us to start characterising the mechanisms through with AMR can emerge within-patients during treatment, and to ask questions on how best we could predict resistance emergence in the future.
Mangrove ecosystems, situated in the tropics, play a critical role in global weather patterns and serve as essential carbon reservoirs. However, they face imminent threats from human activities like altered hydrological regimes and extensive deforestation. The symbiotic relationship between mangrove trees and microbiota within these ecosystems is vital for mutual survival. This study delves into interaction dynamics within degraded mangrove ecosystems due to hydrological connectivity loss. Using 16S rRNA sequencing, we analyzed microbial communities across various degradation stages (conserved, moderate-degradation, and degraded) during dry and rainy seasons at three sediment depths interacting with mangrove roots. Our analysis revealed 11,469 Operational Taxonomic Units (OTUs), indicating that microbial community structure is primarily influenced by degradation degree, with seasonal variations also playing a role. Significant diversity shifts were observed within the upper sediment layers, with conserved sites dominated by Vibrionaceae and moderate-degradation sites by Halomonas and Marinomonas, known halophytes. The presence of Vibrionaceae in conserved sites suggests potential urban contamination, while microbial diversity in moderate-degradation sites correlated with the dry season. Interestingly, a core community of Firmicutes persisted in deeper sediment layers across all degradation scenarios, hinting at its potential as a seed community for ecosystem restoration. These findings provide insights into microbial community responses to human-induced stressors and underscore the role of core microbial communities in guiding restoration strategies for degraded mangrove ecosystems.
Establishing gut microbiota during early life plays a crucial role in the host's physiology and development, with potential long-term effects. In wild birds, the establishment and potential routes of transmission to colonize gut microbiota during early life are poorly known. In the Brown Booby (Sula leucogaster), a long-lived seabird that nests in Islas Marietas, Mexico we (1) characterized the bacterial microbiota diversity, composition and functionality of eggshells, chicks throughout early development (7-56 days of age) and adults, and (2) evaluated the contribution of parental transmission of bacterial taxa to the chick microbiota. We analyzed exudates of cloaca and eggshells, and samples of feces to recover amplicon sequence variants (ASVs) of the 16S rRNA gene V3-V4 region. Gut microbiota of young and adults consisted mostly of the phyla Bacillota, Pseudomonadota, Actinobacteriota, Deferrribacterota, Fusobacteriota, and Bacteroidota. For alfa diversity, the observed ASVs showed a slight decline over time during the period of chick development and were higher in eggshells than in any other stage. Frequent and dominant ASVs did not vary significantly through chick development. However, diversity values from cloacal exudates were significantly higher than those from feces samples. For beta diversity, the microbial communities differed throughout developmental stages and beta diversity values were higher in samples collected from feces than from cloacal exudates. Analyses of variation through the chick development of bacterial communities' functional traits and the importance of parental transmission to the chick gut microbiota are in progress and will be shown at the conference.
Dynamics and change in fitness effects of mutations through long-term bacterial evolution

Olivier Tenaillon

Olivier Tenaillon, Alaksh Choudhury, Zoya Dixit, Mélanie Magnan
CEA (France), CNRS (France), Genoscope, IAME, Inserm, Inserm (France), Institut Cochin, Université Paris Cité

The influence of beneficial mutations takes a center stage in the intricate dynamics of adaptation. Since the early days of population genetics, scientists have ascribed varying degrees of significance to these mutations and, by extension, their impact on populations genetic diversity. On the conservative end, proponents argue for the sporadic emergence of rare beneficial mutations exerting a transient influence on genetic diversity as they independently reach fixation. On the more assertive side, a contrasting viewpoint posits an abundance of highly impactful beneficial mutations that engage in competitive clonal interference, fostering the emergence of multiple high-frequency polymorphisms. To rigorously assess the quantitative contribution of beneficial mutations to population genetic diversity, we turned to the Long-Term Experimental Evolution. In this experiment 12 lineages of Escherichia coli evolved in the laboratory over a span exceeding 70,000 generations. Genetic barcodes were introduced to both the ancestral strain and clones from two populations sampled after 2,000 or 20,000 generations of evolution. The subsequent replay of evolution, spanning up to 600 generations with these barcoded libraries, enabled a comprehensive tracking of the emergence of beneficial mutations at low frequencies. This methodology unveiled a diverse array of patterns, encapsulating the entire spectrum of models elucidated over a century of population genetics. Notably, the findings underscored the substantial impact of beneficial mutations on population genetic diversity throughout the long-term course of adaptive evolution, and pinpoints to the molecular determinants involved in the changes of the beneficial mutation rate in the course of evolution.
Evolution of the human gut resistome across diverse lifestyles and environments

Rahgavi Poopalarajah

Rahgavi Poopalarajah, Muhammad Arshad, Aashish R Jha
The New York University (UAE)

Antimicrobial resistance is an evolutionary adaptation that is modulated by the gut microbiome. The gut microbiome harbours the gut resistome. The gut resistome refers to the antimicrobial resistance genes from pathogenic and non-pathogenic microbes in the gut. Microbes can acquire antimicrobial resistance genes on mobile genetic elements via horizontal gene transfer. The gut flora is a known reservoir of antimicrobial resistance genes, and the transfer of antibiotic resistance genes can occur from commensal gut microbes to opportunistic pathogens in the gut. There are several reports of lifestyle and environmental factors that influence gut microbiome composition and function, yet changes in the gut resistome due to lifestyle and environment are less known. Here, we evaluate the gut resistomes of global populations across lifestyle gradients and environments. Initially, by metagenomic sequencing of 97 fecal samples from healthy and prediabetic individuals, aged 18-70 years, we characterized the gut resistomes of the Western community. Our preliminary findings show that in spite of health status, Western adults constitute a high prevalence of glycopeptide resistance genes (~68% of total antimicrobial resistance genes) in the gut. Surprisingly, we discovered many vancomycin resistance genes. Vancomycin is last resort antibiotic against gram-positive bacteria. Notably, Eubacteriales and Clostridiales contained four or more vancomycin resistance genes, of which vanT was the most abundant gene type. Future work will include identifying gut resistome signatures linked to diverse lifestyles and environments. To conclude, our results may provide more insights into the evolution of drug resistance in the host gut.
Lateral transduction (LT) is a powerful form of horizontal gene transfer involving prophage-mediated transfer of large regions of the bacterial chromosomal downstream of integration sites via delayed prophage excision. Staphylococcus aureus, a major bacterial pathogen, has a strongly conserved core genome with phage integration (attB) sites that are distributed in a way that could enable around 60% of the S. aureus chromosome to undergo LT-mediated mobilisation. LT can both introduce new genes while also removing mildly deleterious mutations and parasitic mobile genetic elements via homologous recombination. Based on the distribution of attB sites and packaging directionality, the S. aureus genome can be differentiated into two regions: the attB-containing R1 region where LT activity occurs, and the attB-free R2 region predicted to be unaffected by LT. In order to examine if R1 and R2 have distinct sequence or chromosomal characteristics reflecting LT v no LT activity, we examined the distribution of pseudogenes, insertion sequences, core and accessory genes, and the genes under positive selection in R1 v R2 regions. A higher frequency of core genes and core genes under positive selection were found in R1. Conversely, pseudogenes, insertion sequences, accessory genes and accessory genes under positive selection were more frequently located in R2. These findings are consistent with the role of LT in removing deleterious mutations, parasitic elements, and maintaining gene conservation, with important impacts on S. aureus genome organisation and diversity.
Plants employ an intricately interplay between responses to abiotic and biotic stresses to cope with environmental challenges such as drought, salinity, extreme temperatures, radiation, and attacks from pests and pathogens. These interactions often result in synergistic or antagonistic effects on plant growth, development, and survival. Plant hormones play crucial roles in regulating responses to both types of stresses. There is significant crosstalk between different hormonal signaling pathways, allowing plants to integrate responses to multiple stresses. However, the ongoing situation of accelerating global climate change represent a challenge for plants that goes in parallel with an increase in the number of new emerging phytopathogens with devastating agronomical results. An alteration in plant homeostasis results in changes in the effectiveness of antiviral responses. Such changes generate new selective pressures upon viral populations, which result in changes in viral fitness and in virulence. Here, I will review recent observations from several plant virus pathosystems which illustrate the interplay between changes in abiotic stresses and virus evolution. The outcome of such interplay ranges from the evolution of more pathogenic viral strains, with expanded host ranges, to the evolution of commensal relationships in which evolved viral strains get a benefit in terms of accumulation while improving plant's tolerance to stresses.
Horizontal gene transfer (HGT) is the acquisition of genetic material outside of a parent-offspring relationship. While it has been detected across all domains of life, its relevance in bacteria is of critical concern if we think of the role it plays in, for example, antimicrobial resistance. From a basic perspective, a horizontally acquired gene may have a positive or negative impact on the bacterium’s fitness, depending on its position in the recipient’s molecular circuitry, and on the energetic cost on a cell via the demands at the DNA/RNA/protein levels that it imposes. Despite being one of the drivers of evolution, to this day, a large-scale computational model of the effect of HGT on bacterial metabolic fitness is lacking. How does an HGT event affect metabolic fitness in the light of this demand/benefit trade-off? Here we developed a computational pipeline to simulate the random transfer of DNA fragments among genomes and estimate their (positive or negative) impact on microbial metabolic phenotypes using constraint-based modelling. We used a set of randomly depleted genome-scale metabolic models (GSMM) to increase its evolvability space. We then simulated random transfers of DNA fragments from 21 species of Bacteria (donors) to this set of GSMMs of E. coli (recipients) and evaluated the impact of such new acquisitions on the recipients’ metabolic fitness. By doing so, we identified trends (involving donor-recipient phylogenetic distance, fragment length, etc.) that illuminate on the role of HGT in microbial evolution and on the factors that influence it.
The evolutionary origins of human ACE2 receptor binding among bat sarbecoviruses

Tyler N Starr

Tyler N Starr, Caroline J Craig, Ashley L Taylor
University of Utah (USA)

Many human viruses trace their origins to zoonoses from animal reservoirs, but we lack clear models for how and why animal viruses evolve molecular traits that facilitate human spillover. Such information would improve prediction of future zoonoses and development of therapeutic reagents with prospective breadth of coverage against the range of animal viruses that might seed future epidemics. SARS-related coronaviruses (also known as sarbecoviruses) circulate in Rhinolophus bats across Asia, Europe, and Africa, but two human spillovers in the past 20 years—SARS-CoV-1 in 2003 and SARS-CoV-2 in 2019—have caused devastating outbreaks. We combined phylogenetic analysis with high-throughput biochemical binding assays to chart the evolution of ACE2 binding specificities among bat sarbecoviruses, revealing that high-affinity binding to human ACE2 evolved as a latent trait in an ancestral bat sarbecovirus long preceding human zoonosis. Here, we seek to understand how and why these bat sarbecoviruses fortuitously acquired this trait that facilitates zoonosis. We characterized paths of evolution within ancestrally reconstructed spike receptor-binding domains via combinatorial mutagenesis. We identify the specific historical substitutions that enabled acquisition of human receptor binding and highlight the divergent genetic and biophysical mechanisms by which human ACE2 binding evolved convergently in bat sarbecoviruses from Asia and Europe. We present preliminary data supporting a model of bat:virus coevolution that explains why evolution of viruses in bats drives fortuitous binding to the human receptor. Together, this work identifies ecological and molecular evolutionary forces driving the acquisition of latent traits that enable zoonosis.
S19 - Evolution of microbial communities: is the sum of parts greater than the whole?
An Integrated Pipeline of Machine Learning Techniques for Analysing Microbial Communities

Adamandia Kapopoulou

Adamandia Kapopoulou, Ianis Vilela, Russel Jasper, Jessica Rieder, Stephan Peischl, Irene Adrian-Kalchhauser, Claudia Bank
Department of Infectious Diseases and Pathobiology, Division of Theoretical Ecology and Evolution, Institute for Fish and Wildlife Health, Institute of Ecology and Evolution, Interfaculty Bioinformatics Unit, Swiss Institute of Bioinformatics (Switzerland), University of Bern (Switzerland), Vetsuisse Faculty

The significance of microbial communities in maintaining natural ecosystems at equilibrium and human and other species’ health is indisputable. With the emergence of numerous microbial datasets, there is growing interest in characterising and understanding the factors that influence microbial composition. Specifically, there is a need for robust methods that relate differences in relative microbial abundance to environmental factors. Recirculating Aquaculture Systems (RAS) are controlled ecosystems that provide an ideal setting for studying microbial communities. They bridge the gap between lab-controlled environments and natural ecosystems. In RAS, we can trace the colonisation of different compartments over time and try to link observed disease outbreaks to the presence of pathogenic microbes.

In a study of the RAS microbiome across space and time, we conducted ultra-deep sequencing of 496 samples from various compartments across six RAS farms at different time points. To analyze this complex dataset, we developed a statistical pipeline leveraging state-of-the-art machine learning tools. Our pipeline focuses on distinguishing between adjacent microbial communities and the environmental factors shaping them, starting with the microbial relative abundance data commonly obtained from metagenomic studies. Our findings validate previous research indicating the significance of pH as a driving force behind microbial variability. Furthermore, our pipeline’s versatility was demonstrated through successful testing on other compartmentalized microbial communities, such as from longitudinally sampled ruminant guts.
Phylo-pangenomics: read mapping and variant detection in huge microbial datasets guided by pandemic scale evolutionary history

Alexander Kramer

Alexander M Kramer, Nicolas M Ayala, Lily M Karim, Sumit Walia, Harsh Motwani, Alan Zhang, Yatish Turakhia, Russell Corbett-Detig
San Diego (USA), Santa Cruz (USA), University of California

Pangenomes, which encode all genetic variation in related genomes, are an increasingly powerful tool in evolutionary genomics. Current pangenome methods catalog variation among observed haplotypes, but neglect the evolutionary processes that generated this variation. In the era of pandemic-scale genomics, huge datasets reflecting microbial evolution are common and not well-served by existing pangenome methods. We present an alternative that, in addition to known variation, also encodes inferred molecular evolutionary history of samples in a phylogenetic tree. Tree pangenomes are compact and accommodate millions of complete microbial genomes, scaling to many more individuals than existing haplotype graph-based approaches. We also developed panmap, a variant caller and genome assembler guided by reference tree pangenomes. From short or long reads, panmap finds a close relative in a tree panggenome with k-mer sketching, aligns reads to that sequence or inferred ancestor, computes genotype likelihoods using a mutation spectrum prior, and optionally generates a consensus assembly of the input sample. We show that entire global phylogenomic datasets for diverse pathogens like Mycobacterium tuberculosis and SARS-CoV-2 can be compressed into reference tree pangenomes and used with panmap to improve variant calling and assembly pipelines for these species. Incorporating evolutionary relationships into pangenomic analyses can significantly enhance the accuracy and efficiency of identifying genetic variants and assembling genomes. This approach holds promise for advancing our understanding of microbial evolution and improving high-throughput analyses across domains like clinical genomics and public health.
Acetic acid bacteria: an evolutionary story of the quest for sugar

Ana Cuesta-Maté

Globe Institute, University of Copenhagen (Denmark)

Acetic acid bacteria are generally considered beneficial microbes for humans and used to produce several fermented foods and beverages. They dominate the microbial communities in acetic fermentation, like vinegar or kombucha. Humans have historically domesticated these bacteria to obtain some of the ferments that we consume, but the question remains on how these microbes got there in the first place. Fermented food researchers find it difficult to sample these bacteria in natural habitats, but our work shows that some of these bacteria are found in social insect guts and environment, in the food storages of their nests, and in the flowers and fruits social insects visit. These bacteria are adapted to somewhat harsh environments (low pH, high osmolarity) where simple sugars are readily available, and so, colonized by many different types of microbes. How has life within insect microbiomes made these bacteria ideal to colonize host-controlled environments, like our nutritious ferments? This project focuses the adaptions of acetic acid bacteria to these dynamic communities, using their presence in fermented food and insect microbiomes to follow their evolution within microbial communities. Genomic adaptions have been observed in acetic acid bacteria domesticated by insects, namely, genome reduction and base compositional bias towards an AT-rich genome. Similar patterns can be observed in acetic acid bacteria in our ferments, but are there further adaptions to life within insect microbiomes that gave acetic acid bacteria an advantage to become our fermented food microbes? Could insect metabolic pathways be already trained for food fermentation?
Adaptive Mechanisms of K1 Killer Toxin Resistance from a Sensitive-Killer Coevolution in Budding Yeast (S. cerevisiae)

Angelina G. Chan

Angelina G. Chan, Michelle Hays, Angela Hickey, Magdalena Pieczynska, Arjan de Visser, Gavin Sherlock
Department of Genetics, Institute of Genetics and Biotechnology, Laboratory of Genetics, Stanford University (USA), University of Warsaw (Poland), Wageningen University (Netherlands)

Killer yeasts, such as the K1 killer strain of S. cerevisiae, express a secreted anti-competitor toxin whose production and propagation requires the presence of two vertically-transmitted dsRNA viruses. Though killer yeasts are immune to their own toxin, infection results in a fitness cost in the absence of toxin-sensitive competitors. Previous experimental coevolution of killer yeast containing the K1 killer virus and sensitive (non-killer) yeast resulted in coadaptation, including emergence of toxin resistance in sensitive ancestor-derived populations. We identified a putative dominant gain-of-function missense mutation in the SSK1 gene, which is sufficient for toxin resistance and fit under laboratory conditions yet is present in a domain strictly conserved in natural isolates. Preliminary data suggest this mutation may impact interactions of Ssk1 with both the HOG (high osmolarity glycerol) and CWI (cell wall integrity) pathways. Ongoing analysis of two separate asymmetrically evolved populations, where the killer population was replaced with an ancestral stock every ~25 generations, suggests both the presence of alternate adaptive pathways independent of SSK1 and increased variability in the K1 resistance phenotype arising in the presence of a non-evolving competitor. Our future work aims to both further understand the mechanism(s) by which Ssk1 mediates toxin resistance, including in fungi associated with pathogenicity such as C. auris (which requires SSK1 for multidrug resistance), and the mechanisms of coadaptation or genetic conflict arising between viral and host genomes in counterevolving killer yeast populations.
The Salmon Microbial Genome Atlas enables novel insights into bacteria-host interactions via functional mapping.

Arturo Vera Ponce de Leon

Arturo Vera-Ponce de Leon, Matthias Hoetzinger, Tim Hensen, Shashank Gupta, Bronson Weston, Sander M Johnsen, Jacob A Rasmussen, Knut Rudi, Lars Snipen, Christian René Karlsen, Morten T Limborg, Stefan Bertilsson, Ines Thiele, Torgeir R Hvidsten, Simen R Sandve, Phillip B Pope, Sabina Leanti La Rosa

NA (Norway), Nofima (Norway), Norwegian University of Life Sciences (Norway), Swedish University of Agricultural Sciences (Sweden), University of Copenhagen (Denmark), University of Galway (Ireland)

Our knowledge about the salmon (Salmo salar) gut microbiota is limited, despite the importance of salmonid aquaculture as one of the most expanding food-production sectors worldwide. This hampers the study of functional interactions between host and bacteria, and the development of dietary strategies aimed at the manipulation of the gut microbiota to improve salmon welfare. Furthermore, the lack of salmon gut-derived microbial genome data makes it difficult to interpret functional omics dataset, as much of the data do not match the available reference bacterial genomes. Combining published and unpublished data, here we present the first version of the Salmon Microbial Genome Atlas (SMGA), originating from fish reared both in fresh water and saltwater. The SMGA comprises 211 high-quality bacterial genomes, recovered by cultivation (n=131) and gut metagenomics (n=80). These genomes were taxonomically assigned into 14 different orders, including 35 distinctive genera and 29 potentially novel species. Benchmarking the SMGA, as a database for metatranscriptomics and metabolomics experiments, functionally characterized key populations in the salmon gut that were detected in vivo. This included the ability to degrade diet-derived fibers and release vitamins and other exo-metabolites with known beneficial effects, which were validated by in vitro cultivation and untargeted metabolomics. Together, the SMGA enables high resolution functional insight into salmon gut microbiota with relevance for salmon nutrition and health.
Disentangling the evolutionary impacts of relatedness and facultative/obligate life cycles during the transition to multicellularity

Autumn Peterson, Anthony Burnetti, William C Ratcliff

Georgia Insitute of Technology (USA), Georgia Institute of Technology (USA)

Complex multicellularity evolved in five lineages, and in each case, these organisms develop clonally and are obligately multicellular. While prior work has shown that clonal development plays a critical role in the evolution of complex multicellularity, none has disentangled this from the impact of obligate vs facultative multicellular life cycles. Theory predicts that selection acting on single cells may impede group-level adaptation, but due to a lack of model systems, this has never been tested. We circumvent this constraint by engineering a strain of Saccharomyces cerevisiae to grow as single cells or clonally-multicellular groups ‘snowflakes’ upon induction with ?-estradiol. We evolved 20 replicate populations of one ancestral genotype with four different life cycles: always unicellular, always multicellular, or alternating between uni and multicellularity (either spending 50% or 75% of their time in a multicellular state). Only the obligately multicellular treatment evolved ‘entrenching’ traits: those which increase group-level fitness at a cost to cell-level fitness. For example, obligately multicellular snowflake yeast convergently evolved to be tetraploid, increasing group size and fitness but reducing cellular growth rates. We show that while tetraploidy should be adaptive in the facultatively multicellular treatments, it never evolved, because selection in the unicellular stage is far more effective than the multicellular stage (~100-fold higher Ne). Our results demonstrate that an obligately multicellular life cycle plays a key role in the transition to multicellularity independent from its effect on genetic relatedness, by avoiding disproportionately powerful cell-level counterselection against novel multicellular traits.
Mixed Wolbachia infections resolve rapidly during in vitro evolution

Cade Mirchandani

Cade Mirchandani, Pingting Wang, Jodie Jacobs, Maximilian Genetti, Evan Pepper-Tunick, William T Sullivan, Russ Corbett-Detig, Shelbi Russell

Cell, Department of Biomolecular Engineering, Department of Molecular, Molecular Engineering Sciences Institute, University of California Santa Cruz (USA), University of Washington (USA), and Developmental Biology

Wolbachia pipientis is one of the most common endosymbionts found in arthropods, and it reached high prevalence in many taxa through its abilities to infect new hosts and their germlines. However, little is known about the ecological and evolutionary interactions that occur when a promiscuous strain colonizes an infected host. Here, we study what occurs when two strains come into contact in host cells following horizontal transmission and infection. We focus on the faithful wMel strain from Drosophila melanogaster and the promiscuous wRi strain from Drosophila simulans using an in vitro cell culture system. Mixing D. melanogaster cell lines stably infected with wMel and wRi revealed that wMel outcompetes wRi quickly and reproducibly. Furthermore, wMel was able to competitively exclude wRi, and we found that this was a nearly deterministic outcome, independent of the starting infection frequency. This competitive advantage was not exclusive to wMel's native cell background, as wMel also outgrew wRi in D. simulans cells. Overall, wRi is less adept at in vitro growth and survival than wMel and its in vivo state, revealing differences between cellular and humoral regulation. These attributes may underlie the observed low rate of mixed infections in nature and the relatively rare rate of host-switching in most strains. Our in vitro experimental framework for estimating cellular growth dynamics of Wolbachia strains in provides the first strategy for parameterizing endosymbiont and host cell biology at high resolution. This toolset will be crucial to our application of these bacteria as biological control agents.
The nature of the Last Universal Common Ancestor and its impact on the early Earth system

Edmund R R Moody


Centre for Ecological Research (Hungary), Eötvös University (Hungary), NA (Hungary), NA (Netherlands), NA (United Kingdom), Okinawa Institute of Science and Technology Graduate University (Japan), Royal Netherlands Institute for Sea Research (Netherlands), Royal Netherlands Institute for Sea Research (United Kingdom), University College London (United Kingdom), University of Amsterdam (Netherlands), University of Bath (United Kingdom), University of Bristol (United Kingdom), University of Exeter (United Kingdom)

The nature of the Last Universal Common Ancestor (LUCA), its age, and impact on the Earth system have been the subject of vigorous debates across diverse disciplines, often based on disparate data. Age estimates for LUCA are usually based on the fossil record, varying with every reinterpretation. The nature of LUCA’s metabolism has proven equally contentious, with some attributing all core metabolisms to LUCA, while others reconstruct a simpler life form dependent on geochemistry. Here we show, using a set of pre-LUCA duplications under a new cross-bracing approach implemented in PAML, calibrated with microbial fossils and isotope records of metabolisms, that LUCA existed ~4.2 Ga. We use 700 prokaryotic genomes to infer a 2.59Mb+ genome encoding at least 2500 proteins, comparable to modern prokaryotes. Our results suggest LUCA was anaerobic, acetogenic and possessed an early immune system. While LUCA is sometimes perceived as living in isolation, we infer LUCA to have been part of an established ecology. The metabolism of LUCA would have provided a niche for other community members, while hydrogen recycling by atmospheric photochemistry could have supported a modestly productive early ecosystem.
Building a reference genomic database of oral microbes for high-resolution oral microbiome analyses

Gabriel Yaxal Ponce Soto, Nicolas Rascovan
CNRS UMR 2000, Institut Pasteur, Microbial Paleogenomics Unit (France), Université de Paris Cité

The oral microbiome situates at the interface between the external environment and the host, exerting significant influence on both local and systemic health. Despite advancements in metagenomic studies, which have provided insights into the taxonomic composition of the oral microbiome, there remains a substantial gap in understanding its variation across various levels (including pangenomic, strain diversity, and functional) and scales (ranging from individual to population, temporal, and geographical). While k-mer-based approaches have been widely employed to classify sequencing reads taxonomically, their utility for more in-depth analyses at the above mentioned levels and scales is limited. To move forward into this direction, we constructed a comprehensive and non-redundant database of oral microbial genomes to be used with mapping-based strategies. Our database and approach allow to gather whole genome data of most species in the oral microbiome, which can be used to i) have better accuracy in detecting taxa with strain-level resolution, ii) perform phylogenomic and phylogeographic analyses, and iii) explore pangenome composition of each species and its genetic functional variation. A main advantage of this strategy is that it also allows analyzing ancient oral microbiomes (e.g., ancient DNA from archaeological dental calculus), which are often too degraded to perform de novo assemblies, thus providing a temporal depth to investigate changes over time. Through these efforts, we aim to deepen our understanding of the oral microbiome’s dynamics and composition, to investigate the evolutionary trajectories of its associated species, and its implications for human health and disease.
Expanded Geographic Distribution for Two Legionella pneumophila Sequence Types of Clinical Concern

JENNAFER HAMLIN

Jennafer A. P. Hamlin, Natalia A. Kozak -Muiznieks, Jeffrey W. Mercante, Lavanya Rishishwa, Emily T. Norris, Anna B. Gaines, Maliha K. Ishaq, Jonas M. Winchell, Melisa J. Willby (USA), Applied Bioinformatics Laboratory (USA), Centers for Disease Control and Prevention, Centers for Disease Control and Prevention (USA), Division of Bacterial Diseases

Legionella pneumophila serogroup 1 sequence types (ST) 213 and 222, closely related variants, were first detected in the early 1990s in the Midwest US and subsequently in the Northeast US and Canada. These STs are associated with both healthcare- and community-acquired Legionnaires’ Disease (LD). We aimed to assess the trend in cases linked to these STs and identify any genetic distinctions. Analyzing data from the Centers for Disease Control and Prevention (CDC) spanning 1992 to 2020, we found a yearly increase of 0.15 cases of ST213/222-associated LD across the Mountain West to Northeast region. Most cases occurred in Michigan (26%), New York (18%), Minnesota (16%), and Ohio (10%). ST222 caused at least five LD outbreaks in the US between 2002 and 2021, while no outbreaks were attributed to ST213. Genetic analysis of 230 isolates revealed high nucleotide identity within each ST (99.92% for ST222, 99.9299.92% for ST213), indicating low diversity within and between these types with a minimum of 99.5% and a maximum of 99.87%. Despite identifying genomic features like plasmids and CRISPR-Cas systems, no specific association explained the increasing prevalence and spread of ST213 and ST222. However, our study highlights the expanded geographical distribution of these STs in the US.
Microalgae release organic compounds into their surroundings that support specific bacterial communities known as the phycosphere microbiota. In exchange, these communities provide benefits to their host by supplying nutrients or enhancing resistance to environmental stressors. The recent sequencing and culturing of phycosphere bacteria from the microalga Chlamydomonas reinhardii has enabled the design of synthetic microbial communities for long-term gnotobiotic experiments. Despite these advances, we lack a mechanistic understanding of how these and other complex communities evolve and interact over extended periods. To address this, we developed a reductionist experimental system to explore long-term ecological interactions, metabolic exchanges, and strain-specific adaptation processes in synthetic phototrophic communities under varying environmental conditions over prolonged periods of time. Here, we report the results of a long-term artificial evolution experiment, where multiple independent synthetic communities were grown for over 500 days using photobioreactors. By combining amplicon sequencing and metagenomics with high-resolution phenotypic data, we have generated one of the most extensive longitudinal host-microbiome datasets to date, demonstrating that long-term artificial selection experiments using complex synthetic communities are feasible. Our analysis reveals distinct ecological networks of co-occurrence and co-exclusion, likely driven by metabolic interactions among community members. Furthermore, based on sequencing results and analysis of re-isolated bacteria, we identified mutations associated with nutrient uptake and antimicrobial compound production, among others, demonstrating strain-specific adaptation processes across different taxa. Our findings provide insights into the ecological factors that contribute to the evolutionary landscape of host-associated microbial communities.
In polymicrobial communities, bacteria fight over nutrients and space. One common weapon is the Type VI Secretion System (T6SS), a nano-harpoon which injects toxins into neighboring cells. While much is understood about mechanisms of T6SS-mediated toxicity, less is known about mechanisms of defense against this attack. We previously subjected Escherichia coli populations to 30 rounds of attack (500 generations) by Vibrio cholerae's T6SS, with a period of exponential growth in between attacks. During these 500 generations, E. coli evolved evolve moderate resistance to T6 attack, through two independent genetic pathways, but this came with a substantial cost to their growth rate. Here, we doubled the duration of the experiment. Surprisingly, at least one E. coli lineage adapted even more rapidly during generations 500-1000, with resistance increasing 4.7 times as much as in the first phase, reaching an ultimate survival of 2400-fold higher than the ancestor. Moreover, our highly resistant E. coli does not grow slower than earlier isolates, breaking the trade-off between greater T6SS-survival and reduced growth rates. These results suggest that trade-offs make the early evolution of T6SS resistance difficult for bacteria that do not often encounter T6SS attack, while more sustained periods of interaction lead to an easier gain in survival. Our experiment highlights an interesting non-additivity of bacterial adaptation to T6SS attack. T6SS, relative to many other strongly-selective antimicrobial agents, is surprisingly robust, requiring a greater degree of sustained selection prior to the evolution of low-cost resistance.
How does nutrient environment and varying mutational constraint impact regulatory rewiring in bacteria?
Louise M Flanagan

Gene regulatory networks are fundamentally important for organism development and survival, ensuring that the right complement of genes is expressed in the right place at the right time. Despite their critical role, they are also subject to change, and many factors drive their evolution. Here I consider how environmental differences can pull evolution down different mutational paths even when selecting for the same phenotype, and the effects of varying mutational constraint and accessibility. Using an established model system in the bacterium Pseudomonas fluorescens, where I re-evolve a lost swimming phenotype (due to flagellar regulator deletion) through starvation selection in different nutrient environments, I show how mutational spectra vary in complex and minimal nutrient environments. By introducing a strong mutational hotspot that restores swimming, I compare the effects of strong and reduced mutational constraint to regain the swimming phenotype and the consequences of this. Having a mutational hotspot can lead to faster restoration of a swimming phenotype, but it can prevent access to other viable mutations that produce faster swimmers. I also explore how the swimming phenotype varies between mutants and investigate the apparent loss of swimming when placed in an alternative nutrient environment, and how this links to regulatory architecture and gene expression. This work highlights the importance of considering the environment and mutational accessibility in evolutionary studies. Incorporating these variables into predictive evolution models is crucial for gaining a comprehensive understanding of the evolutionary dynamics of regulatory networks.
Most statistical methods for microbial data tend to analyze just one or two microbiome samples at a time, limiting their capacity to measure complex dynamics across multiple samples. For example, such methods are ill-suited to measuring the stability of microbial communities over long timescales, the heterogeneity of microbiomes across multiple host individuals, and the repeatability of microbial community assembly. To assess these aspects of microbial community dynamics, we need a measure of compositional variability across two or more microbiome samples at once. We introduce FAVA, an FST-based Assessment of Variability across vectors of relative Abundances. FAVA extends the diversity-partitioning framework of the population-genetic statistic FST, with microbiome samples playing the role of “populations” and microbial taxa playing the role of “alleles.” We demonstrate FAVA in a longitudinal analysis of the gut microbiomes of 22 healthy adults taking an antibiotic. We find that the antibiotic destabilizes microbial communities, using FAVA to quantify both the magnitude and the timescale of the antibiotic perturbation. In particular, we show that FAVA can identify a return to stability regardless of whether the community returns to its initial compositional state. We have implemented FAVA in an R package useful for quantifying the evolutionary and ecological dynamics of microbial communities over time, space, or host individuals.
Diazotrophic endophytic community in Cordia dodecandra phyllosphere according to management in agroforestry systems

Miguel Angel López-Garrido

Miguel A López-Garrido, Miriam M Ferrrer, Yuri J Peña-Ramírez, Natalia Y Labrín-Sotomayor, José R Pacheco-Arjona, Hector Estrada-Medina
CONAHCYT-Universidad Autónoma de Yucatán (Mexico), ECOSUR (Mexico), Universidad Autónoma de Yucatan (Mexico), Universidad Autónoma de Yucatán (Mexico)

The phyllosphere represents the most extensive terrestrial habitat, approximately twice the land-earth area. Bacteria dominate the phyllosphere, playing an essential role in nitrogen fixation and host's growth. Since management affects the diversity of the microbiota, assessing the composition of diazotrophic endophytic communities in trees within systems that differ in management practices can help us understand their role in nitrogen fixation. In Cordia dodecandra, the core microbiota comprises species with diazotrophic potential. However, the diversity of actors and their variation with management have not been assessed. For this purpose, microbial DNA was extracted from the leaves of two trees in a forestal system (reduced irrigation and weeding) and four in a silvopastoral system (goat grazing, star grass cover, besides irrigation and weeding). Additionally, the canopy cover was also measured. The DNA was amplified for the nifH gene by Qpcr, mass sequencing of the nifH genes, and V3-V4 region of the 16S rRNA. Alpha and beta diversity was analyzed for sequences. Similar Shannon and Chao1 index values were found in both systems. However, the communities composition differed between the systems. The relative abundance of nifH genes was positively correlated with canopy cover. Aureimonas, Quadrisphaera, Actinomycetospora, and Pseudokineococcus were enriched in the forestal system. Diazotrophic species identified (p > 99%) with nifH genes were Alcaligenes faecalis, Ensifer americanum, Methanosaeta concilii, Chroococcidiopsis thermalis, Hassallia byssoidea and Nostoc commune. A small proportion of the C. dodecandra phyllosphere microbiota is nitrogen-fixing, although the role of the majority of species systems remain unknown.
Horizontal Gene Transfer (HGT) is a defining characteristic of microbial evolution in natural microbial communities. While HGTs importance in microbial evolution has been clearly demonstrated by comparative genomics, there are still only a handful of studies that have incorporated HGT into evolution experiments. We have carried out a series of experiments capturing the dynamics and fitness effects of horizontally transferred genetic variants in experimental populations in the naturally competent bacteria H. pylori and A. baylyi. I would like to present evolution experiments, genome sequence and large-scale sequencing fitness assay data that shows how recombination can be beneficial. In particular, I will discuss how Fisher-Muller dynamics act in microbial populations, how HGT can reverse the evolution of antibiotic resistance, and present exciting new work showing how we can carry out experimental tests of pangenome evolution.
This study explored the evolutionary dynamics of multi-drug strains of Klebsiella spp. and Escherichia coli isolated from Aguascalientes hospitals between 2020 and 2023. We identified 65 multidrug-resistant strains. These strains were subjected to biotyping and sequencing to elucidate their genomic landscapes and antibiotic resistance profiles. Comprehensive antibiotic susceptibility testing was conducted using broth microdilution, including the detection of β-lactamases and carbapenemases. Our study revealed a significant increase in the diversity and prevalence of resistance genes among multidrug strains of Klebsiella spp. and Escherichia coli over the years, particularly affecting third- and fourth-generation antibiotics. Genotype-phenotype correlations identified discordances, indicating potential adaptation mechanisms in response to antibiotic exposure, whereas phylogenetic analyses demonstrated temporal divergence and evolutionary relationships within the strain populations. Comparisons with pre-pandemic samples emphasized the accelerated pace of genetic evolution under antimicrobial pressure. Additionally, biosynthetic gene cluster (BGCs) analysis provided insights into secondary metabolite production. Our study provides a comprehensive investigation of the evolution of antibiotic resistance in a hospital environment, demonstrating how genetic makeup alters and resistance genes evolve. These findings underscore the importance of monitoring genetic alterations in bacteria to inform medical treatment strategies and to devise effective interventions. Our study highlights the pivotal role of genomic surveillance in shaping clinical strategies and in promoting the development of targeted interventions. By understanding the evolution of bacterial resistance, health care professionals can make informed decisions and take proactive measures to combat the escalating threat of antibiotic resistance.
Communities, lineages, and the evolution of evolvability
Paul B. Rainey

Presented by self
Max Planck Institute for Evolutionary Biology (Germany)

The term, evolvability, first coined almost 100 hundred years ago, has entered common usage to the extent that today many accept the idea that life has evolved to evolve as self-evident. However to many, the notion is highly contentious. Skeptics point to the fact that evolvable systems require foresight and that this is incompatible with natural selection as a blind process. They also point to the fact that evolvability is a property of lineages and that selection beyond the level of individuals is extremely weak. I will describe the results of a substantive selection experiment that documents, in real time, the de novo emergence of evolvability. Key to experimental realisation was a lineage-level birth-death dynamic, where lineage reproductive success depended upon capacity to mutate between two target phenotypic states, each optima in a repeating cycle of environments. Fuelled by variation in evolutionary potential associated with unique mutational paths, lineages capable of mediating rapid and reliable transitions between states through local mutational bias emerged. The mechanism is analogous to that underpinning highly mutable "contingency loci" in pathogenic bacteria. Drawing upon detailed knowledge of the evolutionary history of lineages, I will describe key steps in construction of the locus, including changes to the genotype-phenotype map, and input from mutations that elevated transcription and concomitantly, rates of localised frame-shifting. I will also report ancillary advantages to rapid phenotype switching. Our results provide a detailed mechanistic account of the adaptive evolution of evolvability.
Eco-evolutionary feedbacks promoted by hypermutation rate in a predator-prey microcosm

Pu Wang

Pu Wang, Michael Travisano
University of Minnesota (USA)

Ecological and evolutionary dynamics can operate at the same time. The eco-evolutionary feedback could intensify the adaptation and lead to further diversification of both communities and environments. Eco-evolutionary dynamics are dependent on genetic variation. Mutation, the source of genetic variation, could play a role in influencing eco-evolutionary dynamics. Our project used a predator-prey microcosm with the ciliate Tetrahymena thermophila consuming Pseudomonas fluorescens bacteria. Five Pseudomonas strains with distinctive pre-existing genetic variation including mutator and maladapted genetic constraints were subjected to predatory selection. We observed the intensification of adaptation in prey populations due to fitness landscape changes via eco-evolutionary feedback. The evolutionary responses of the five strains were compared against each other to determine the roles of genetic constraints and mutation rates in eco-evolutionary interactions. The pre- and post-predation population genome sequencing data revealed that evolved populations with lower genetic diversities had higher phenotypic diversities. These results reflect the challenges of genotype-phenotype mapping. We genome-sequenced certain colonies with new phenotypes to investigate the underlying evolutionary mechanisms. We found that potential adaptations could be limited by genetic constraints, which could be overcome by higher mutation rates. The results suggested mutation rates had a role in determining the evolutionary trajectories, thus mediating the eco-evolutionary feedback.
How ecological interactions shape microbial mutation rates: a story of collective detoxification

Rowan Green

Rowan C Green, Rok Krašovec, Christopher G Knight
University Of Manchester (United Kingdom), University of Manchester (United Kingdom)

The ultimate source of genetic variation, de novo mutagenesis, is a highly dynamic trait responsive to multiple ecological factors. Evolution therefore depends on the ecological environment not only for selection but also in determining the variation available in a population. One such environmental dependency is the inverse relationship between mutation rates and population density in many microbial species. Using dynamical computational modelling and in vitro mutation rate estimation we show that the negative relationship between mutation rate and population density arises from the collective ability of microbial populations to control concentrations of hydrogen peroxide (H2O2). Our in silico modelling approach to hypothesis generation allows us to efficiently explore the idea-space of possible mechanisms for this mutation rate plasticity through altering model structures and parameters. This generates select hypotheses, which we test using high-throughput mutation rate measurements alongside direct measurements of environmental H2O2. We demonstrate a loss of this density-associated mutation rate plasticity in Escherichia coli populations deficient in the ability to degrade H2O2 however, the reduction in mutation rate in denser populations can be restored in peroxide degradation-deficient cells by the presence of wild-type cells in a mixed population. These model-guided experiments provide a mechanistic explanation for density-associated mutation rate plasticity, applicable across all domains of life, and frame mutation rate as a dynamic trait shaped by microbial community composition.
The influence of subsistence strategies on oral microbiome composition and function in ancient populations

Sarah J Johnson

Sarah J Johnson, Tanvi P Honap, Cara Monroe, Alan Leventhal, Monica V Arellano, Douglas J Kennett, Keith M Prufer, Cecil M Lewis
Center for Stable Isotopes (USA), Center for the Ethics of Indigenous Genomic Research (USA), Laboratories of Molecular Anthropology and Microbiome Research (USA), Muwekma Ohlone Tribe (USA), San Jose State University (USA), University of California (USA), University of New Mexico (USA), University of Oklahoma (USA)

Through the analysis of shotgun metagenomes recovered from archaeological dental calculus, we present an investigation into the effects of diet on oral microbiome composition and function in human populations from the pre-European Contact Americas. These include ancestral Ohlone individuals from California, USA, and ancestral Maya individuals from rock shelters in Belize occupied for over 10,000 years. MetaPhlAn4 and HUMAnN3 were used to estimate taxonomic composition and functional potential, respectively, and SourceTracker analysis and DNA damage patterns were used to authenticate recovery of ancient oral metagenomes. In the ancestral Maya, the transition from a pre-agricultural to maize-based diet is expected to mark a change in carbohydrate metabolism. In this study, we demonstrate that this shift in subsistence pattern was accompanied by a significant reduction in abundance of the oral bacterium, Pseudoramibacter alactolyticus, estimated to be the primary contributor to carbohydrate metabolism. Following maize adoption, several Streptococcus species, including known opportunistic pathogens implicated in infective endocarditis, became dominant contributors to carbohydrate metabolism. The ancestral Ohlone, primarily hunter-gatherers with a starch-rich diet of acorns and seeds, did not have a dominance of Streptococcus species contributing to carbohydrate metabolism, but did show a significantly higher abundance of microbial gene families involved in starch metabolism, such as alpha-amylase and starch phosphorylase, as compared to pre-staple maize ancestral Maya and other ancient agricultural populations. These data reveal a complex metabolic response to dietary changes highlighting the interplay between diet, microbial ecology, and health over the course of human evolution.
Harmful blooms caused by diazotrophic (nitrogen-fixing) Cyanobacteria are becoming increasingly frequent and negatively impact aquatic environments worldwide. Many of these diazotrophic strains are multicellular, and form filaments. Due to the sensitivity of the nitrogenase (the enzyme responsible for nitrogen fixation) to oxygen, nitrogen fixation cannot be performed at the same place and the same time with photosynthesis. Multicellularity enabled the evolution of special cells called heterocysts, in which nitrogen fixation is being performed, while photosynthesis is performed in the other vegetative cells. Cyanophages (viruses infecting Cyanobacteria) can potentially regulate cyanobacterial blooms, yet Cyanobacteria can rapidly acquire mutations that provide protection against phage infection. Here I will present results that provide novel insights into cyanophage:Cyanobacteria interactions by characterizing the resistance to phages in two species of heterocystous diazotrophic Cyanobacteria: Nostoc sp. and Cylindrospermopsis raciborskii. These results demonstrate that phage resistance is associated with a fitness tradeoff by which resistant Cyanobacteria have reduced ability to fix nitrogen and/or to survive nitrogen starvation. Furthermore, whole-genome sequence analysis of 58 Nostoc-resistant strains allows the identification of several mutations associated with phage resistance, including in cell surface-related genes and regulatory genes involved in the development and function of heterocysts (cells specialized in nitrogen fixation). Furthermore, phylogenetic analyses show that most of these resistance genes are accessory genes whose evolution is impacted by lateral gene transfer events. Together, these results further our understanding of the interplay between multicellular diazotrophic Cyanobacteria and their phages and suggest that a tradeoff between phage resistance and nitrogen fixation affects the evolution of cell surface-related genes and of genes involved in heterocyst differentiation and nitrogen fixation.
Expanding the Plant Microbiome
Sheila Roitman

Sheila Roitman, Haim Ashkenazy, Detlef Weigel
Max Planck Institute for Biology (Germany)

Eukaryotic organisms harbour large communities of microorganisms forming an holobiont, considered to be a single ecological and evolutionary unit. In recent years, bacterial community dynamics and their effect on the plant holobiont have been the subject of many studies. However, little is known regarding the role that bacteriophages play in shaping those bacterial communities. In my work, I intend to set the basis for understanding the role of the microvirome in plant colonization and development, by studying Arabidopsis thaliana associated bacteria and phages, in laboratory and natural settings. I isolated lytic viruses of Pseudomonas viridiflava, an environmentally relevant plant-associated microbe. These bacteriophages are generalists, infecting more than one strain of P. viridiflava, including pathogenic and commensal bacteria. Lysogenic phages are ubiquitously found in the genomes of these microbes, and are likely playing a role in defense against lytic phages and other microbes. These findings will serve as the basis for a better understanding of the impact phages have on the bacterial community assembly and evolution in the plant.
The lipopolysaccharide specificity of mammalian TLR4/MD-2 complexes has fluctuated randomly over evolutionary time

Sophia Phillips

Sophia Phillips, Maggie Barry, Laurel Moneysmith
University of Oregon (USA)

The animal innate immune system continuously evolves to recognize and respond to microbes. Toll like receptor-4 (TLR4) and co-receptor MD-2 form a complex, found on the surface of immune cells, that mediates inflammation in response to lipopolysaccharides (LPS) from the outer membrane of gram-negative bacteria. Different bacteria produce LPS that vary in number of acyl chains, which allow them to activate or evade TLR4/MD-2. We hypothesized that host species maintain specificity for their unique bacterial landscapes via coevolution of TLR4/MD-2 that recognize these LPS molecules. To test this hypothesis, we characterized LPS specificity of TLR4/MD-2 complexes across mammals. We found that, while the ability to recognize LPS variants with six or seven acyl chains is strongly conserved, recognition of LPS with four or five acyl chains was repeatedly gained and lost during evolution. A simple explanation for this would be that natural selection altered specificity as microbial communities changed over time. Consistent with this hypothesis, the co-receptor MD-2 (where LPS binds) has a higher ratio of sites with elevated dN/dS compared to TLR4 (6.3% vs 2.7%, respectively). However, mutagenesis of sites in the MD-2 binding pocket experiencing diversifying selection were not sufficient to switch ligand specificity. Instead, the genetic background was a stronger predictor of effect than identity of mutation. This suggests that the primary evolutionary constraint on mammalian TLR4/MD-2 is to recognize larger LPS molecules, while the ability to recognize small LPS molecules fluctuates stochastically over evolutionary time.
S20 - Pushing the frontiers of conservation genomics.
Prey in Peril: Impact of Habitat Alterations on two Large Herbivore species in Central Indian Landscape
Abhinav Tyagi

Abhinav Tyagi, Uma Ramakrishnan
National Centre for Biological Sciences (India)

In current times, habitat fragmentation and loss are the major threats to biodiversity worldwide. Reduction of habitat to smaller patches and increased distance between patches, lead to small isolated populations that pose an increased risk of loss of genetic diversity, inbreeding and genetic load. Maintaining gene-flow among these fragmented habitat patches is critical for long-term species persistence. Several natural and anthropogenic landscape features impede animal movement. Identifying these features is crucial for effective conservation planning. Genetic approaches aid in quantifying connectivity but most studies have focussed on a single species, usually, a large carnivore. Despite playing important role in ecosystem functioning and habitat maintenance, herbivores have been largely neglected. Here, we address genetic connectivity of two large herbivores, Gaur (Bos gaurus) and Sambar (Rusa unicolor) in central India. We collected fecal samples to generate genome-wide SNP data using ddRAD sequencing for 124 Gaur and 99 Sambar individuals. We demonstrate that gaur population in central India is fragmented and exhibit high genetic differentiation, especially in small populations like Umred Karhandala WLS. Although Sambar shows low genetic structure, small population in Bor Tiger Reserve exhibits slight differentiation. Our results suggest that although forest degradation and roads restrict animal movement, the extent of the impact varies with species ecology. Our findings reveal that different species exhibit varied responses to various landscape features. We identify small and isolated populations that need conservation intervention. We opine a shift from large and charismatic species-focused conservation to a multi-species landscape conservation approach.
Genetic variation, adaptation, and conservation of guenons across multiple natural environments

Adeola Oluwakemi Ayoola

Adeola Oluwakemi Ayoola, Amy Goldberg
Duke University (USA)

Guenons (Cercopithecus), are the most widely dispersed non-human primates across Africa's tropical forests, exhibiting diverse taxonomic, phenotypic, and ecological behaviors. Here, we focus on guenons of Nigeria at the edge of the diverse Upper Guinea-Congolian forest. Multiple of these species are on the IUCN Red List resulting from bushmeat hunting, habitat loss, and fragmentation. Across three field seasons, almost one-third of bushmeat we tested belonged to guenon species, including Mona (31), Sclater's (10), Red-eared (1), and White-throated guenon (4), emphasizing a role for hunting pressures in conservation. We next considered the parasite pressures shaping guenon genetic variation and conservation risk. As our close phylogenetic relatives, multiple monkeys act as a reservoir of and model for, malaria that can infect human populations. Using genomic data, we previously found a signature of adaptation and selection at the malaria-interacting gene G6PD, notably, only in a subset of populations. To understand the selective landscape and genomics of adaptation to malaria, we collected and sequenced ~400 fecal samples from multiple ecologies—tropical rainforest, mangrove, and Guinea savanna—using a newly designed genus-specific capture. We tested for population-specific and shared signatures of selection and identified novel and shared variants across guenons and humans. Further, we tested for multiple malaria species to reconstruct the geographic range and natural occurrence of parasites compared to local human populations informing the risk for zoonoses. Importantly, this work is synergistic with educational components and community engagement towards building long-term conservation and public health initiatives.
Heavily fished Malawi cichlids rapidly adapt to anthropogenic pressures.

Alexander Hooft van Huysduynen

Alexander Hooft van Huysduynen, Hannes Svardal
University of Antwerp (Belgium)

In some remarkable cases, natural populations are able to persist through rapid evolutionary adaptation or phenotypic plasticity despite significant anthropogenic pressures. While ecological consequences are prioritized in conservation, the genomic impact of human-induced environmental change is often under-prioritised.

The Malawi cichlid radiation is renowned for its species richness and ecological diversity, but it is also a critical source of dietary protein and livelihood for millions of people. Over the past 40 years, exploitative fishing of some Malawi cichlid populations has led to local extinctions, dramatically reduced population size, and has also induced shifts in life history traits, such as in the Lake Malombe, where Copadichromis mloto has undergone a rapid reduction in size at maturity of 50% compared with only 20 generations ago.

To comprehensively explore the genomic signatures of this anthropogenically induced evolutionary adaptation, we used whole-genome resequencing data from 223 Copadichromis mloto samples. Our analysis identified several selective sweeps in the genomes of heavily fished populations and differential allele frequency patterns between heavily and weakly fished populations consistent with rapid evolutionary adaptation. Candidate genes were found to be related to life history traits such as gonad development and whole organism growth.

Our results demonstrate the genomic consequences of fisheries-induced evolution in a non-model study system and present signals of very recent adaptation to localized anthropogenic pressures. As WGS datasets grow, we will become more capable of investigating the genomic scars of anthropogenic pressures and their consequences for the future resilience of natural populations.
The California Gulf (CG) is a closed sea with a high environmental variation, thus is a region candidate to present ecologically divergent natural selection in many marine species. Two species of butterfly rays (Gymnura marmorata and Gymnura crebripunctata), cohabit in the GC, both present morphology similitud, thus are usually confused. Accordingly, we use population genomics to assess the phylogeographic break between species. Add to this, the intraspecific genomic variation was evaluated. We use the technique ddRADseq to obtain SNPs from samples of both species obtained across the CG and the Pacific Ocean. To assess the phylogeographic break, we obtained 768 SNPs for both species, which showed a phylogeographic break in the south coast of Sonora. This pattern can reflect an adaptive divergence associated with bathymetry, chlorophyll, and temperature in the GC. The preliminary intra-species genomic structure suggests a genomic structure in G. marmorata between localities from the CG and the Pacific, due to the Baja California peninsula acting as a geographic barrier; while G. crebripunctata shows divergence between the mouth and center of the CG, this could be due to changes in the oceanographic conditions between zones and/or fidelity to reproductive sites. These results can be used as basic biological information for the design of fisheries management and conservation strategies in the GC.
Genetic management for an endangered carnivore: the spatiotemporal impacts of supplementation.

Andrea L Schraven

Andrea L Schraven, Catherine E Grueber, Elspeth A McLennan, Katherine A Farquharson, Andrew V Lee, Samantha Fox, Carolyn J Hogg
Department of Natural Resources and Environment (Australia), The University of Sydney (Australia), Toledo Zoo and Aquarium (USA)

Management of threatened species must be planned and executed, especially where prevailing threats, like disease, remain within the species’ range. Supplementations (or population reinforcements) involve releasing conspecifics into an existing population, to alleviate small population pressures, i.e. inbreeding and genetic drift. The Tasmanian devil (Sarcophilus harrisii) has suffered dramatic declines since the emergence of a contagious cancer, devil facial tumour disease, which is unmitigated in the wild. The Save the Tasmanian Devil Program (STDP) is responsible for managing devils in the wild and have been trialling releases of individuals into wild and diseased sites since 2015. The aim of our study was to quantify changes in genetic differentiation among devil sites before and after supplementation and assess the impact of this action on 1) genetic diversity and 2) relatedness within each site over time. We used genome-wide SNPs from nine years (2014 to 2022) of sampling across eight devil sites, four supplemented and four not supplemented. We found a statistically significant decline in genetic differentiation between supplemented sites compared to between not supplemented sites. On average, genetic differentiation decreased by 11.99% among not supplemented sites while genetic differentiation among supplemented sites decreased by 45.29% between 2014 and 2022. We compared changes in genetic diversity and relatedness over time between not supplemented and supplemented sites and found substantial variability among all sites. Our study sheds light on both the spatial and temporal genetic ramifications of reinforcing wild populations - a management strategy previously avoided where disease presence cannot be mitigated.
Counting invisible elephants. Using non-invasive DNA sequencing approaches to monitor wild elephants.

Andrew J Tighe

Earth’s remaining elephants are keystone species in their ecosystems, and are a crucial element in maintaining biodiversity and even restoring degraded landscapes. By selectively browsing certain plant species and spreading the seeds of others, elephants play an outsized role in shaping forest ecosystems. Effective conservation and management strategies for wild elephants require accurate information on population numbers and dynamics, inbreeding potential, disease risks, parasite loads and diet. Increasingly there is also a push to be able to obtain such information non-invasively to avoid unnecessary stress to the target animals. This study highlights how eDNA and scat sampling can provide such information, with a specific focus on a population of savannah elephants (Loxodonta africana) living in Arabuko Sokoke forest in Kenya. Arabuko Sokoke forest is the largest remaining fragment of East African dry coastal forest and contains an under studied population of elephants with an unknown population number. The very dense nature of the forest makes traditional counting methods impossible, and so this project will use a novel SNP based approach to generate a population size estimate for the elephants living in the forest, using DNA collected non-invasively from scat samples. By individually genotyping every elephant which visits the one remaining waterhole during the dry season, this project will then be able to extrapolate the number of elephants surviving in the forest. Further the project will utilize DNA metabarcoding methods to uncover both the diet of the elephants and what parasites they carry.
Genomic Consequences of Strong Bottlenecks in a Previously Endangered Bird
Anna Maria Calderon

Anna Maria Calderon, Andrew W Wood, Zachary A Szpiech, David Toews
- (USA), The Pennsylvania State University (USA)

Runs of homozygosity (ROH) are homozygous chromosomal segments that arise from population bottlenecks, isolation, and inbreeding, when two related individuals mate and pass on identical by descent (IBD) haplotypes to an offspring. In addition to providing insights into levels of inbreeding and genetic diseases, quantifying the distribution and prevalence of ROH can also illuminate population history such as bottlenecks and population expansions. In the early 1970's Kirtland's warblers (Setophaga kirtlandii) experienced population declines as a result of jack pine habitat degradation and nest parasitism by the brown-headed cowbird. The population crash caused it to be listed under the 1973 Endangered Species Act. Even though the population is above the recovery goal, the existence of this species relies heavily on conservation management. Using whole genome sequencing of contemporary and historical samples of this previously endangered bird, we compare the distribution of runs of homozygosity as well as allele frequencies pre-bottleneck and post-bottleneck recovery to understand the long-term genomic consequences of reduced population size.
Pedigree reconstruction informs conservation genomics: Inbreeding and mating patterns of a wild tiger population

BV Aditi Prasad

BV Aditi, Anubhab Khan, Kaushal Patel, Megan Aylward, Sampurna Roychoudhury, Uma Ramakrishnan
Indian Institute of Science Education and Research (India), NA (India), National Centre For Biological Sciences-TIFR (India), National Centre for Biological Sciences (India), National Centre for Biological Sciences-TIFR (India)

Inbreeding and its fitness consequences could pose a threat to endangered species that exist in small and isolated populations. Inbred populations show varying temporal trajectories, with some populations declining or going extinct, whereas other populations survive over time. From a conservation genetic perspective, it is crucial to monitor the genetics of such populations on a temporal scale. Tigers have undergone extensive decline, and individuals in Northwestern India show genomic signatures of inbreeding and isolation. We conducted individual-based whole genome resequencing and analyses from 32 invasive and 32 non-invasive shed hair samples to i) understand the matrilineal contributions of founding females, ii) track inbreeding and relatedness over five generations; and iii) establish paternity to estimate the relatedness between mate pairs. Whole mitochondrial genomes reveal closely related matrilineal founders, with three matrilines surviving in current individuals. We find a small (but insignificant) increase in inbreeding over these five generations. Paternity assignments suggest that no mate pairs are first-degree relatives. Based on the genomics of non-invasive (and invasive) samples, we synthesize an unprecedented wild pedigree over five tiger generations and highlight potential inbreeding avoidance in an endangered species.
The new global Convention of Biological Diversity will be voted on in December 2022, finally bringing together a new post-2020 framework. A primary goal of the new convention is that by 2050, the integrity and resilience of natural ecosystems is enhanced and protected by at least 15% of the current area; the rate of human-induced extinctions is halved; and genetic diversity and adaptive potential is safeguarded with at least 95% of genetic diversity among and within populations maintained. These targets have led to a global push in recent years to develop methods that will permit us to measure genetic diversity in multiple species at scale. Part of this initial push has been to develop reference genomes to accurately determine and measure genetic diversity and adaptive potential. At the same time, new methods and sequencing technologies have vastly expanded the field of conservation genomics. We have moved from measuring genetic diversity using AFLPs and microsatellites to reduced representation sequencing and more recently whole genomes. As we increase the resolution of our conservation genomic resources, we increase the challenge of our bioinformatic capabilities, data sharing and storage, coupled with the suite of technical challenges each new species brings. Using examples of Australian species, I will highlight several of the current challenges we face in conservation genomics and how we have overcome some of these. From developing reference genomes for an extinct in the wild reptile and a critically endangered frog, generating whole genome data from samples collected across time and space under various circumstances, to the challenge of calling haplotypes in complex gene families where genetic variation is extremely low. Key to all these challenges, and our innovative solutions, have been the questions we are wanting to ask of the dataset at this time and how to future-proof it against technological innovations so we can maximise our conservation applications, today and tomorrow.
The influence of gene flow on population viability in an isolated urban caracal population

Christopher C Kyriazis

CHRISTOPHER C KYRIAZIS, Laurel Serieys, Jacqueline M Bishop, Marine Drouilly, Storme Viljoen, Robert K Wayne, Kirk E Lohmueller
Panthera (USA), San Diego Zoo Wildlife Alliance (USA), UCLA (USA), University of Cape Town (South Africa)

Wildlife populations are becoming increasingly fragmented by anthropogenic development. Small and isolated populations often face an elevated risk of extinction, in part due to inbreeding depression. Here, we examine the genomic consequences of urbanization in a caracal (Caracal caracal) population that has become isolated in the Cape Peninsula region of the City of Cape Town, South Africa and is thought to number ~50 individuals. We document low levels of migration into the population over the past ~75 years, with an estimated rate of 1.3 effective migrants per generation. As a consequence of this isolation and small population size, levels of inbreeding are elevated in the contemporary Cape Peninsula population (mean FROH>1Mb=0.20). Inbreeding primarily manifests as long runs of homozygosity >10Mb, consistent with the effects of isolation due to the rapid recent growth of Cape Town. To explore how reduced migration and elevated inbreeding may impact future population dynamics, we parameterized an eco-evolutionary simulation model. We find that if migration rates do not change in the future, the population is expected to decline, though with a low projected risk of extinction. However, if migration rates decline or anthropogenic mortality rates increase, the potential risk of extinction is greatly elevated. To avert a population decline, we suggest that translocating migrants into the Cape Peninsula to initiate a genetic rescue may be warranted in the near future. Our analysis highlights the utility of genomic datasets coupled with computational simulation models for investigating the influence of gene flow on population viability.
Critically endangered Rice's whale exhibits an unexpectedly complex demographic history
Diana Aguilar Gomez

Diana Aguilar Gomez, Jacqueline Robinson, Sergio Nigenda, Chris Kyriazis, Nicole Vollmer, Lynsey Wilcox Talbot, Patricia Rosel, Phillip Morin, Kirk Lohmueller

California State University San Marcos (USA), Cooperative Institute for Marine and Atmospheric Studies (USA), National Marine Fisheries Service, National Oceanic and Atmospheric Administration (USA), San Diego Zoo Wildlife Alliance (USA), Southeast Fisheries Science Center, Southwest Fisheries Science Center, University of California Los Angeles (USA), University of California San Francisco (USA)

The Rice's whale (Balaenoptera ricei) was not formally recognized as a distinct species until 2021. They are the sole resident baleen whale species in the Gulf of Mexico, an area heavily impacted by industrialization. Rice’s whales are highly endangered with an estimated population of fewer than 100 individuals. Principal threats to affect their population recovery include ship strikes, oil spills, climate change, noise, fisheries, and other human disturbances. To support conservation efforts, we are conducting a population genetic study to assess the genomic health of these whales. We sequenced whole genomes of 20 individuals at ~30x coverage. Our findings reveal a highly inbred population, with most sequenced individuals showing significant relatedness and extensive runs of homozygosity, comprising approximately 40% of the genome. Furthermore, we also found that their demographic history is complex, suggesting drastic population fluctuations and/or genetic admixture to account for observed diversity patterns. Specifically, we inferred the existence of a ghost population from which the Rice’s whale diverged approximately 100,000 years ago, followed by a recent bottleneck. This bottleneck likely stems from anthropogenic causes, such as historical whaling. While cetaceans commonly exhibit introgression, no genetically close species near the Gulf of Mexico serve as viable candidates to validate this hypothesis, deepening the mystery. Information gathered from this genomic analysis will be integral for better understanding this rare and endangered species, and ensuring effective conservation and management planning.
Conservation Genomics of Neotropical Wild Cats

Eduardo Eizirik

Presented by self
Pontifícia Universidade Católica do Rio Grande do Sul (Brazil)

This is a very exciting time for conservation genomics, as new methods are developed and improved at a rapid pace, old questions are revisited with unprecedented amounts of data, while new questions emerge and become the focus of attention. Genomic data are now enabling a wide variety of studies targeting threatened taxa, serving as a basis for improved conservation planning and management actions on their behalf. The Neotropical region harbors a remarkable and threatened biodiversity, including many species of endemic carnivorans (Mammalia, Carnivora). Several of these species have been the focus of conservation genetic and genomic studies conducted by our research group over the last 20-30 years. For this talk, I will focus on two of these systems: small cats of the genus Leopardus and jaguars (Panthera onca). We have been studying Leopardus spp. for almost 20 ears, including phylogenetic, phylogeographic, population genetic and ecological studies, in recent years transitioning to studies based on whole-genome data. One of the main foci in this system has been the discovery and characterization of a hybrid zone between L. geoffroyi and L. guttulus in southern Brazil. I will focus specifically on an ongoing study that employs low-coverage whole-genome-sequence data for over 380 individuals sampled over a 4,000 km transect covering most of the parental ranges and the hybrid zone, which has allowed unprecedented insights on the structure, dynamics and conservation implications of this inter-species admixture process. The second focus will be on jaguars, a big cat that faces multiple threats including habitat loss and fragmentation, human-wildlife conflict and poaching for an expanding illegal market. I will briefly cover the history of genetic studies of Brazilian jaguar populations, moving into the current phase of analyses based on whole genomes and their conservation implications. I will then describe the recent development of a genome-enabled SNP panel for jaguar forensics and molecular ecology, and show results on its performance for geographic assignment (which is critical for forensic analyses of confiscated jaguar parts), sexing, individual identification and kinship assessments. I will conclude by discussing a broader framework for conservation genomics, connecting high-quality genome sequences to population-level studies and direct, integrated applications in wildlife forensics, population monitoring and genetic management.
Sea lamprey, an ancient vertebrate, is of conservation concern across its native range despite it being an invasive pest in the Laurentian Great Lakes. As an anadromous species, it is subject to habitat degradation, barriers to migration, and the increasing effects of climate change. We aim to inform conservation efforts by using genome-wide information to account for past and present effective population size and future vulnerability under climate change. We re-sequenced whole genomes for 155 sea lamprey from locations across their entire native range. We estimated the trajectory of effective population size in the recent past for each genetically distinct population and per-site contemporary effective population size. We then calculated genetic offsets to predict maladaptation to future climate change. All sea lamprey populations were found to have experienced severe effective population size declines in the recent past. Contemporary effective population size estimates identified locations of conservation concern. Genetic offsets forecasted locations at which sea lamprey are most likely to be maladapted and therefore more vulnerable under future climate change. Using these complementary methods, we make recommendations regarding conservation priorities and the potential for implementing genetic rescue as a conservation measure. This study shows the strength of applying population genomic methodology across different evolutionary timescales for the entirety of a species’ range.
Contrasting levels of genetic diversity in two endemic axolotl species from Michoacan Mexico.

Esther Denice de la Cruz Santos

Denice de la Cruz-Santos, Eria A Rebollar, Yurixhi Maldonado-López, Maria D Basanta, Anahi Canedo-Téxon, Eugenia Zarza, Gabriela Castellanos-Morales

Biology Integration Institute, Centro de Ciencias Genómicas, Ciencias y Tecnologías (Mexico), Consejo Nacional de Humanidades, El Colegio de la Frontera Sur, Instituto de Investigaciones sobre los Recursos Naturales, Universidad Michoacana de San Nicolás de Hidalgo (Mexico), Universidad Nacional Autónoma de México (Mexico), University of Nevada Reno (USA), unidad Tapachula (Mexico), unidad Villahermosa (Mexico)

Mexico is a center of diversity for axolotls (Ambystoma sp.) and most of the species are endemic to our country and are considered endangered species. In particular, obligate neotenic species, which never transition through metamorphosis, are highly threatened since their distributions are restricted to only one lagoon per species. Thus, exploring their genetic diversity is relevant to propose conservation measures. In the present study, we used a genotyping by sequencing (GBS) approach to compare the levels of genetic diversity and genetic differentiation between two critically endangered species, Ambystoma andersoni (N = 30) and Ambystoma dumerilii (N = 15), which are endemic to the Lago de Patzcuaro and Laguna Zacapu, respectively. We performed a reference-based SNP calling, using the available genome of A. mexicanum to map our reads, and retained 28,816 SNPs after filtering. We found that each species constitutes a well-defined genetic cluster without evidence of gene flow. Ambystoma dumerilii showed higher levels of genetic diversity (HE = 0.278) than A. andersoni (HE = 0.097). Lack of genetic diversity in A. andersoni raises awareness about the urgent need to conduct in situ and ex situ conservation action to raise levels of genetic variation and protect this species. For A. dumerilii habitat protection and restoration is needed to promote its permanence in the mid to long term.
Metagenomic analysis in tropical fresh and wastewater

El Colegio de la Frontera Sur (Mexico), Instituto de Biotecnología, Investigadora CONAHCYT (Mexico), UNAM (Mexico)

Highly biodiverse tropical environments are threatened due to habitat transformation and climate change among other threats. Environmental DNA and RNA methods are promising tools to uncover and monitor tropical ecosystems, but their implementation is still in early stages in these crucial environments. Here we performed a metagenomic analysis of fresh and wastewater from southern Mexico where we compared viral community composition. We detected different taxonomic profiles for viruses that reflect water use. We identified 38 virus families infecting a wide taxon range. We detected 15 different phytopathogenic viruses that belonged to the family Virgaviiridae, genus Tobamovirus. Pepper mild mottle virus and tomato brown rugose fruit virus had the largest relative abundance. Applying phylogenetic methods on assembled sequences of the coat protein (CP) we inferred the origin of plant infecting viruses. We observed a close relationship between the solanaceous infecting viruses and the reference sequences found in GenBank. On the other hand, Fabaceae and Cucurbitaceae-infecting viruses did not show such a close relationship when compared with GenBank reference sequences. These results suggest that the detected viral sequences probably originated from new strains or species of local viruses likely from the consumption of locally cultivated produce, whereas those infecting solanaceous plants might come from consumption of processed or imported food to southern Mexico. This study highlights that global food and plant trade can influence local viral communities, especially in areas altered by human activities with the potential of affecting interactions within the ecosystems and the introduction of pathogens.
Epigenome-wide scans identify DNA methylation markers for monitoring sublethal thermal stress in endangered sea turtles

Eugenie Charley Yen

Eugenie C Yen, James D Gilbert, Alice Balard, Inês O Afonso, Kirsten Fairweather, Débora Newlands, Artur Lopes, Sandra M Correia, Albert Taxonera, Stephen J Rossiter, José M Martín-Durán, Christophe Eizaguirre

Instituto do Mar (Cape Verde), Project Biodiversity (Cape Verde), Queen Mary University of London (United Kingdom)

As global warming progresses, there is a pressing need to quantify temperature-mediated sublethal effects on individuals, since they reduce average population fitness without increasing mortality. Detecting these cryptic effects in the wild is however challenging. Conservation epigenomics emerges as a promising discipline because epigenetic variation represents a direct, molecular link between the genome and the environment. DNA methylation in particular can be thermally induced and stably maintained. Here, we focused on endangered loggerhead sea turtles (Caretta caretta) that nest in Cabo Verde (West Africa). We set up a split-clutch field experiment in an in-situ hatchery, where nests of wild loggerheads were buried at different depths to expose them to different temperatures in biologically relevant conditions. Upon emergence, we collected blood samples from hatchlings for whole genome bisulfite sequencing. We identified over 700 differentially methylated sites between incubation treatments, including on genes with neurodevelopmental and cytoskeletal functions. These methylome differences were linked to their performance in multiple fitness tests of locomotion ability. Hatchlings that developed in the warmer treatment were weaker than their siblings from the cooler treatment, with direct correlations found between methylation levels and fitness measurements. Overall, we identified a set of biomarkers for monitoring exposure to sublethal thermal stress and showcase the applicability of this epigenomic approach for conservation. After successfully trialling nanopore sequencing in the field, our work also contributes to the movement towards precision conservation for protecting the health and resilience of endangered species.
Nearly-neutral theory in Asian Elephants and the conservation implications of deleterious mutation accumulation.

Gabe D O'Reilly

Gabe D O'Reilly, Rebekah Rogers, Michael Campana, Jesus Maldonado, Natalia A Prado-Oviedo, Jeroen Kappelhof
Adelphi University (USA), Biology Department, Center for Conservation Genomics, Department of Bioinformatics, Smithsonian's National Zoo and Conservation Biology Institute (USA), UNC Charlotte (USA), Wageningen University & Research - Animal Breeding and Genomics (Netherlands)

Nearly Neutral Theory states that a population's ability to purge a deleterious mutation is a function of the strength of selection and the effective population size of a population. The larger a population, the more readily it can purge deleterious mutations. However, under intense bottle necks populations can lose the ability to purge those mutations. In bottle-necked and fragmented populations, this could allow for slightly deleterious mutations to accumulate in the genome. These dynamics of nearly-neutral evolution have previously been documented in extinct elephantids. Modern populations of Asian Elephants have experienced parallel population declines and may be at similar risk of detrimental mutations. Deforestation, large infrastructure projects and regional conflicts have fragmented the Asian elephant population, leaving them endangered. Using short read genomic data from 100 Asian elephant genomes, we have investigated the population structure of Asian Elephants in South-to-South East Asia and are cataloguing their structural variation. With the aim to identify any accumulations of deleterious mutations and identify exactly how key genes are being altered by those mutations. These results will give us a much clearer idea of how the Asian Elephant population has been evolving during the Anthropocene and help clarify the circumstances under which genomic decay may occur, offering empirical bounds on evolutionary theory. Additionally, these genetic results can help steer conservation efforts for Asian elephants, a species of cultural and ecological importance in Southeast Asia.
The Role of Population History in Shaping the Mutational Load of Structural Variants Relative to SNPs, in Distinct Island versus Continental Lagopus Lineages

Gabriel David

Gabriel David, Patrik Rödin Mörch, Jacob Höglund, Theodore Squires, Kristinn Pétur Magnússon
University of Akureyri (Iceland), Uppsala universitet (Sweden)

While the importance of specific demographic histories in shaping patterns of mutational load conferred by deleterious single nucleotide polymorphisms (SNPs) has received considerable attention in the recent past, few studies have investigated the corresponding fitness consequences of structural variation in distinct evolutionary lineages. We performed heuristic-based filtering and rapid automated curation of short-read-discovered SVs callsets from 99 re-sequenced individuals across two recently (~2 million years) diverged ptarmigan (Lagopus) species. High-confidence SV callsets reveal that the relative proportion of deleterious structural variants is consistently greater in small effective population sizes, but that the relative frequency of deleterious variants differs between populations having experienced temporary bottlenecks versus longer-term low Ne. Despite the Svalbard rock ptarmigan population exhibiting the lowest coalescent Ne (Watterson’s theta) estimate for all populations, it did not carry the highest masked and realised load for both SVs and SNPs. Crucially, these differences in genetic load may reflect differences in the type of bottleneck: the Svalbard population has likely experienced a founder event during post-glacial colonisation of the islands, while the Pyrenean rock ptarmigan population likely experienced a longer-term gradual decline with recent inbreeding. Our findings demonstrate the importance of considering the nuances of population history when interpreting the potential effects of small effective population size on mutation load. Furthermore, similar to SNPs, we find that many canonical SV classes (deletions, duplications and inversions) may largely conform to nearly-neutral expectations.
The genetic architecture of an adaptive phenotype conditions vulnerability and populations evolutionary response to climate change

Genís Garcia-Erill

Genís Garcia-Erill, Francisco Campuzano-Jiménez, Perrine Gauthier, Palle Villesen, Andreas H. Faber, Marc Palmada-Flores, Sylvain Santoni, John D. Thompson, Bodil K. Ehlers, Thomas Bataillon

Aarhus University (Denmark), Bioinformatics Research Center, CEFE, CIRAD, CNRS, Departament de Biologia, Department of Ecoscience, EPHE, INRAE, IRD, Institut Agro (France), Laboratori Ictiologia Genetica, UMR AGAP Institut, Univ Montpellier, Univ Paul Valéry Montpellier (France), Universitat de Girona (Spain)

Common thyme (Thymus vulgaris) is a dominant species in the Mediterranean ecosystem that influences the composition of its associated plant communities. Thyme populations exhibit genetically determined individual chemotype variability at a very local scale. Chemotypes can be grouped by their chemical composition into phenolic and non-phenolic ecotypes, which have been shown to be adaptive for different local habitats. Ongoing climate change is already resulting in shifts in the frequency of locally adaptive ecotypes in the Mediterranean region. Understanding the evolutionary history of this polymorphism and its maintenance is important for predicting the vulnerability to climate change of thyme populations and thyme associated species in Mediterranean ecosystems and ultimately guiding conservation action. We sequenced 900 thyme individuals from over 50 sites spanning 9 gradients of altitude and environmental conditions across Southern France. Combining genetic, environmental and phenotypic data allowed us to catalogue and link genetic and phenotypic diversity to environmental variation across both local and regional spatial scales. Using a newly assembled draft reference genome, we mapped the loci that determine chemotype in thyme populations. These include structural variants at loci encoding for terpene synthases involved in the biosynthesis of the monoterpenes that define the chemotypes. We use a combination of population and ecological genetics to understand how past demography, genetic connectivity and parallel adaptations shape thyme’s functional and neutral genetic diversity. Finally, we explore with simulations how SNPs and environmental data can be used to predict the vulnerability of thyme populations and their potential to respond to climate change.
Developing new strategies to characterise genomic variation for biodiversity conservation

J. Camilo Chacón-Duque

J. Camilo Chacón-Duque, Karin Norén, Tomas Marques-Bonet, Love Dalén
CSIC-Universitat Pompeu Fabra (Spain), Centre for Palaeogenetics (Sweden), Department of Zoology, Institució Catalana de Recerca i Estudis Avançats (ICREA) (Spain), Institut de Biologia Evolutiva, Stockholm University (Sweden)

Genetic diversity is one of the three basic elements of biodiversity and understanding the full extent of this diversity within species is fundamental for their conservation. However, despite the ongoing advances in DNA sequencing technologies, conservation genomics is lagging behind in using whole-genome data efficiently. In contrast, other fields, such as human genomics, have developed theoretical and computational approaches to tackle this. If applied to endangered species, these approaches could provide a better understanding of their genetic diversity, particularly regarding the structure of this diversity across populations and its effect on adaptation to changing environments.

By leveraging two types of genetic data currently overlooked in conservation genomics – haplotype information and structural genomic variation, we are developing new analysis strategies to characterise genomic variation in endangered species. Using the Arctic fox as a model species, we are using whole-genome data from both present-day and historical samples to: 1) infer subtle patterns of population structure and gene flow, 2) quantify temporal changes in genomic structural variation and 3) assess the ability of these types of genetic data to inform on adaptation to local environments.

With these strategies, we expect to provide not only a comprehensive characterisation of genomic diversity in Arctic foxes – with a direct impact on the Swedish Arctic fox conservation programme, but also new analytical tools for conservation genomics, unlocking the full potential of whole-genome data for biodiversity conservation.
Genomic insights into a forest reserve reveal ecological stratification of R genes and CYP450 family genes

Jia Jun Ngiam

Jia Jun Ngiam, Matti Asser Niissalo, Sitaram Rajaraman, Joan Jong, Shawn Lum, Gillian Khew, Jarkko Tapani Salojärvi
Nanyang Technological University (Singapore), National Parks Board (Singapore), University of Helsinki (Finland)

The tropical rainforests in Southeast Asia are amongst the world’s most hyper-biodiverse and species-rich regions. One prominent hypothesis for the maintenance of species-rich assemblages in the tropics is the Janzen-Connell hypothesis, where plant diversity is driven and maintained by plant-pathogen coevolution. However, the eco-evolutionary dynamics of plant immunity with respect to forest ecological succession is relatively understudied, due to the paucity of baseline genomic information. Here, we sequenced and assembled draft genomes of 499 angiosperms from a conserved forest reserve in Singapore. Our analyses reveal that plant investment to defenses differs across varying plant lifeforms with a strong phylogenetic signal, suggesting shared evolutionary histories of immune responses. Interestingly, we also found a strong stratification of resistance genes and defense-related secondary metabolism across different forest successional zones. Our findings contribute novel genomic perspectives into ecological sorting in relation to the diversification of plant immunity and secondary metabolism in tropical forest ecosystems.
Elucidating the evolutionary process of animal adaptation to deserts is key to understanding adaptive responses to climate change. Here we generated 82 individual whole genomes of four fox species (genus Vulpes) inhabiting the Sahara Desert at different evolutionary times. We show that adaptation of new colonizing species to a hot arid environment has probably been facilitated by introgression and trans-species polymorphisms shared with older desert resident species, including a putatively adaptive 25 Mb genomic region. Scans for signatures of selection implicated genes affecting temperature perception, non-renal water loss and heat production in the recent adaptation of North African red foxes (Vulpes vulpes), after divergence from Eurasian populations approximately 78 thousand years ago. In the extreme desert specialists, Rueppell's fox (V. rueppellii) and fennec (V. zerda), we identified repeated signatures of selection in genes affecting renal water homeostasis supported by gene expression and physiological differences. Our study provides insights into the mechanisms and genetic underpinnings of a natural experiment of repeated adaptation to extreme conditions.
The evolutionary genomics of population’s responses to climate change
Jonas Andres Aguirre Liguori

Jonás Andrés Aguirre-Liguori, Santiago Ramírez-Barahona, Abraham Morales Cruz, Brandon S Gaut
Instituto de Biología, Lawrence Berkeley National Laboratory (USA), UNAM (Mexico), Universidad Autónoma de Yucatán (Mexico), University of California Irvine (USA)

Climate change is becoming a serious threat to biodiversity. Populations can respond to this threat by migrating to suitable areas or adapting locally to persist in the future. Usually, the impact of climate change is evaluated using species distribution models. However, these models omit evolutionary processes. Recently, researchers have used genomic tools and landscape genomics to incorporate demographic history and adaptive genetic diversity in the predictions of how populations will adapt to climate change. However, other evolutionary processes are still missing in the future projections of climate change. In this presentation, we first show how demographic history, local adaptation, gene flow, population dispersal and genomic load can be combined to predict the potential response of specific populations to climate change. Second, we analyze 5 wild Vitis species to exemplify how these layers of information can be used to identify crop wild relatives (CWRs) of Vitis species (grapes) that will be resistant in the future. We find five V. mustangensis accessions that are expected to respond adequately to climate change. Finally, we project which of these five V. mustangensis accessions could be adapted in the future if they were actively moved to locations where V. vinifera might be at risk if climate continues to change. In theory, these accessions, after experimental validations, could potentially be used to introduce climate adaptive traits in V. vinifera.
Genome connectivity of juvenile sharks Sphyrna lewini in potential nursery areas of the southern Gulf of California

Karla L. Velarde-Mota

Karla L. Velarde-Mota, Daniela G. Félix-López, Nancy C. Saavedra-Sotelo
CONAHCYT (Mexico), Division of Cardiovascular Sciences, Facultad de Ciencias del Mar, Faculty of Biology Medicine and Health, Investigadoras e Investigadores por México, The University of Manchester (United Kingdom), Universidad Autónoma de Sinaloa (Mexico)

The scalloped hammerhead shark Sphyrna lewini is a highly migratory top predator distributed in tropical and subtropical oceans worldwide. Unlike males, females are mainly distributed in coastal zones using mangrove estuaries as nursery areas (providing protection and food for pups). In many cases, females exhibit philopatry to these nursery grounds promoting population genetic structure. In the Gulf of California (GC) S. lewini is an important fishery resource, particularly there is a high abundance of juveniles and young-of-the-year (YOY) individuals. This work aims to evaluate the use of the GC as nursery areas for S. lewini. In this study, we collected muscle tissue samples from thirty-seven juvenile and YOYs from artisanal fisheries in four localities in the GC. We will assess the sibling relationships and adaptive variation through single nucleotide polymorphisms (SNPs) from the ddRADseq technique. Sibship analysis will be performed to evaluate the degree of connectivity between potential nursery areas. We will also evaluate the adaptive genetic structure as a proxy for residency from sharks within the GC. The results obtained may have implications for the conservation and management of S. lewini in the GC, due to a high degree of residency in nursery areas could indicate that the GC is a critical region for the survival of this species. Key words: Population genomics, Philopatry, Residence, Local adaptation, Shark fishery.
Estuaries serve a wide-range of functions including natural environmental buffers and other ecosystem services. These ecosystems are also critical habitats for numerous marine species, including the juvenile stages of taxa critical to marine food webs and multi-billion dollar fishing and tourism industries. However, the persistence of taxa that utilize estuarine habitats is increasingly threatened globally by rapid environmental changes resulting from widespread coastal development and urbanization. Here we use environmental DNA (eDNA) to quantify the presence of marine fishes utilizing estuarine habitats along an environmental gradient in Okinawa, Japan that spans dense human population centers to a UNESCO world heritage site. Specifically, we integrate data on species occurrence, water quality parameters, oceanographic data and human population data to assess trends in marine fish community and phylogenetic diversity. Our results reveal spatiotemporal trends across this landscape with critical implications for estuarine stewardship and the management of living marine resources.
Genomics of Brazilian howler monkeys reveals adaptation to malaria

Katherine McVay

Katherine McVay, Erick Figueroa-Ildefonso, Katharine Korunes, Luana Portela, Victor Y Guimarães, Ligia S. L. Silveira da Mota, Cauê Monticelli, João Valsecchi, Fernando A Perini, Alcides Pissinatti, Patricia Domingues de Freitas, Amy Goldberg (USA), Duke University (USA), Federal University of Minas Gerais (Brazil), Federal University of São Carlos (Brazil), Mamirauá Sustainable Development Institute (Brazil), Primatology Center of Rio de Janeiro (Brazil), São Paulo State University (Brazil), São Paulo Wildlife Coordination (Brazil)

Malaria is one of the strongest selective pressures in human evolution, with dozens of genes implicated in adaptation to different Plasmodium parasites. Over 30 Plasmodium species infect non-human primates, including regular parasite sharing with humans. Plasmodium is hypothesized to have arrived in South America ~500 years ago through European Colonization and the Trans-Atlantic Slave Trade. Thus, South American primates provide an opportunity to understand adaptation in a malaria-naïve primate system. In particular, howler monkeys (genus Alouatta) are the primary reservoirs for P. simium, a vivax-like zoonotic malaria infecting humans in southern Brazil, and regularly share P. malariae/brasilianum with humans. We conducted whole-genome sequencing on 88 howler monkeys (average coverage ~11X) to understand broad patterns of shared or unique adaptations to malaria between humans and multiple howler species. Based on 43 host genes known to interact with malaria (adapted from Ebel et al. 2017, PLoS Genetics), we find preliminary evidence for shared and species-specific variation and adaptation in genes associated with red blood cell structure and invasion using a combination of allele-frequency- and haplotype-based summary statistics. Additionally, we detect species-level infections with Plasmodium parasites in 14 howler individuals using KrakenUniq, with a higher proportion of infections in the Legal Amazon than other regions. We further reconstruct broader population structure and demographic history of 4 howler monkey species from across Brazil, finding differences in rates of inbreeding and population size by species.
Structural genomic variation in the inbred Scandinavian wolf population contributes to the realized genetic load but is positively affected by immigration

Linnéa Smeds

Linnéa Smeds, Lars SA Huson, Hans Ellegren
Penn State University (USA), Uppsala University (Sweden)

The Scandinavian wolf population consists of 500 individuals, and is threatened by large hunting pressure and governmental aims to further decrease the population size. Due to a small number of founders and geographical isolation, Scandinavian wolves are extremely inbred, and a previous study using SNP data have shown that the realized genetic load has increased over generations since the founding event. In this study we expand our work to structural variants (SVs), an under-explored type of mutations potentially affecting a larger fraction of the genome than SNPs. We analyzed whole-genome, short-read sequences from 212 wolves sampled in Scandinavia and from neighboring populations in Finland and Russia. More than 26,000 high-confidence variants were detected after stringent filtering and manual curation. The majority of variants were shorter than 1 kb, with a distinct peak in the length distribution at 190 bp, corresponding to insertion events of SINE/tRNA-Lys elements. The site frequency spectrum of SVs in protein-coding regions was significantly shifted towards rare alleles compared to putatively neutral variants, consistent with purifying selection. The realized genetic load of SVs in protein-coding regions increased with inbreeding levels in the Scandinavian population, while immigration provided a genetic rescue effect by lowering the load and reintroducing ancestral alleles at loci fixed for derived SVs. Our study shows that structural variation comprises a common type of in part deleterious mutations in endangered species and that establishing gene flow is necessary to mitigate the negative consequences of loss of diversity.
Whole genome sequencing of Greenland muskox (Ovibos moschatus) individuals using surface-deposited skeletal elements

Mijin Park

Mijin Park, Juhyeon Lee, Donghee Kim, Raffaela Bianco, Johannes Krause, Tae-Yoon S. Park, Won Young Lee, Choongwon Jeong
Department of Archaeogenetics, Division of Glacier & Earth Sciences, Institute for Data Innovation in Science, Korea Polar Research Institute (Korea, Max Planck Institute for Evolutionary Anthropology (Germany), Republic of), School of Biological Sciences, Seoul National University (Korea)

The muskox (Ovibos moschatus) is a large-bodied ungulate species adapted to the tundra climate in Canada and Greenland. The current natural population in Greenland is assumed to have expanded from the Canadian Arctic Archipelago to north and northeast Greenland approximately 2,000 years ago through a serial founder event. However, their initial expansion process to Greenland as well as subsequent gene flow has not been thoroughly studied due to limited accessibility to muskox samples suitable for genome sequencing. In this study, we collected surface-deposited muskox skeletal elements from Ella Island in northeast Greenland and assessed their suitability for the non-disruptive source of genome information. Firstly, by comparing 13 libraries from 10 individuals, our analysis revealed that muskox horn shafts exhibit superb level of DNA preservation, making them easily accessible and high-quality targets for sampling in the field. Secondly, genomes of the newly sequenced Ella island individuals (n=9; 0.1-43.8x coverage) showed that they form a clade with previously published individuals from nearby sites, supporting that the surface-deposited horn samples can yield high-quality genomic data. Finally, using the newly produced high-coverage genomes, we quantified the extreme reduction in genetic diversity over time in the Greenland populations through the Multiple Sequentially Markovian coalescent (MSMC) analysis. In summary, our study presents an accessible non-disruptive sampling strategy for muskox conservation genomics, along with high-quality whole genome sequences from eastern Greenland.
Genetic structure of Shorea curtisii and S. leprosula (Dipterocarpaceae) in Southeast Asia using MIG-seq

Misato Ogasahara

Misato Ogasahara, Rahayu Sukmaria Sukri, Faizah Metali, Kenzo Tanaka, Reiji Yoneda, Mohamad Alias Azani, Eyen Khoo, Melvin Gumal, Bibian Diway, Iskandar Z. Siregar, Koichi Kamiya
Asian School of the Environment, Department of Silviculture, Development and Innovation Division (RDID), Ehime University, Faculty of Forestry and Environment, Faculty of Science, Forest Department Sarawak (Malaysia), Forestry and Forest Products Research Institute (Japan), Graduate School of Agriculture, IPB University (Indonesia), Institute for Biodiversity and Environmental Research, Japan International Research Center for Agricultural Sciences (Japan), Matsuyama (Japan), Nanyang Technological University (Singapore), Research, Sabah Forestry Department (Malaysia), Sarawak Forestry Corporation (Malaysia), Shikoku Research Center, Universiti Brunei Darussalam (Brunei Darussalam), Universiti Putra Malaysia (Malaysia)

In the last glacial period, it is thought that tropical rainforests were distributed widely on the exposed Sunda Shelf by lower sea level, but the savanna corridor interfered with east-west dispersal of rainforest species. Population genetic studies revealed a genetic divergence between “Sumatra and the Malay Peninsula” and “Borneo” for the Dipterocarpaceae, a major tree family in Southeast Asia. Divergence time and presence or absence of migration after divergence can differ between species, possibly reflecting restrictions on gene flow in the past. To understand details of vegetation in Southeast Asia in the past, population genetic analysis of NGS data for several species widely distributed in this region were utilized. In this study, MIG-seq was used to assess the geographic distribution of genetic variation in S. curtisii and S. leprosula. The analysis showed that a clear pattern of genetic structure was found in Bornean populations, but such a clear structure was not shown among populations in the Malay Peninsula. This suggests that the extent and distribution of refugial populations and population size changes differed between the two regions.
Mutations in different parts of genes can have very different effects on phenotypes and can be important for understanding how species adapt to local environments. Changes in regulatory regions of genes can shape time-dependent cellular and behavioral processes that underly pre and post-zygotic reinforcement mechanisms. Gopherus agassizii and G. morafkai are two recently speciated desert tortoise lineages that reside in adjoining but seasonally distinct deserts. The species are differentially adapted to rainfall and brumation patterns and have different disease instances. We used whole genome sequencing of 21 individuals to characterize differences between the two species. Eighty percent of mutations found in the 1% most diverged portion of the genome mapped to intergenic regions. Mutations within genic regions were largely found in promoter regions. Gene enrichment of promoter changes revealed UV nucleotide excision repair and the circadian/circannual regulation of gene transcription to be enriched. Changes to the regulation of these genes could affect the timing of biological functions related to key reproductive behaviors like brumation, mating, and laying, reinforcing isolation. Genomic regions with high exonic divergence included mucosal genes. Further investigation found extensive expansion of Mucin 5 genes with differential duplication, pseudogenization, and retention between the two lineages. Mucosal genes are important in osmoregulation and host defense against bacterial pathogens. Overall, we show regulatory divergence could underly transcriptional cascades that underlie the offset of reproductive behaviors between the species, and gene expansion that could play a role in the differential disease response among these species of conservation concern.
Exploring the Bobcat Holobiome: Applications of Conservation Multiomics at the Wildland Urban Interface

Natalie Payne

Natalie Payne, Desiree Andersen, Robert Davis, Cheryl Mollohan, Kerry Baldwin, Albert L. LeCount, Koenraad Van Doorslaer, Leigh Combrink, Melanie Culver

Arizona Cooperative Fish and Wildlife Research Unit, Arizona GIS, Bobcats in Tucson Research Project © (USA), Cancer Biology Graduate Interdisciplinary Program, Department of Immunobiology, Genetics Graduate Interdisciplinary Program, LLC (USA), School of Natural Resources and the Environment, The Bios Institute, U.S. Geological Survey, UA Cancer Center, University of Arizona (USA)

Rapid advances in genomics in the 21st century provide higher resolution for answering questions related to wildlife conservation. However, conservation genomic studies focusing on host organisms alone provide incomplete pictures of biological processes affecting population viability. Characterization of interactions between individuals and within ecosystems of wildlife holobionts is crucial for understanding organismal health and can yield important insights into disease spread and environmental change. Here, we describe the use of a multiomics dataset from a population of radiocollared bobcats (Lynx rufus) straddling the wildland urban interface (WUI) in the Tucson Mountains, Arizona, USA. We aimed to investigate how kinship and urbanization influence bobcat social dynamics and microbial communities. Integrating paired host genomic and longitudinal GPS data uncovered extensive home range sharing of adult mother-daughter pairs. As with other felids, female bobcats typically maintain exclusive home ranges. Our results provide evidence of a more complex matrilineal social structure in this solitary felid. Virome characterization has revealed numerous feline and canine pathogens with the potential to impact wild and domestic carnivore health at the WUI. Models of virome and microbiome (prokaryotic and eukaryotic) composition along with host location and genomic datasets reveal the roles of kinship and urbanization in shaping communities of host-associated taxa. We highlight the importance of incorporating genomic research into field studies to investigate ecological hypotheses. With the current rate of anthropogenic ecosystem alteration, investigating population changes at a microbial level reveals early indicators of threats to wildlife population viability and the health of humans and domestic animals.
Biodiversity loss is reaching critical levels, with over 42,100 species currently facing extinction. Understanding the genetic composition of inbred populations is crucial for guiding conservation strategies, especially in the face of rapid environmental changes induced by climate and human activities. Yet, our understanding of genetic variation’s effects remains limited. Particularly, identifying the distribution, prevalence, and impacts of pathogenic variants in non-model organisms is largely unexplored. The rapid increase in genomic data, coupled with new developments in statistical modeling, has opened the door to new modeling opportunities to identify pathogenic variants. Capitalizing on the vast repository of genomic data available, we leverage insights from sequence variation distribution across species and employ deep probabilistic models to identify and categorize pathogenic variants in endangered species. This approach yields predictions enriched by a comprehensive understanding of protein context and a notion of fitness across the vast spectrum of proteins and their evolutionary histories. Using the black-footed ferret (Mustela nigripes) as a case study—a highly inbred and endangered species—we uncover distinct trends in pathogenicity scores among pre- and post-founder event samples. We also observe sex differential mutation burdens and an enrichment of pathways related to reproduction in predicted pathogenic variants. By doing so, we not only illuminate the genetic health of endangered species but also test the broader utility of unsupervised deep learning models in cross-species variant effect prediction. Ultimately, this research not only enhances our understanding of genetic variation’s impacts but also informs conservation strategies to safeguard biodiversity.
A paleogenomic approach to reconstruct historical responses of coral reefs to anthropogenic change
Raul A Gonzalez-Pech

Raul A Gonzalez-Pech, Stephanie Marciniak, Colin Howe, Raphael A Eisenhofer-Philipona, Sofia Roitman, Julia E Cole, Laura Weyrich, Mateo Lopez-Victoria, George Perry, Monica Medina
McMaster University (Canada), New England Biolabs (USA), Pontificia Universidad Javeriana (Colombia), The Pennsylvania State University (USA), University of Copenhagen (Denmark), University of Michigan (USA)

Coral reefs are declining worldwide due to anthropogenic-driven environmental change. The foundation and health of these ecosystems rely on the harmonic functioning of all members of coral holobionts, i.e., cnidarian host, symbiotic microalgae (families Symbiodiniaceae and Ostreobiaceae), and associated microbiome. Coral stress responses often involve shifts in the taxonomic identity of their symbionts and microbiomes. Tracing back changes in coral holobiont composition over prolonged time periods can help us reconstruct health history of reefs and gain a better understanding of coral response to current stressors. Here, we focused on a major Caribbean reef-builder coral, Orbicella faveolata, from the Varadero Reef, Colombia. This reef has undergone extensive freshwater sediment discharge for decades as result of urbanization. We show, for the first time, that a paleogenomic approach can be used to reconstruct historical, microbial shifts of coral holobionts, potentially associated with stress.
A major goal of conservation biology is to maintain high genetic diversity to ensure the resilience and adaptive potential of species given rapid global changes. However, high variation may also support deleterious alleles, or genetic load, increasing the risk of inbreeding depression if population sizes decrease. New sequencing technologies and analytical methods allow the measurement of genetic load in non-model organisms, and the purging of deleterious variation has been demonstrated in some threatened species. However, less is known about the costs of declines and inbreeding in species with large population sizes and high genetic diversity even though this encompasses many species globally that are expected to undergo declines. Caribou is a species of ecological and cultural significance in North America with a wide distribution and extensive phenotypic variation but with some populations undergoing significant declines resulting in their at-risk status in Canada. We assessed intra-specific genetic variation, adaptive divergence, inbreeding, and genetic load across populations with different demographic histories using a chromosome-scale reference genome and 66 whole-genome sequences. We found caribou to have high genetic diversity, comparable to the most diverse mammal species, and adaptive diversification of genes, but also high genetic load among lineages. We found highly divergent levels of inbreeding across individuals, including genomic erosion and the loss of alleles by drift but not increased purging in inbred individuals. Our results highlight the “double-edged sword” of genetic diversity that may be representative of other species at risk affected by anthropogenic activities.
Low Rates of Gene Duplication in an Endangered Freshwater Mussel
Rebekah L. Rogers

Rebekah L. Rogers, Gabe O'Reilly, Jeffrey T Garner, John Wares
Alabama Department of Conservation (USA), UNC Charlotte (USA), University of Georgia (USA)

Nearly neutral theory predicts that evolutionary processes will differ in small populations compared to large populations, a key point of concern for endangered species. The nearly-neutral threshold, the span of neutral variation, and the adaptive potential from new mutations all differ depending on Ne. To determine how genomes respond in small populations, we have created a reference genomes for a US federally endangered IUCN Red List freshwater mussel, Elliptio spinosa and compare it to genetic variation for a common and successful relative Elliptio crassidens. We find higher rates of background duplication rates in E. spinosa consistent with proposed theories of duplicate gene accumulation according to nearly-neutral processes. Along with these changes we observe fewer cases of adaptive gene family amplification in this endangered species. However, TE content is not consistent with nearly-neutral theory. We observe substantially less recent TE proliferation in the endangered species with over 500 Mb of newly copied TEs in Elliptio crassidens. These results suggest a more complex interplay between TEs and duplicate genes than previously proposed for small populations. They further suggest that TEs and duplications require greater attention in surveys of genomic health for endangered species.
Leveraging synteny to generate reference genomes for conservation: Assembling the genomes of Hector\'s and M?ui dolphins

Massey University (New Zealand), Oregon State University (USA), Technical University of Munich (Germany), University of Auckland (New Zealand), University of Otago (New Zealand)

Escalating concern regarding the impacts of reduced genetic diversity on the conservation of endangered species has spurred efforts to obtain chromosome-level genomes through consortia such as the Vertebrate Genomes Project. Assembling reference genomes for many threatened species remains challenging due to the difficulty in obtaining optimal input samples, that can characterize long-term conservation collections. Here, we present a pipeline that leverages genome synteny to construct high-quality genomes for species of conservation concern despite less-than-optimal samples, demonstrating its use on Hector\'s and M?ui dolphins. These endemic New Zealand dolphins are threatened by human activities. Hector\’s dolphins are classified as endangered by the IUCN, while the M?ui dolphin is among the most critically endangered marine mammals. To assemble reference genomes for these dolphins, we created a pipeline combining de novo assembly tools with reference-guided techniques, utilizing chromosome-level genomes of closely related species. The pipeline assembled highly contiguous chromosome-level genomes (scaffold N50: 110 MB, scaffold L50: 9, miniBUSCO completeness scores >96.35%), despite non-optimal input tissue samples and sequencing data. We demonstrate that these genomes can provide insights relevant for conservation, including historical demography revealing long-term small population sizes, with subspecies divergence occurring ~20kya, linked to the Last Glacial Maximum. M?ui dolphin heterozygosity was 40% lower than Hector\’s and comparable to other cetacean species noted for reduced genetic diversity. Through these exemplar genomes, we demonstrate that our pipeline can provide high-quality genomic resources to facilitate ongoing conservation genomics research.
Cryptic diversity in Eurasian minnows: Implications for the conservation of German freshwater fishes

Temitope Opeyemi Oriowo

Temitope O Oriowo, Ioannis Chrysostomakis, Nils Sternberg, Astrid Böhne, Madlen Stange
(Germany), Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig, Museum Koenig (Germany)

Recent studies have revealed previously unknown cryptic diversity within Eurasian minnows (Leuciscidae – Phoxinus phoxinus), challenging its monotypic classification in Europe (Palanda?i? et al., 2015, 2017). These findings have profound implications for conservation and taxonomy, particularly in Germany, where multiple Phoxinus species coexist. The recognition of multiple genetically distinct lineages within Phoxinus led to a revision of German freshwater conservation laws to reflect the conservation needs of each recognised species (Bundesamt für Naturschutz, 2023). The existence of several species in Germany and the historical stocking and reintroduction programmes prompted us to understand the degree of differentiation and hybridisation between different Phoxinus species and populations, with a focus on the Middle Rhine. We conducted population structure analyses using our annotated, chromosome-resolved reference genome (Oriowo et al. 2024) and whole-genome datasets of over 300 Phoxinus individuals from various catchments in and around Germany. The analysis of more than 6 million SNPs revealed highly differentiated populations in various river catchments, such as the Meuse, Seine, and Rhône, with minimal to no hybridization with non-native species. In the Upper and Middle Rhine, we identified a distinct lineage within lineage ‘5b’ (P. csikii), genetically distinct from the ‘Danube lineage 5b’. In the Lower Sieg, a tributary to the Middle Rhine near Bonn that experienced high degrees of stocking, an analysis of hybridisation revealed a high degree of interbreeding between lineages ‘10’ (P. phoxinus) and ‘5b of the Rhine type’. These findings underscore the importance of whole-genome analyses in conservation relevant species.
Towards genome-wide data of Arctic species obtained from snow footprints

Theo Phanu Serivichyaswat

Theo Phanu Serivichyaswat, Samantha López-Clinton, Chenyu Jin, Anja Mezger, Marc Friedländer, Tom van der Valk
- (Sweden), Centre for Palaeogenetics (Sweden), Department of Bioinformatics and Genetics, Department of Molecular Biosciences, Department of Zoology, National Genomics Infrastructure (Sweden), Science for Life Laboratory, Stockholm University (Sweden), Swedish Museum of Natural History (Sweden), The Wenner-Gren Institute

Environmental DNA (eDNA) analysis has emerged as a powerful and non-invasive tool for monitoring species without direct observation, but the genomic data obtained has extremely low coverage and is thus insufficient for population genetic studies. This study focuses on utilizing single-cell sorting, whole genome amplification techniques and Nanopore sequencing to assess the possibility of obtaining genome-wide data from environmental samples. Through sampling snow footprints, we aim to non-invasively obtain whole genomes of multiple arctic species. Our methodology involves sample preservation, single-cell sorting, genome amplification, next-generation sequencing and bioinformatic analysis of genome-wide data. By coupling the current techniques with population genomic analysis, we aim to gain invaluable insights into the spatial distribution, population dynamics, and habitat preferences of arctic species, aiding in their monitoring and conservation efforts. Our findings will underscore the utility of this novel approach as a tool for biodiversity and population monitoring, particularly for species in the Arctic that are challenging to study using traditional methods.
Using comparative transcriptomics to understand response to heat stress in polar fishes
Tina Cai

Tina Cai, Samuel N Bogan, Joanna L Kelley
Santa Cruz (USA), University of California

Anthropogenic climate change is one of the most pressing issues globally, increasing risks of species extinction and biodiversity loss. While the polar and sub-polar regions are experiencing an accelerated impact due to declining sea-ice cover and warming, species that have adapted to these extremely cold and stable environments might be the most susceptible to the drastic changes. Many fish species have adapted to live in the subfreezing water temperatures, with adaptations including the ability to produce antifreeze proteins and the loss of an inducible heat shock response. There is currently a limited understanding of how polar fish species will respond to increased climatic disturbances. Here, we conduct a reanalysis of existing transcriptome studies to understand how polar fishes and these specialized functions respond to elevated heat stress. We compare gene expression among different fish species under varying levels of heat exposure across life stages, tissues, and types of antifreeze proteins. Transcriptomic data provides an opportunity to investigate whether gene expression responses to heat stress are conserved across polar fish lineages. We expect to observe a difference in sensitivity to heat stress across tissue types depending on the magnitude and duration of thermal exposure. Activation of genes linked to DNA repair and metabolism response may be essential for the preservation of cellular functions across lineages. Our findings will elucidate the mechanisms by which polar fishes respond to heat stress, contributing to a better understanding of their capacity to adapt to rising global temperatures.
Decoding the population structure and history of the world’s deadliest cat: (Felis nigripes) the black-footed cat

Victoria B Grant

Victoria B Grant, Shreya Banerjee, Jonas Oppenheimer, Katherine A Solari, John J Baczenas, Beth Shapiro, Dmitri Petrov, Ellie E Armstrong, Molly Schumer

Davis (USA), Santa Cruz (USA), Stanford University (USA), University of California, Washington State University (USA)

Advances in genomic sequencing have opened new opportunities for studying species’ genetic diversity and evolutionary past to address conservation concerns. Listed as vulnerable to extinction by the IUCN Red List, black-footed cats (Felis nigripes) are one of Africa’s rarest cats. The only cat residing in the arid and semi-arid parts of southern Africa, very little is understood about black-footed cats because of their small size and nocturnal behavior. Moreover, black-footed cats live in isolation and their intraspecific interactions and demographics are not well described. In such cases where ecological knowledge is lacking, range-wide population genomics can improve our understanding of population dynamics. Here, we sequenced whole genomes of black-footed cat individuals from across their species’ range in the first genomic study of free-roaming individuals. To do so, we incorporated whole genomes 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 for population analyses. We found evidence of larger populations in the past with a prolonged population decline in the last 10,000 years. Furthermore, we found remarkably low genome-wide diversity (θ ~ 0.0004) suggesting a recent bottleneck. We compared individuals from across the range to evaluate patterns of population structure and found higher genetic similarity between individuals in close geographic proximity. Overall, these results provide a range-wide and long-term baseline informing the conservation of this enigmatic species.
S21 - Decoding the past to safeguard the future.
Ancient environmental DNA from the Pliocene High Terrace deposits in the Canadian High Arctic

Bianca De Sanctis

Bianca De Sanctis, Alexandra Rouillard, Kurt H Kjær, Mikkel W Pedersen, Alberto Reyes, Natalia Rybczynski, Lasse Vinner, Eske Willerslev
Canadian Museum of Nature (Canada), Umeå University (Sweden), University of Alberta (Canada), University of Copenhagen (Denmark), University of Santa Cruz (USA)

The Pliocene High Terrace deposits on Ellesmere Island in Canada’s High Arctic are known for their exceptional preservation and high fossil yield, especially the Beaver Pond and Fyles Leaf Beds sites. Extensive paleoecological study of these sites has uncovered a larch-dominated boreal forest with an array of wetland flora, grasses, extinct species of beaver, rabbit and wolverine, and even a High Arctic camel. These strata date to approximately 3.9 million years ago, when mean annual temperatures in the Arctic were 14-22C higher than they are today. Here, we report preliminary results of DNA isolated and shotgun sequenced from sediment samples collected from Beaver Pond and Fyles Leaf Beds.
Metagenomic reconstruction of the plant community in the stomach content of a steppe bison living over 48,000 C14 years ago

Chenyu Jin

Chenyu Jin, Adrian Forsythe, Miklós Bálint, Frank Kienast, Pavel Nikolskiy, Katerina Guschnanski, Tom van der Valk

-- (Sweden), Centre for Palaeogenetics (Sweden), Department of Bioinformatics and Genetics, Department of Ecology and Genetics, Department of Insect Biotechnology, Department of Zoology, Evolutionary Biology Centre, Geological Institute, Institute of Ecology and Evolution, Justus-Liebig University (Germany), LOEWE Centre for Translational Biodiversity Genomics (Germany), Russian Academy of Sciences (Russian Federation), School of Biological Sciences, Senckenberg Biodiversity and Climate Research Centre (Germany), Senckenberg Research Institute and Natural History Museum (Germany), Stockholm University (Sweden), Swedish Museum of Natural History (Sweden), The University of Edinburgh (United Kingdom), Uppsala University (Sweden)

By reconstructing past ecosystems, we can gain valuable context for understanding how extant taxa are distributed and how ecosystems may respond to changing climate in the future. During the last cold stage (115 - 11.7 kya), Arctic Eurasia exhibited a distinctive plant composition consisting of steppe and xerophilous arctic plants adapted to a dry climate. This combination formed the Pleistocene mammoth steppe, which provided a habitat for large herbivores like mammoths (Mammuthus primigenius) and steppe bison (Bison priscus). In this study, we take a unique opportunity to analyze a plant paleo-community of the Arctic mammoth steppe through DNA sequencing of a well-preserved stomach content of a steppe bison, which was discovered in the Kolyma Lowlands in northeast Russia in 2009 and dated to >48,000 C14 years before present. Macrofossil analysis of this sample revealed a floral composition resembling a saline meadow. In line with macrofossil evidence, we identified alkali grass (Puccinellia sp.) and Larch (Larix sp.) in the DNA data of the steppe bison sample. We also confirmed the presence of plant genera such as Artemisia, Allium, Silene, Poa, Carex, and Dryas, which are typically associated with arid soil conditions, whereas the detected genera Puccinellia, Festuca, Poa, and Potamogeton are indicative of high salinity. This list includes taxa previously not identified by macrofossil analyses and allows us to obtain a more comprehensive view of the mammoth steppe flora and steppe bison dietary preferences.
Phylogenomic reconstruction and evolutionary history of the Nothofagus of South America
Gabriela Narváez

Gabriela Narváez, Daly Noll, Josefina Jorquera, Roberto Nespolo, Juliana Vianna
Center of Applied Ecology and Sustainability (CAPES) (Chile), Facultad de Ciencias, Institute Millennium Center for Genome Regulation (CRG) (Chile), Instituto de Ciencias Ambientales y Evolutivas, Millennium Institute Biodiversity of Antarctic and Subantarctic Ecosystems (BASE) (Chile), Millennium Institute for Integrative Biology (iBio) (Chile), Millennium Nucleus Limit of Life (LiLi) (Chile), Pontificia Universidad Católica de Chile (Chile), Universidad Austral de Chile (Chile), Universidad de Chile (Chile)

The Nothofagus genus has attracted considerable attention due to its complex biogeographic history in the Southern Hemisphere. However, the lack of genome-wide data and the limitation of obtaining samples representative of the geographical distribution of the different species still represent a challenge to understanding speciation in forest environments and tree plants in general. In South America, the Nothofagus forests are mainly conformed by the subgenus Nothofagus, widely distributed across the Andes Mountains of Chile and Argentina. We generated high-coverage whole genome resequencing data from 10 representative individuals of the five species of subgenus Nothofagus (N. pumilio, N. antarctica, N. dombeyi, N. nitida and N. betuloides). We have reconstructed the evolutionary history of this group, the phylogenetic tree, and raised the contribution of introgression. We have also compared the gene content and orthologous relationships between these species and other Nothofagus trees. Overall, we highlight interesting features of this subgenus and suggest that the resources presented herein contribute to the complete understanding of the speciation and reconstructions of deep phylogenetic relationships such as those that have been unresolved and in constant debate in all genus Nothofagus. Therefore, this project will raise high-impact scientific knowledge for Nothofagus, contributing to a better understanding of our local biodiversity, and thus allowing research opportunities to open in this area from the global south.
The domestic pigeon (Columba livia) is one of the most important model organisms, playing a key role in developing Darwin’s Theory of Evolution. In contrast, its parental species, the rock dove, remains largely understudied, and questions regarding its evolution, taxonomy, and conservation status remain unresolved. Additionally, experts warn of the rock dove’s imminent risk of extinction due to genetic replacement by the almost ubiquitous feral pigeon. In order to describe, for the first time, the rock dove’s genetic diversity prior to the feral pigeon worldwide expansion, we generated whole-genome sequencing data from 65 historical samples representing all recognized rock dove subspecies. Remarkably, the characterization of the rock doves’ diversity and gene-flow patterns revealed a highly divergent population from the West of Africa that potentially represents a different species. We hypothesize that this population diverged as a result of a long history of allopatric cycles, which occurred during the quaternary glacial and interglacial periods. We also detect high levels of inbreeding and low heterozygosity across the West African rock doves, highlighting the need for its conservation status revaluation. Finally, our historical dataset revealed admixture signals between domestic/feral pigeons and wild rock doves, confirming that the expansion of the feral birds, favoured by ongoing urbanization and environmental deterioration, represents the main threat to the rock dove populations. This study represents an important step toward the conservation of the species, allowing the identification of genetically relevant populations on which to focus future conservation strategies to ensure the rock dove’s preservation.
The impact of whaling in the Northeast Atlantic right whale population (Eubalaena glacialis) using palaeogenomics

Gonçalo E. Themudo, Alba Rey-Iglesia, Paula F. Campos
CIIMAR, The Globe Institute, University of Copenhagen (Denmark), University of Porto (Portugal)

The North Atlantic (NA) right whale (Eubalaena glacialis) was a primary target for early industrial whalers, particularly the Basques, who extensively hunted them for their high-quality oil. Once abundant on both sides of the North Atlantic, inhabiting its temperate and subpolar waters, these whales are now functionally extinct in the Northeast Atlantic and critically endangered according to the IUCN Red List. Despite their fragmentary remains making identification challenging, bones are abundant in the archaeological and historical records and reveal a long history and extensive tradition connected to the exploitation of these marine resources. In this study, using state of the art ancient DNA techniques we aim to genetically characterize the functionally extinct population of NA right whales from the Northeast Atlantic. We sequenced 17 full genomes from historical specimens collected from the seabed near Medieval whaling ports in the Cantabrian Sea, dating from the 13-18th century. Data gathered from these specimens was compared with genomes of modern individuals from the extant population in the Northwest Atlantic. Our results highlight differences in genetic diversity over time, shedding light on the evolutionary impact of whaling on this species.
Genomic erosion through time in avian species undergoing population collapse

Hernán E. Morales

Hernán E Morales, Georgette Femenling, Claudia Fontsere, Xufen Liu, Samuel Speak, Ester Milesi, Barry Hartup, Andy Caven, Ross Crates, Rachel Bristol, Molly Cassatt-Johnstone, Beth Shapiro, Jim Groombridge, Cock van Oosterhout, Tom Gilbert

Australian National University (Australia), DTU (Denmark), INTERNATIONAL CRANE FOUNDATION (USA), Independent consultor (Seychelles), McGill University (Canada), University of California Santa Cruz (USA), University of Copenhagen (Denmark), University of East Anglia (United Kingdom), University of Kent (United Kingdom)

The escalating biodiversity crisis threatens species with the loss of essential genomic diversity needed to survive and adapt to future environmental change. Of particular concern is that population decline may have a delayed effect on the genomes of threatened species, relative to demographic processes. This delay can result in genomic erosion, even after population size recovery, a phenomenon known as the 'genetic drift debt.' Temporal genomics can offer the necessary resolution to understand these changes. We compared re-sequenced whole genomes from pre-decline (100+ year-old museum samples) and post-decline (modern samples) populations in three closely monitored endangered bird species: the Seychelles paradise flycatcher (recovered from N=28), the whooping crane (recovered from N=16), and the regent honeyeater (declining N<300). These species experienced distinct trends of population decline and conservation management actions over recent decades. Employing a population genomics approach we investigated how different types of genomic variation (deleterious, functional, neutral) have responded to population declines and subsequent conservation efforts. We show how despite their demographic recovery, past declines have left a strong genomic signature, with loss of genetic diversity, inbreeding and accumulation of genetic load. We showcase that the loss of genomic diversity, and the risk of extinction are influenced by demographic trajectories, the severity of their decline, connectivity, and conservation management strategies. Using forward-in-time simulations, parameterized with empirical temporal data, we make predictions of future trajectories of genomic erosion and identify how we can implement effective conservation actions.
Human-mediated animal translocation has significantly altered global ecosystems. In remote Wallacean and Melanesian islands, the origin of wild pigs (genus Sus), ecologically important yet sometimes considered pests, remains uncertain due to potential human introduction and natural dispersal. Determining the extent of human involvement in the spread of pig species and the timing of these events is crucial for conservation strategies. Genomic analyses of over 100 wild pigs genomes, including from museum and archaeological specimens, revealed a mosaic ancestry east of the Wallace Line, reflecting complex natural dispersal and human translocation events. Ancestry deconvolution revealed evidence for natural dispersal in some parts of Wallacea, such as Flores, where wild pigs coexist with predators like Komodo dragons. The major ancestry components of Wallacea and Melanesian wild pigs, however, can be traced to mainland Chinese domestic pigs introduced over 3,000 years ago during the Austronesian expansion, and to European domestic pigs introduced during colonization. Our analyses provide evidence to establish the origin of wild pigs and the duration of their presence in these tropical ecosystems to inform appropriate conservation strategies.
Harnessing the past to predict the future: the evolution and future adaptive potential of seasonal camouflage in white-tailed jackrabbits

Mafalda Sousa Ferreira

Presented by self
Stockholm, Stockholm University, Sweden

Adaptation from standing genetic variation is a critical component of evolutionary responses to rapid environmental change. However, the difficulty of identifying the genetic basis of fitness-relevant traits in natural populations has limited the direct incorporation of genotype-to-phenotype information into conservation efforts. We studied the evolution of adaptive winter color variation in the white-tailed jackrabbit (Lepus townsendii), a North American species undergoing population declines. Using extensive museum records, we show that winter pelage color closely tracks dynamics of snow cover across the range of white-tailed jackrabbits, suggesting that geographic variation for the trait is maintained by strong selection. Using whole genomes of specimens collected during winter, we show that seasonal camouflage variation was primarily determined by additive genetic variation at three pigmentation genes. Using ecological and genetic modeling and forecasted environmental parameters, we predict that future declines in snow cover will strongly favor darker winter phenotypes across much of the white-tailed jackrabbit distribution. We also predict that low levels of standing adaptive variation should enable severely mismatched populations to adapt to this shift in snow cover conditions. However, adaptation to future snow cover may be impeded by ongoing population declines that appear to differentially threaten adaptive standing genetic variation. Our study illustrates how evolutionary genomics can be used to identify functional genetic variation of critical importance for climate change adaptation.
Reconstructing the population history of the endangered Hispaniolan solenodon, one of the most evolutionary distinct mammals on earth

Maria H Zicos, Selina Brace, Laurent Frantz, Samuel Turvey

Institute of Zoology at the Zoological Society of London (United Kingdom), Ludwig Maximilian University of Munich (Germany), Natural History Museum (United Kingdom), Queen Mary University of London (United Kingdom)

Island systems like the Caribbean have been focal in researching evolutionary patterns and processes. Islands are particularly relevant for reconstructing colonisation and radiation events, but also for preserving ancient populations and biodiversity with important conservation implications for understanding the evolutionary history of insular species. The insular Caribbean was colonized by non-volant land mammals that diversified into more than 120 different species, however the region suffered from a massive Holocene extinction event and few terrestrial mammalian species remain. The extant fauna includes two representatives of the Solenodontidae, the Hispaniolan solenodon (Solenodon paradoxus) and Cuban solenodon (Atopogale cubana). Solenodons are nocturnal insectivores and the only living mammals with dental venom delivery systems, forming an ancient lineage that diverged from other living mammals c.73 million years ago. Solenodons are considered endangered and coupled with their evolutionary distinctiveness and current threats to survival, both species have been identified as global priorities for mammal conservation. To explore the population history of these high priority species this study has generated six new solenodon genomes using ancient DNA methodologies: five for the Hispaniolan solenodon, one of which is the first genome for the previously described south-western population, and the first genome for the Cuban solenodon. Our genome-wide data supports structure into three populations as previously suggested from mitochondrial only data, with limited gene flow between biogeographical regions of the island. Further, low genome-wide heterozygosity was identified, comparable to that of endangered eulipotyphan relatives. A more contiguous genome assembly will allow investigations into genomic erosion.
The Mexican mormoopids inhabit one of the most biologically complex regions globally, leading to distinct genetic structuring patterns among species distributed partially or entirely within this area. Recent diversification within the genus Pteronotus bats in Mexico has given rise to P. mesoamericanus, which diverged from P. mexicanus around 0.7 million years ago. Despite being recognized as a distinct species, the evolutionary history of P. mesoamericanus in Mexico remains unclear. Our objective was to employ a comprehensive approach, integrating various methods to evaluate the evolutionary history and genetic patterns influenced by environment. To accomplish this, we utilized data from 67 sequences (21 localities) of the hypervariable domain HVII of the mitochondrial control region, and SNPs from 89 individuals (18 localities) exclusively from Mexico. Our analysis encompassed phylogeographic structure, historical demography, as well as ecological niche modeling and a phyloclimatespace approach. Our findings reveal distinct genetic groups: Gulf of Mexico and the Southeast of the country. This genetic structuring is most pronounced in the Southeast but at the same time with the highest genetic diversity. Notably, in Yucatán Peninsula, we observed a greater number of lineages with unique haplotypes and alleles. Furthermore, our analysis did not indicate isolation by distance, suggesting that the differentiation of these groups is primarily driven by environmental factors, particularly temperature fluctuations from the past. Understanding the presence of distinct lineages, each geographically and environmentally isolated from one another, provides us with insights into the potential fate of these lineages under climate change.
Reviving from four birds: post-bottleneck overview of temporal genomics of Mauritius kestrel

Xuejing Wang

Xuejing Wang, Jim Groombridge, Cock van Oosterhout, Ken Norris, Simon Tollington, Vikash Tatayah, Hernán Eduardo Morales

Centre for Agri-Environmental Research, Chester Zoo (United Kingdom), Division of Human and Social Sciences, Durrell Institute of Conservation and Ecology, GLOBE Institute, Mauritian Wildlife Foundation (Mauritius), North of England Zoological Society, Policy and Development, School of Agriculture, School of Anthropology and Conservation, School of Environmental Sciences, University of Copenhagen (Denmark), University of East Anglia (United Kingdom), University of Kent (United Kingdom), University of Reading (United Kingdom)

In many species, population declines can damage gene pools through the loss of genetic diversity and increased genetic load. This process, known as genomic erosion, increases the extinction risk, even after populations recover from a bottleneck. The Mauritius kestrel is one of the most extreme examples of population decline, having experienced a population bottleneck of four individuals in 1974. Through intensive conservation management, including captive-breeding programs and artificial nest boxes, the population rebounded to 350-400 individuals in the early 2000s and declined to <250 individuals. Here we analyze a whole-genome temporal dataset of 59 Mauritius kestrels from 1986 to 2021, supplemented with genomes from six outbreed and inbreed sister species. Demographic inference indicated that the kestrel population was small even before human arrival to Mauritius. Despite marked differences in population viability, we detected no genetic divergence between the subpopulations on the eastern and western parts of Mauritius. Their remaining genetic diversity is extremely low, with two mitochondrial genotypes found in early samples (1986-1995) and one genotype remaining in recent individuals (2012-2021). Genome-wide heterozygosity has remained stable, though at a minimal level – 5% of that found in widely distributed kestrel species. Runs of homozygosity (ROH, >1 Mb) constitute over 40% of the genomes, and the fraction of long ROHs (>5Mb) increased in recent times, indicating ongoing inbreeding within the recovered population. Our findings underscore the dramatic impact of the bottleneck on deleterious genetic variation, highlighting the population’s heightened vulnerability to future challenges.
S22 - Impact of environmental changes on agrobiodiversity and strategies for resilience.
It's increasingly clear that microbes influence host health, fitness, and even important developmental transitions like metamorphosis in mosquitoes and flowering time. The timing of these transitions is central for organisms responses in changing environments. However, we don't know which microbes alter such traits and how they do it, whether it's adaptive for the microbes or a consequence of their metabolism. Understanding the role of microbial communities in developmental timing of other organisms can help us better predict, and even enhance adaptive responses to changing signals. Using life-history models and community selection, our goal is to understand microbial effects on plant developmental timing. One intriguing hypothesis is that since microbes are indicators of environmental and ecological conditions, they can provide information about optimal phenology to the plant hosts. We are using Arabidopsis thaliana to determine how microbial communities can affect developmental timing. We sampled microbial communities from several environments and inoculated A. thaliana seeds to see the effect of those communities on development. Using directed evolution, we serially propagated microbial communities in planta and selected only those communities that promoted extreme phenotypes on developmental timing. After four transfers, we stabilized the community to obtain microbes that reliably accelerate, delay, or maintain flowering time with respect to uninoculated controls. Using omics approaches, we are characterizing the community and the metabolic functions that are associated with either late or early flowering and incorporating our insights into predictive life-history models to evaluate possible effects of microbes on plant adaptation to climate change.
The improvement of shattering traits in key climate-resilient under-utilised crops using Crispr-Cap9 technology.

Amy C. Jackson

Amy C. Jackson, George Burton, Emma Wallington, Caspar Chater, Rafal Gutaker
NIAB (United Kingdom), RBG Kew (United Kingdom), University of Sheffield (United Kingdom)

There are many examples of underutilized crops that display drought-resistant traits which could play a crucial role in improving food security, especially in regions with challenging agricultural environments. However, they often display a lack of domestication characteristics, such as non-shattering, which hinders their potential as a global sustainable food item for this changing world. With increasing efforts to publish genomic resources for under-utilised crops, genetic improvement of agronomic traits is increasingly possible. Here we focused on white fonio (Digitaria exilis), a small grain cereal crop native to West Africa and rice bean (Vigna umbellata), a legume cultivated across Asia. Through orthologous searches, we have identified homologous seed-shattering genes, SEED-SHATTERING1 (Sh1) and QUANTITATIVE SHATTERING1 (qSh1) in fonio and MYB26 gene in rice bean. Phylogenies and multiple sequence alignment were performed for each gene between closely related species to identify mutations, conserved regions, and gene structure. We aim to develop a reduced-shattering phenotype in both species using Crispr-Cas9-mediated genome editing technology. This poster highlights the utility of homolog search for accelerated domestication of crops with underutilized potential.
Strategies for mining useful alleles for climate change adaptation

Ayelet Salman-Minkov

Ayelet Salman-Minkov, Daniel Runcie, Jeffery Ross-Ibarra

UC Davis (USA)

Climate change and global warming pose significant challenges to future food security, as elite lines of major crop species are susceptible to heat and drought. A potential solution may lie within traditional varieties. Cultivated by local farmers worldwide for hundreds to thousands of years, traditional varieties possess adaptations to diverse climatic conditions. We would like to leverage these adaptations, identify alleles beneficial in future environments, and integrate them into the elite lines of crop species. Our objective is to determine the most effective strategy for identifying and incorporating valuable alleles into elite lines, utilizing both geo-referenced and genomic data. To represent diverse, realistic genetic architectures, we utilize simulated genomic data representing populations evolving across various demographies and landscapes. We evaluate three breeding approaches: 1) genotype-environment association (GEA) followed by marker-assisted selection (MAS), 2) genomic prediction of local adaptation combined with QTL mapping in a simulated future environment followed by MAS, and 3) genomic prediction of local adaptation combined with genomic selection in a simulated future environment. We anticipated that the GEA approach would be effective when there are common alleles with large-effect sizes, but the QTL mapping approach might be more effective when the population has many large-effect alleles, but all are rare. We found that the GEA approach was rarely better than the alternatives, while the QTL approach was best in nearly 50% of scenarios, regardless of genetic architecture. This suggests that the QTL approach may be the preferred choice for mining beneficial alleles.
Domestication is a complex, non-linear process in which humans and natural evolutionary processes, including genetic drift and natural selection, shape genetic diversity. Wild and feral populations, landraces, breeding lines, and even genetically modified lines coexist in the centers of domestication, influencing each other. Growing evidence shows that gene flow from wild relatives has contributed to the genetic diversity, including adaptive variation, of domesticated forms. To face the current and future challenges associated with climate change, new and resistant pathogens, and reduced crop diversity, we need to further explore the genetic diversity present in crop gene pools, including wild relatives and landraces adapted to diverse biotic and abiotic conditions. I present evidence of highly diverse traditional cultivars of the scarlet runner bean (Phaseolus coccineus), an orphan crop from Mesoamerica. We found evidence of a severe genetic bottleneck associated with the onset of domestication, followed by demographic expansion and asymmetric gene flow from the wild to the cultivated populations that contributed to the recovery of genetic diversity. In lentils, a cool-season grain legume, landraces from different regions of the Middle East and South Asia are highly differentiated, and the genetic groups of the wild lentil (Lens orientalis) show no geographic pattern, but climatic variables explain their distribution, suggesting local adaptation to specific conditions. Fostering wild relatives and landraces preserves useful genetic diversity and agrobiodiversity. Despite this, there are significant gaps in gene bank collections, and many orphan crops and their wild relatives remain unexplored.
Climate change is a threat to both biodiversity and crop sustainability. One path to maintain crop productivity is via introgression with crop wild relatives (CWRs). Historically, CWRs have contributed alleles that improve resistance to pathogens, confer abiotic tolerance to stressors such as salinity and temperature, and bolster yields and flavor profiles. But what is the conservation outlook for CWRs in the face of climate change, and how can we utilize them more effectively in the face of an urgent climate crisis? In this talk, I will summarize our recent work that focuses on using, as a model, the domesticated grapevine (Vitis vinifera) and its ecological diverse CWRs from North America. Our work merges genomic information on landscape scales with population genetic inference and climate models. With these approaches, we are: i) attempting to predict CWR germplasm that will be agronomically useful for viticulture in future climates, ii) studying key parameters like genetic load, which affects conservation outlook, and iii) producing insights into related phenomena, such as the predicted geographic distribution of a devastating bacterial pathogen.
Traditional crops as reservoirs of genetic diversity: the case of pumpkin Cucurbita pepo ssp. pepo in Mexico

Carmina NA Martínez-González

Carmina Martínez-González, Guillermo Sánchez-de la Vega, Josué Barrera-Redondo, Emanuel Villafán de la Torre, Cuauhtémoc Jacques Hernández, Erika Aguirre-Planter, Rafael Lira-Saade, Luis E. Eguiarte

Centro de Biotecnología Genómica (Mexico), Departamento de Ecología Evolutiva, Facultad de Estudios Superiores Iztacala, Instituto de Ecología, Instituto de Ecología A.C. (INECOL), Max Planck Institute for Biology (Germany), Red de Estudios Moleculares Avanzados (REMAv) (Mexico), UBIPRO, Universidad Nacional Autónoma de México (Mexico)

Landrace crops (razas or variedades criollas in Spanish) are considered the second most important reservoir of genetic diversity for crop improvement, only after the wild crop relatives. Landrace crops are the product of centuries of selection by farmers and have locally evolved in response to both the environmental conditions of each region and different social and cultural aspects. In Mexico, the traditional landraces of pumpkins and squashes of C. pepo ssp. pepo were domesticated, cultivated and managed under the traditional cultivation system – the milpa – starting at least 10,000 years ago, which has resulted in the development of high morphological and genetic diversity. We analyzed 8,643,037 SNPs obtained by whole genome sequencing at high coverage from 96 individuals of different landraces of C. pepo ssp. pepo representing all Mexico and Guatemala. We also included 13 individuals of the ancestor of this cultivated species, C. pepo ssp. fraterna, from the only three wild populations that we found. The traditional landrace varieties of Mexico exhibit a genetically differentiated group from their wild populations and have higher levels of genetic diversity (Cultivated ?= 0.155; Wild ?= 0.058), possibly due to the very small population sizes of the extant wild populations. We found a clear geographic differentiation among the landraces, forming four groups. We also detected 28 genes under selection. We conclude that in contrast to what is generally reported in the literature, the traditional landraces of C. pepo ssp. pepo are the most important reservoir of genetic diversity for this species.
Disentangling the Evolution of Grape Crop Wild Relatives
Christopher J Fiscus

Christopher J Fiscus, Jonás A Aguirre-Liguori, Jacob B Landis, Garren RJ Gaut, Abraham Morales-Cruz, Brandon S Gaut
Cornell University (USA), Irvine (USA), Lawrence Berkeley National Laboratory (USA), Universidad Autónoma de Yucatán (Mexico), University of California

The Vitis genus comprises ~80 species distributed throughout North America and Eastern Asia that thrive in a variety of environments from lush forests to dry deserts. The rapid adaptive radiation of this genus makes it an ideal system to examine the genetic and environmental factors contributing to species divergence. Understanding adaptation in wild Vitis also benefits the breeding and improvement of the crop species V. vinifera, as wild species possess alleles pre-adapted to abiotic and biotic stress that can be utilized agronomically. To study the evolution of Vitis, we cataloged genetic variation in over 550 individuals from 56 species, with special emphasis on species endemic to North America. We first used this dataset to clarify species relationships, which have been obscured due to rampant introgression and hybridization between species. We detected several clades of hybrid origin, suggesting complex speciation histories. Next, we identified genotype-environment associations (GEA), identifying alleles associated with either warm/wet or cool/dry climates, and we examined the evolutionary history of these alleles through recent evolutionary time. Finally, we leveraged our dense sampling of V. arizonica, a species endemic to the American Southwest, to model the relationship between climatic and genetic variation using novel machine learning algorithms that were modified to account for collinear predictors, which otherwise distort predictor importance. Our work contributes to ongoing efforts to decipher the genetic basis of environmental adaptation and supports initiatives aimed at bolstering the sustainability of essential food crops through harnessing the potential of crop wild relatives.
The utility of landrace environmental data for climate-adaptive maize breeding

Forrest Li

Forrest Li, Daniel J Gates, Edward S Buckler, Sarah J Hearne, Jeffrey Ross-Ibarra, Daniel E Runcie
Center for Population Biology, Cornell University (USA), Department of Evolution and Ecology, Department of Evolution and Ecology and Department of Plant Sciences, Department of Plant Sciences, Genome Center, International Maize and Wheat Improvement Center (CIMMYT) (Mexico), Ithaca, NY and Institute for Genomic Diversity, Robert W. Holley Center for Agriculture and Health, USDA-ARS, University of California Davis (USA), and University of California Davis (USA)

Maintaining crop production yields in the face of climate change is a major challenge facing plant breeding today. Considerable adaptive genetic variation to improve breeding programs exists in ex situ germplasm collections of traditional landraces, but effective utilization of this diversity requires identifying adaptive alleles or testing the agronomic performance of varieties in different environments. To evaluate approaches for identifying environmentally adaptive alleles, we collected climatic variables from the environment-of-origin of each of nearly 4000 previously genotyped traditional maize landraces representing a breadth of maize diversity. We evaluated three different approaches of using genetic and climate data to identify useful diversity based on phenotypic predictive ability in field trials. First, we performed gene-environment association (GEA) studies identifying genetic loci associated with environmental heterogeneity, such as hsftf9, a putative heat shock protein that may provide adaptive benefit in high temperatures, as well as the locally adaptive large scale inversion Inv4m, previously characterized as locally adaptive to highland conditions. Second, we used existing genotype data to predict optimal climate niches that may inform selection of landraces better adapted to target environments as well as demonstrate the limitations of prediction of environmental variables across spatial ranges. Third, we tested the ability of genomic prediction using population structure and high-resolution SNP data for landrace collections. We find that genomic prediction capturing population structure and genetic relationship continues to have the highest predictive ability compared to models employing climate-of-origin variables or models focused on large-effect loci associated with climate.
The domestication process of two Agave species used for mezcal production with different propagation methods and evolutionary histories.

Irene Martínez Velasco

Irene Martínez Velasco, Jaime Gasca Pineda, Anastasia Klimova, Karen Ruiz Mondragón, Cécile Truchot Taillefer, Guillermo Sánchez de la Vega, Erika Aguirre Planter, Rafael Lira, Alfonso Valiente Banuet, Luis E. Eguiarte Fruns

Facultad de Estudios Superiores Iztacala, Instituto de Ecología, Laboratorio de Recursos Naturales, Unidad de Biotecnología y Prototipos (UBIPRO), Universidad Nacional Autónoma de México (Mexico)

In Mesoamerica, in situ management of wild plants is common and plants with different use to humans can exhibit varying degrees of domestication and be subject to diverse practices. In economically important species, such as agaves, direct extraction of wild plants along with different management practices alters the genetic composition of the populations. Currently, many Agave species are intensively propagated for mezcal production. Two of them, Agave karwinskii in the states of Puebla and Oaxaca, and Agave cupreata from Guerrero and Michoacán are facing strong extraction in their wild populations and artificial selection processes. The prevailing propagation method of many Agave species— including A. karwinskii— involves underground rhizomes, facilitating the establishment of commercial plantations. However, species like A. cupreata do not produce basal shoots or bulbils, depending entirely on seed production, making plant nurseries indispensable for mass plant production. We analyzed populations of both species under different degrees of management, from wild to commercially exploit. We studied 300 individuals of A. karwinskii, and 153 of A. cupreata using genomic data from GBS, with 12,888 and 15,880 SNPs, respectively, to obtain diversity statistics and describe the genetic differentiation patterns. We found a high expected heterozygosity in A. karwinskii (0.299) and A. cupreata (0.236), as well as a significant decrease in genetic diversity associated with the degree of agricultural management in both species. Our results indicate that, in addition to the propagation method, the type and direction of anthropogenic pressure are significant modifiers of their genomic biodiversity populations.
Accelerating climate change underscores the urgency of comprehending evolutionary responses to heat stress. In this context, one of the most significant challenges facing organisms today is global warming. There is an urgent need to gain a better understanding of the long-term evolutionary effects of rising temperatures on biodiversity. The primary objective was to assess the capacity of populations to adapt to heat stress. For this purpose, we conducted a long-term evolution experiment using the powerful microbial model system, budding yeast (Saccharomyces spp.). We selected eight species with distinct temperature preferences, ranging from 15 to 35°C, isolated from diverse ecological and geographic locations around the world. Multiple populations of each species evolved for 600 generations under increasing temperature conditions, ranging from 25 to 40 °C. We observed that not all species have the ability to adapt to high temperatures, while cold-tolerant species did not exceed 34°C. In contrast, heat-tolerant species surpassed 36°C. However, growth rate measurements at the final evolved temperature revealed that cold-tolerant species exhibited the greatest increase in fitness compared to their non-evolved strains. Finally, we performed comparative genomics analyses to dissect the genetic architecture of thermal adaptation. Whole-genome sequencing demonstrated signatures of selection that explained this greater fitness. The large genomic diversity and evolutionary time frames covered by this work provide the unique opportunity to study responses to future climate change, in real-time, with high replication, and across multiple species.
Genetic diversity of *Zea diploperennis* and strategies for its conservation

José Carlos Moreno Juárez, Alicia Mastretta-Yanes, Ana Laura Wegier, Diana María Rivera, José de Jesús Sánchez González

Campus Tlajomulco (Mexico), Centro Universitario de Ciencias Biológicas y Agropecuarias, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Mexico), Departamento de Ciencias Básicas, Instituto de Biología, Jardín Botánico, Tecnológico Nacional de México, UNAM (Mexico), Universidad de Guadalajara (Mexico)

*Zea diploperennis*, a perennial teosinte species and wild relative of maize, exhibits a distinctive range and is characterized by unique agricultural traits. These traits hold substantial potential for future maize breeding programs. Additionally, it holds cultural significance for traditional farmers in the Sierra de Manantlán region. However, *Z. diploperennis* populations face threats from land-use changes and the decline of traditional agriculture, closely linked to its existence. Consequently, a comprehensive assessment of the conservation status of its genetic diversity is imperative. This evaluation, aligned with the Kunming-Montreal Global Biodiversity Framework ratified in 2022, utilized Genotyping-by-Sequencing (GBS) data from all known *Z. diploperennis* localities. Genetic clustering analysis revealed two distinct clusters corresponding to Jalisco and Nayarit localities, indicating significant genetic structure. Current and historical effective population sizes were estimated, revealing both clusters with effective population sizes below 500 and a recent decline in effective sizes. This suggests a present low genetic diversity, further diminished in recent generations. Urgent in situ conservation strategies are warranted for both clusters to mitigate potential inbreeding depression.
Population genomics of tequila and mezcal (Agave spp.): Unexpected findings and lessons for these Mexican industries in the face of climate change.

Luis E. Eguiarte

Universidad Nacional Autónoma de México (Mexico)

Agave is a diverse genus, with more than 280 species and reports of use by humans for at least 140 species, usually as fibers, as food, and in particular for making alcoholic drinks as pulque (just fermented), mezcal or tequila (both distilled). Recently, there has been an impressive boom in the interest and production of both mezcales and tequila. Here we present the results of our populations genomics studies with tequila agave, A. tequilana in the state of Jalisco, espadín agave of Oaxaca (A. angustifolia), bacanora from Sonora (also A. angustifolia), tobalá agave (A. potatorum) from Oaxaca and Puebla, tobasiche, cirial, cuije, etc. agave, all different names for A. karwiskii, from Oaxaca and Puebla, agave tepextate or pizometl (A. marmorata) also from Oaxaca and Puebla, and agave papalomé or ancho (A. cupreata) from Guerrero and Michoacán. We found very high levels of genetic variation and low geographic differentiation, and in some cases more inbreeding than expected, but diversity decreases with the level of management (the more intensively managed populations usually have less overall variation). In tequila and espadín, we detected that the plants are clonally propagated, but in these selected clones each plant is highly heterozygous. We are using these genomic data and detailed studies on the potential distribution of the species in the future along their genetic variation (i.e., genomic offset analyses) to propose areas where plants could be grown and cultivated in the future, which we believe will be important results for the producers of tequila and mezcal.
The distribution of self-incompatibility systems in cultivated angiosperms, facing global changes

Miriam Monserrat Ferrer, Mirley Arlyn Verde-Cáceres, Sara Victoria Good
The University of Winnipeg (Canada), Universidad Autónoma de Yucatán (Mexico), University of Manitoba (Canada)

Theoretical and experimental studies have highlighted that angiosperms harbouring self-incompatibility (SI) systems are more susceptible to inbreeding depression when a transient breakdown in the SI system is achieved either due to selective forces (e.g. mate limitation) or environmental effects (e.g increased temperature). Domesticated plants are usually thought to be self-compatible, although, in a recent survey of 5686 hermaphroditic angiosperms, we found that 443 out of 907 cultivated species displayed self-incompatibility, partial self-compatibility, or self-sterility. Here, we summarize existing experimental work examining the effect of increased temperature on the breakdown of SI, as well as review the molecular and genetic basis of SI systems in cultivated species. Next, we present an analysis of the association of cultivated species with i) phylogenetic distribution, ii) presence of SI, self-compatibility (SC), partial self-compatibility (PSC) or self-sterility (SS); iii) growth form, iv) propagation method and v) use of the crop. We discuss how clonal propagation of individuals enables the maintenance of self-incompatible and partially self-compatible cultivars even for fruit production, but these maintaining fertility and SI lines present unique challenges not present for self-compatible cultivars. Finally, we contextualize this work within the ongoing dialogue about the particular vulnerabilities of self-incompatible cultivars in the context of global change, particularly the impact that increasing temperature.
A boaring story of hybridization and adaptation in wild boars, domestic pigs and feral swine
Mirte Bosse

Mirte Bosse, Niek Barmentlo, Timothy J Smyser
National Wildlife Research Center (USA), VU University (Netherlands), Wageningen University (Netherlands)

The ecological threat posed by invasive species is well established, however evolutionary processes that contribute to the invasion potential of a given species are not as well understood. Wild pigs in the US serve as an ideal model to evaluate genomic and evolutionary processes associated with invasiveness given the unique molecular resources available for Sus scrofa, profound ecological and economic destruction caused by the species, and the global extent of their invasive spread. Within the United States, wild pigs overwhelmingly descend from the historical hybridization of wild boar and domestic pigs and have expanded rapidly over the past 40 years to invade as many as 38 states. We sought to evaluate the importance of genetic enrichment through hybridization to the perceived heightened invasiveness of the species, using thousands of genotyped feral swine. The vast majority of feral swine were of mixed ancestry, with dominant genetic associations to Western heritage breeds of domestic pig and European populations of wild boar. Mixed ancestry from wild and domestic lineages clearly contributes to the rapid expansion of invasive feral swine populations. We discovered several genomic regions under selections that encode for phenotypic traits likely to impact fitness in the wild such as coat coloration and craniofacial development. These results elucidate genetic processes that facilitate establishment and subsequent evolutionary responses that contribute to invasive spread.
Domestication gradient in Capsicum annuum chile pepper’s center of origin: implications under climate change scenarios drawn from niche models and genomic diversity.

Natalia Elena Martínez-Ainsworth


Iowa State University (USA), National Autonomous University of Mexico (UNAM) (Mexico), Ohio State University (USA), University of Hawai'i (USA)

Among crops domesticated in Mexico, chile pepper (Capsicum annuum) offers an ideal model to analyze the intersection of domestication status and management regimes over diverse environmental conditions and how these categories might face the challenges imposed by anthropogenic climate change. Here we generate a large C. annuum collection (n=1226) spanning four domestication categories throughout Mexico. In addition, we recovered coordinates from bibliography, public repositories and developed proxies to perform detailed ecological niche models for each category under present day conditions as well as eight future climate change scenarios, three time points and two emissions forecasts. Our main findings include an expansion to higher altitudes and retraction from dry-hot areas of the early domesticate landraces followed by a regaining of these lost areas by commercial varieties purportedly through irrigation. Environmental envelopes differed across domestication categories, notably for temperature seasonality and precipitation of the warmest quarter. There were several key geographic areas suitability was lost in both wild and landrace under future niche suitability projections this is a potential threat to geneflow between these groups which has helped maintain genetic diversity. We further obtained genome-wide genotyping (GBS) data on a sample of 61 wild, 45 semiwild, 164 landraces and 5 commercial samples. By comparing genomic diversity estimates and exploring environmental structuring of such diversity we have found that backyard and milpa management practices are a key potential source of climate adaptive traits and thus should be targeted by conservation efforts.
Genomic landscape of introgression in Indonesian cattle and its implication for their adaptive potential
Sabhrina Gita Aninta

Xi Wang, Casia Nursyifa, Laura Bertola, Genis Garcia-Erill, Sabhrina Gita Aninta, Anubhab Khan, Josiah Kuja, Kristian Hanghøj, Jonas Meisner, Thomas Bøggild, Corey Bradshaw, Alam Persada Putra, Sendi Dwi Priyono, Yuli A. Tribudi, Jiang Yu, Johannes Lenstra, Reagan Sims, Darren E. Hagen, Michael P. Heaton, Timothy P. L. Smith, Mikkel-Holger S. Sinding, Laurent Frantz, Greger Larson, Dedy Duryadi, Muhammad Agil, Bambang Purwantara, Rasmus Heller (Australia), Animal & Food Sciences, College of Animal Science and Technology, College of Science and Engineering, Department Population Health Sciences, Department of Agricultural Sciences, Department of Biology, Faculty of Agriculture, Faculty of Biology, Faculty of Veterinary Medicine, Fakultas Kedokteran Hewan, Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor (Indonesia), Institut Pertanian Bogor (Indonesia), Laboratorium Genetika Konservasi Hewan, Northwest A & F University (China), Oklahoma State University (USA), School of Archaeology, Texas State University (USA), U.S. Department of Agriculture (USA), Universitas Gadjah Mada (Indonesia), Universitas Tanjungpura (Indonesia), University of Copenhagen (Denmark), University of Munich (Germany), University of Oxford (United Kingdom), Utrecht University (Netherlands)

Admixture between species is increasingly recognised to be prevalent in many species, especially within domesticated cattle breeds as local people aim to improve their genetic variation. In Indonesia, cattle breeds are generally known to be in high genetic diversity due to having an admixed ancestry from zebu (Bos indicus) and banteng (Bos javanicus). We performed whole genome sequencing of 233 bovids, representing local cattle from Aceh, Pesisir, Pasundan, Jabres, Madura, Sumba Ongole, Javan banteng, as well as Bali cattle in Indonesia and Australia. After comparing with other cattle where published genomic data is available, Indonesian cattle breeds are found to have the highest genetic diversity of any cattle populations worldwide. This is largely due to amounts of banteng introgression, reaching up to ~30% in the Madura breed. We also found regions with pronounced peaks and valleys of introgression, suggesting selection has been acting on the introgressed haplotypes. While the genomic landscape of introgressed ancestry was significantly correlated between breeds, extreme outliers were predominantly private to each breed, consistent with breed-specific artificial selection pressures. However, we found evidence that some phenotypes, such as coat colour, might experience convergent adaptive introgression in all breeds with banteng introgression. These findings unveil much of the unknown part of the evolutionary history of Asian cattle and processes occurring after introgression into a livestock species, in addition to discovering a rich source of previously unknown genetic variation in cattle that is relevant for genetic improvement.
S23 - Exploring the Frontiers of Single-Cell Biology in Diverse Organisms.
LatinCells: empowering research capacity in Latin America through community engagement, infrastructure development, and single-cell RNA profiling of diverse tissues for the Human Cell Atlas.
Adolfo Rojas

Domenica Marchese, Adolfo Rojas-Hidalgo, Ricardo Verduco, Alanis Huañaco, Alicia Colombo, Gerardo Donoso, Katherine Marcelain, Marcelo Fonseca, Diego Pérez-Stuardo, Emiliano Vicencio, Sebastián Leiva-Navarrete, Guillermo Barreto, Carla Gallo, Patricia Severino, Patricia Possik, Hugo Guerrero-Cázares, Lucía Spanenberg, Daniela Robles-Espinoza, José Antonio Corona-Gómez, Carlos Ortiz-Ramírez, Andrés Moreno-Estrada, Vinicius Maracaja-Coutinho

Advanced Genomics Unit (UGA-LANGEBIO) (Mexico), Brazilian National Cancer Institute (INCA) (Brazil), Hospital Israelita Albert Einstein (Brazil), Institut Pasteur de Montevideo (Uruguay), Mayo Clinic (USA), Universidad Arturo Prat (Chile), Universidad Nacional Autónoma de Mexico (Mexico), Universidad Peruana Cayetano Heredia (Peru), Universidad de Chile (Chile), Universidad de Talca (Chile), Universidad del Valle (Colombia)

Latin America boasts unparalleled ethnic and genetic diversity, a result of centuries of admixture between indigenous populations and other continental ancestries. However, this richness remains largely unrepresented in the Human Cell Atlas, impeding the development of tailored medical interventions and perpetuating healthcare disparities. The LatinCells initiative seeks to rectify this by constructing comprehensive cellular maps with samples from diverse indigenous and admixed populations across Brazil, Chile, Colombia, Mexico, Peru, Uruguay, and among U.S. Latinos. We have implemented a multifaceted approach, which includes actively engaging with indigenous communities in remote and rural locations, deploying advanced technological solutions for scRNAseq data generation from immune cells and gallbladder samples, and developing bespoke computational analysis pipelines. We are also building local capacity through the creation of training materials and workshops focused on single-cell RNA sequencing analysis, tailored to the needs of researchers in Latin American countries. Our preliminary results are focused on pilot samples for which we have single-cell expression atlases of immune cells derived from various Latin American ancestries. Through LatinCells, we aim not only to bridge the gap in representation within the Human Cell Atlas but also to foster a collaborative network that empowers local communities and researchers to strengthen research capacity across Latin America.Keywords: Single-Cell; Immune System; Transcriptomics; Bioinformatics Education. Funding: This project is funded by the Chan Zuckerberg Initiative (CZI).
Evaluation of a deeply conserved vertebrate signal of neuronal activation in plethodontid salamanders
Apoorva Ganesh
Apoorva Ganesh, Stella M. Lai, Damien B. Wilburn
Ohio State University (USA)

Chemical signals for intraspecies communication are important to the reproductive strategies of many vertebrate taxa and likely evolve by sexual selection. Vertebrate olfactory systems possess many sensory neurons tuned to detect species-specific pheromones emitted by conspecifics, but biochemically identifying pheromone-receptors pairs is challenging. Consequently, little is known about the molecular co-evolution of pheromones and their receptors. Plethodontid salamanders are a valuable model system where multiple families of male protein pheromones can be recombinantly synthesized and alter female mating behavior via stimulation of olfactory neurons. The physiological response of neuronal activation in salamanders is still under research. Further, verified methods to measure this activation are limited. The current best approach requires harsh fixation methods that prohibit further isolation of genetic material from activated neurons to identify putative pheromone receptors. Here we evaluate the phosphorylation of Ribosomal Protein S6 (RPS6) as a potential probe for neuronal activation in plethodontid salamanders. C-terminal phosphorylation of RPS6 was the first discovered protein phosphorylation and has historically been associated with neuronal activation in other vertebrate taxa. Phylogenetic comparisons of plethodontid RPS6 support deep evolutionary conservation of the C-terminus including key serine residues. Compared to established methods, biochemical and immunohistochemical experiments with commercial anti-pRPS6 show similar levels of neuronal activity in response to pheromone treatment. Immunohistochemical labeling with anti-pRPS6 requires relatively gentle formaldehyde fixation of olfactory tissue that will enable future studies to specifically isolate pheromone-activated neurons for identification of receptor genes.
Cells are the fundamental units of life, however we know relatively little about their evolution and the underlying molecular mechanisms. We aim to explore three fundamental questions regarding the evolution of new cell types and functions: (i) how transcriptional states can diverge between genetically identical cells, allowing for specialization and heterogeneity? (ii) how genetic variation at the population level impacts such transcriptional divergence? and (iii) can co-option of transcriptional networks trigger rapid evolution of new cellular functions? To answer these questions, we are comprehensively profiling cell types and cell states in multiple models. For example, using unicellular yeast we provide evidence that transcriptional heterogeneity emerges spontaneously in uniform cell cultures, contributing to stress adaptation. Additionally, by combining single cell seq with human population genomics we are cataloging genetic variants associated with cell type-specific expression signatures across Latin America. This will inform to what degree genetic diversity drives functional adaptation at the cell type and population level, and which cell types contribute more to such variation. Finally, using a plant model we describe the co-option of transcription factor activity leading to the evolution of new photosynthetic functions on a vascular cell, resulting in improved growth and productivity.
How one becomes another: exploring the role of exaptation on the evolution of gene regulation

Cauã Antunes Westmann

Cauã Antunes Westmann, Leander Goldbach, Andreas Wagner
Department of Evolutionary Biology and Environmental Studies, Swiss Institute of Bioinformatics (Switzerland), The Santa Fe Institute (USA), University of Zurich (Switzerland)

Exaptation, the co-option of existing traits for new functions, is a central process in Darwinian evolution. However, the molecular changes leading to specific exaptations remain unclear. Here, we investigated the potential of bacterial transcription factor binding sites (TFBSs) to evolve exaptively, examining three global transcription factors (TFs) from Escherichia coli: CRP, Fis, and IHF. Using a massively parallel single-cell reporter assay, we mapped three combinatorially complete adaptive landscapes, encompassing all intermediate sequences between three pairs of strong binding sites for each TF. Our results revealed that these landscapes are smooth and navigable, with a monotonic relationship between mutations and their impact on gene regulation. Starting from a strong TFBS for one of our TFs, Darwinian evolution can thus create a strong binding site for another TF through a small number of individually adaptive mutations. Notably, most intermediate genotypes are prone to transcriptional crosstalk – gene regulation mediated by both TFs. Because our landscapes are smooth, Darwinian evolution can also easily create binding sites that show such crosstalk whenever it is adaptive. Our study presents the first in vivo evidence that new TFBSs can evolve exaptively through multiple small and adaptive mutational steps. They also highlight the importance of regulatory crosstalk for the diversification of bacterial gene regulation.
At NGI, we provide sequencing services, library preparation, and bioinformatic support to researchers in Sweden and abroad. One of our areas of focus is transcriptomics, including single-cell and spatial methods. These methods primarily target human or mouse samples, but we have successfully processed samples from other species including salamander, cockroaches, and sponges. Single-cell RNAseq methods allow the characterisation of gene expression of hundreds/thousands of cells per sample, while retaining cellular resolution. Starting from a cell suspension, each cell is compartmentalised into a droplet via microfluidics (Chromium 10X) or FACS-sorting in a 384-well plate (Smart-seq3) and the transcriptome is barcoded individually. These methods can be used on non-human/mouse samples if quality requirements are met. Spatial transcriptomic methods add a spatial dimension to a tissue’s sequencing data. The 10x Visium assay for fresh-frozen tissue samples relies on capturing the polyadenylated RNA from a tissue section which has been placed on a slide coated with barcoded probes. The library obtained from this sample can be mapped spatially and combined with a histological image to gain spatial resolution of the tissue. Among the Visium assays available, only this one can be used for non-human/mouse samples. The quality of the sample is vital for obtaining high-quality data. When sequencing non-standard samples, it is essential to understand the challenges that each specific method may pose, so discussing possible limitations with us in the beginning of a project can be of great benefit.
Cross-species comparison of gene regulatory networks using single cell sequencing

Ines Hellmann

Anita Térmeg, Fiona Edenhofer, Johanna Geuder, Wolfgang Enard, Ines Hellmann
Ludwig-Maximilians University (Germany)

Many phenotypic differences between primate species must be established during development. Hence, in order to understand them we need to investigate the underlying gene regulatory networks (GRNs) and their evolution. Single-cell RNA-seq methods finally allow us to reliably infer GRNs. However, most comparative analyses of GRNs have so far focused on qualitative aspects - a quantitative comparison as required for evolutionary analysis remains challenging. We designed a computational pipeline to perform quantitative cross-species network comparisons to identify conserved and diverged gene modules and applied it to scRNA-seq data of an early neuronal differentiation experiment, with iPS cells from human and cynomolgus macaque. Surprisingly, we found the network of the otherwise conserved pluripotency regulator OCT4 among the most diverged networks. Closer inspection of the genes that show the highest degree of differential regulation included most prominently two genes under the control of LTR7 elements. A completely different approach is to infer GRNs from perturbation screens. To this end, we established inducible KRAB-dCas9 iPSC lines for humans and macaques. In our first CRISPRi screen, we target 27 transcription factors for each of which we designed multiple guide RNAs, allowing for a quantitative comparison between the species. Using the log fold changes between the perturbed and unperturbed conditions as edge weights, we can then construct GRNs and apply our established statistics to find diverged and conserved co-expression modules. This analysis could e.g. confirm the most diverged genes of the OCT4 network identified in unperturbed cells.
The human immune system is under constant evolutionary pressure, primarily through its role as first line of defence against pathogens. Accordingly, population genomics studies have shown that immune-related genes have a high rate of adaptive evolution. These studies, however, are mainly based on protein-coding genes without cellular context, leaving the adaptive role of cell-types states uncharted. Inferring the rate of protein-coding genes adaptation in human immune cells at cellular resolution, we found cell-types from lymphoid and myeloid compartments to harbour significantly increased adaptation rates (eg. foetal Pre-Pro B cells and adult Trm CD8+). Focusing on the Lung, we found an enrichment of cell-types with high adaptation in barrier tissues, suggesting an adaptive response to external challenges such as respiratory pathogens. We further analysed iPSC-derived macrophages responding to various challenges, including pro- and anti-inflammatory cytokines or bacterial and viral infections, the latter simulating the evolutionary arms race between humans and pathogens. Here, we found adaptation in early immune responses, suggesting host benefits to adapt to early infection stages to control pathogen spread. Together, our study reveals spatio-temporal and functional biases in human immune populations with evidence of rapid adaptive evolution, providing a retrospect of forces that shaped the complexity, architecture, and function of the human body.
Improved characterization of single-cell RNA-seq libraries with paired-end avidity sequencing

John Chamberlin

John Chamberlin, Austin Gillen, Aaron Quinlan
University of Colorado Department of Medicine (USA), University of Utah Department of Biomedical Informatics (USA), University of Utah Department of Human Genetics (USA), University of Utah

Here, we present an unconventional paired-end alignment strategy enabled by the improved homopolymer handling of avidity sequencing (Element Biosciences). Sequencing past the poly(T) primer with the forward read allows direct inspection of the putative site at which the primer hybridized to the RNA molecule. We show that this increases confidence in annotation optimization strategies by ruling out internal priming or other artifacts as well as directly assigning the recovered reads to polyadenylation sites. We also share important adaptations to the alignment workflow needed for accurate paired end alignments.
Single-cell RNA sequencing for phylogenomic analyses of uncultivable microbial eukaryotes - A case study with marine planktonic protists (Protista, Ciliophora, Oligotrichia)

Luciana Santoferrara

Shahed U. A. Shazib, Ragib Ahsan, Marie Leleu, George McManus, Laura Katz, Luciana Santoferrara
Hofstra University (USA), Purdue University (USA), Smith College (USA), University of Connecticut (USA), University of Massachusetts Amherst (USA), na (USA)

Protists comprise most branches of the eukaryotic tree of life, still their astonishing diversity and billion-year-old evolutionary history remain poorly understood. Phylogenomics is a powerful tool to study evolution, but poses multiple challenges for protists (e.g., large and complex genomes, lack of dedicated databases and bioinformatic tools) in addition to the standard ones (e.g., collecting, storing, and analyzing large amounts of data). We have developed a phylogenomic workflow based on single-cell RNA sequencing data, integrating key steps from cell isolation to species tree inference. We assessed the effectiveness of our pipeline using publicly available and newly generated transcriptomes (11 and 28, respectively) from the Oligotrichia, a diverse group of ciliated protists. This group has been relatively well-studied based on ribosomal RNA gene (rDNA) markers, which we reconstructed by read mapping of transcriptome sequences and used both to supplement species identification and to compare rDNA phylogenies and phylogenomic inferences. We also compared phylogenomic analyses from well-curated orthologs (single-copy genes) and in the presence of paralogs (multi-copy genes), and evaluated the effect of missing data. For this, we analyzed multiple subsets of up to 1,014 conserved protein-coding genes by both concatenated gene trees and Asteroid species tree inferences. Both approaches yielded similar results, and most of the phylogenetic relationships among the Oligotrichia were consistent and strongly supported. Our analyses demonstrate that Asteroid provides robust phylogenetic support, while simplifying curation steps and maximizing the number of gene families represented in the inferences.
Evolutionary impact of transcriptomic differences in X and Y carrying gametes

Meritxell Riera

Meritxell Riera, Sofia Boeg Winge, Marjoline Rouselle, Florent Murat, Noe Mbengue, Samuele Soraggi, Christina Hvilsom, Henrik Kaessmann, Kristian Almstrup, Mikkel Heide Schierup (Germany), Aarhus University (Denmark), Bioinformatics Research Centre, Center for Molecular Biology (ZMBH), Copenhagen University Hospital (Denmark), Copenhagen Zoo, DK 2000 Frederiksberg (Denmark), DKFZ-ZMBH Alliance, Department of Growth and Reproduction, Heidelberg University, Heidelberg University (Germany), Rigshospitalet, University Hospital (Denmark)

Spermatogenesis is a highly complex developmental process by which germ cells divide and differentiate to generate haploid spermatozoa. Genes expressed in the testis show signals of rapid evolution in mammals and apes, most likely driven by differences in the reproductive success of spermatozoa. Recently, in a comparative study of single-nuclei RNA-seq (snRNA-seq) data in mammals we showed that this accelerated rate of molecular evolution is driven by genes expressed in late spermatogenesis, and in particular haploid spermatids. Haploid cells differ in whether they carry an X or a Y chromosome, which can lead to conflict over transmission of sex chromosomes to the next generation. In this study, we compare the transcriptomic profiles of haploid cells carrying an X or Y chromosome using snRNA-seq data, which we spatially validate using single-molecule RNA in situ hybridization (smiSH). We find both lineage specific differences and conserved expression patterns across primates which we correlate with signals of molecular evolution. Notably, we identify genes specifically expressed in X-carrying spermatids that show a human specific increase in post-meiotic gene expression and strong signals of natural selection. These genes are also found in Neanderthal deserts, suggesting that they could have been associated with reproductive barriers between humans and Neanderthals when they met and interbred around 60,000 years ago. Our study uncovers the post-meiotic expression of sex-linked genes across primates and provides valuable insights into possible cases of sexual antagonistic effects that shaped human evolution.
Disentangling the gene regulatory network controlling intrinsic antiviral immunity in stem cells

Mónica Padilla Gálvez

Mónica Padilla Gálvez, Qing Yang, Daniel Blanco-Melo

Fred Hutchinson Cancer Center (USA), Postbaccalaureate program, Vaccine and Infectious Disease Division

Stem cells defend from viral infections by intrinsically expressing antiviral genes independent of viral recognition or signal transduction. As these cells differentiate, these intrinsically expressed antiviral genes (IAGs) become interferon signaling dependent. Previous work indicates an incompatibility of pluripotency with interferon response, highlighting both the need for IAGs expression and that these processes might have evolved from similar pathways. However, two main questions remain: (1) how do stem cells regulate IAGs? (2) What regulatory switch during differentiation makes IAGs expression interferon signaling dependent? Here, we aim to elucidate the gene regulatory network controlling IAGs expression dynamics during early differentiation using bulk RNA-seq and multiome analysis. We first identified putative IAGs comparing gene expression between human embryonic stem cells (hESCs) at baseline and infected differentiated cells. We further classified them by expression levels and characterized their behavior in hESCs upon influenza A virus infection. We found 20 transcription factors (TFs) regulating IAGs, including key pluripotency factors, such as KLF4 – a Yamanaka factor shown to inhibit interferon signaling. Our analysis also highlights that some IAGs could be further induced upon infection, including IFITM1 and 28 genes known to have a strong antiviral effect in ESCs. Moreover, co-expression with IFN during infection suggests the underlying IAGs network converges with the canonical interferon pathway. In general, our results indicate that maintaining IAGs expression is a key property of pluripotency. This regulatory network permits limited IAG inductions via interferon signaling upon infection, highlighting a unique antiviral immune response mounted by stem cells.
Exploring single-cell gene expression of immune adaptations in Indigenous Communities from México and Colombia

Octavio Zambada-Moreno

Octavio Zambada-Moreno, Jose Antonio Corona-Gomez, Aaron Espinosa-Jaime, Carlos Humberto Ortiz-Ramirez, Andres Moreno-Estrada, Marisol Espitia-Fajardo, Angel Zamudio-Jacome, Juan Pablo Peña-Sanchez, Yuberlis Gonzalez-Ipuana, Guillermo Barreto

Centro de Investigación y de Estudios Avanzados del IPN (Mexico), Universidad del Valle (Colombia)

Understanding the genetic basis of adaptive phenotypic variation has been a long goal in evolutionary biology. In humans, this has been particularly challenging because genomics studies and resources have been focused mainly on European ancestries, missing most of the available genetic variation in Latin American populations. To help close this gap, we are profiling both genetic and regulatory variation at the level of single-cell gene expression in indigenous Mexican and Colombian populations, representing diverse geographic and cultural regions. Since the immune system is one of the most responsive to rapid adaptation events, we will analyze gene expression changes in immune cell types as a function of diet, environment, and thus, different pathogen exposure. We are genotyping and comparing transcriptional profiles across 300 individuals from communities in México and Colombia through a single-cell RNAseq approach. To reach communities in remote and isolated locations, we have successfully implemented a new protocol for the isolation and cryopreservation of peripheral blood mononuclear cells (PBMCs) on-site, eliminating the need for specialized equipment and minimizing infrastructure constraints, providing efficiency and flexibility during fieldwork. We aim to leverage both genomic and transcriptional information to identify expression quantitative trait loci (eQTLs) at the cell-type level. This will enable enhanced resolution in quantifying ancestry-specific abundances of crucial molecules controlling immune function, such as receptors and immune response activators. This work is part of a broader project led by the LatinCells consortium, which aims to create a comprehensive human cell map of Latin American diversity.
The molecular bases of red-to-yellow color variation in parrots
Roberto Arbore

Roberto Arbore
CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, University of Porto (Portugal), Washington University in St. Louis (USA)

Parrots produce stunning plumage colors by the endogenous synthesis of a unique class of pigments known as psittacofulvins. A polyketide synthase has been shown to be required for psittacofulvin biosynthesis, but the molecular mechanisms explaining variation in psittacofulvin coloration remain to be elucidated. We found that red and yellow colors in parrots result from the differential deposition during feather development of two classes of psittacofulvins differing by a simple chemical modification. Then, we integrated genetic mapping with gene expression and functional experiments to identify a housekeeping enzyme as the key regulator of a red/yellow polymorphism in wild parrot populations. Finally, through fine mapping and single-cell genomic techniques (scRNA-seq and snATAC-seq) we traced this polymorphism to a single point mutation in a conserved regulatory element exclusive to differentiating keratinocytes in developing feathers. The simplicity of the proposed genetic and enzymatic mechanisms offers an explanation for the exceptional evolutionary lability of psittacofulvin-based colors throughout parrot evolution. Our study illuminates the molecular mechanisms governing the colorful plumage of parrots, offering insights into the evolutionary origins of key innovations via the recruitment of genes with essential cellular functions.
Characterizing neuronal evolution between recently diverged species at single-cell resolution

Taylor L. Cooper

Taylor L. Cooper, Haowen He, Kathryn N. Leatherbury, Nikesh M. Kumar, Jeffrey T. Streelman, Patrick T. McGrath
Georgia Institute of Technology (USA)

Evolution of neuronal cell types contributes to the remarkable diversity of behaviors between species. The cellular heterogeneity of vertebrate brains and complex regulation of gene expression within these cell types can now be analyzed using single nucleus RNA- and ATAC-sequencing. To understand how neural cell types evolve, we utilize simultaneous multi-omic techniques in the Lake Malawi cichlid flock, a classical example of an evolutionary radiation that produced >800 species distinguished by social and feeding behaviors which allowed the exploit of diverse ecological niches. We investigate cell type differences in three brain regions, the telencephalon, hypothalamus, and pituitary, across 15 species (one male and one female each) representing four of the seven major ecogroups in the lake. We find differences in cell-proportion in behaviorally relevant neurons, supporting a role for neurogenesis in the evolution of these species. We also identify differentially expressed genes within conserved cell types that distinguish lineages within the lake. Further, we will discuss the association of large putative 'supergenes' that segregate within the lake with chromatin accessibility and gene regulation. Our work provides a framework to understand the remarkable diversity of social and feeding behaviors that characterize this impressive adaptive radiation.
Animals exhibit a rich diversity of behaviors and perceptual capacities to interact with the complex environment. Mate preference is one such rapidly evolving adaptive behavior during the speciation process. Closely related but reproductively isolated species often evolve unique mating cues and the mate preference for conspecific mating cues. Heliconius butterflies represent a classic example of adaptive radiation with extraordinary color pattern divergence. These diverse wing colorations are also used as mating cues for male mate choice. Previous studies have shown that Heliconius males with divergent mate preferences differ at multiple layers of the peripheral visual system, including photoreceptor peak sensitivity and inter-photoreceptor connections. To better understand the transcriptomic changes underlying this color vision diversification, we conducted single-nuclei RNAseq (snRNAseq) and profiled more than 80,000 nuclei from both adult male and female Heliconius butterflies, spanning from a polymorphic population to species of subgenus-level divergence (~14 Mya). We found that while most cell type abundances are similar across species and sex, a group of glial cells is highly variable in males and its abundance is associated with the inter-photoreceptor connection variation. Surprisingly, this glia cell type is marked by sens-2 expression, which is a top candidate gene in the genome-wide association study (GWAS) for male mate preference in Heliconius. Overall, our snRNAseq results provide a comprehensive cell atlas for butterfly retina and enable us to identify cell type abundance and regulatory network changes in a group of butterflies undergoing rapid color vision evolution.
Comparative single-cell regulome reveals evolutionary innovations in neural progenitor cells during primate corticogenesis

Xin Luo

Xin Luo, Yuting Liu, Yiming Sun, Cheng Li, Bing Su
Kunming Institute of Zoology (China), Peking University (China)

The cellular and genetic mechanism underlying the human-specific features of cortex development remains unclear. We generated a cell-type resolved atlas of transcriptome and regulome of the developing macaque and mouse prefrontal cortex, and conducted evolutionary analyses with the published complementary human data. We discovered a primate-specific expansion of two neural progenitor subclasses, glia-committed radial glia (RG) and truncated RG. Specifically, the human neural progenitors show extensive transcriptional rewiring in the growth factor and extracellular matrix pathways. Expression of the human-specific progenitor marker ITGA2 in the cortex of fetal mouse promotes progenitor proliferation and an increased upper-layer neuron proportion. We demonstrate that these transcriptional divergences are primarily driven by the activity changes of the distal regulatory elements in the genome. Markedly, the chromatin regions with human-gained accessibility enrich the human-fixed sequence changes, as well as sequence polymorphisms associated with intelligence and neuropsychiatric disorders. Our results uncover evolutionary innovations in neural progenitors and gene regulatory mechanism during primate cortex evolution.
Evolutionary origin of the chordate nervous system revealed by amphioxus developmental trajectories
Yichen (Serena) Dai

Yichen Dai, Yanhong Zhong, Rongrong Pan, Yongheng Fu, Liang Yuan, Juan Du, Meng Li, Xiao Wang, Huimin Liu, Chenggang Shi, Gaoming Liu, Pingfen Zhu, Sebastian Shimeld, Xuming Zhou, Guang Li
Chinese Academy of Sciences (China), Institute of Zoology, University of Oxford (United Kingdom), Xiamen University (China), Xinjiang Normal University (China)

Amphioxus is phylogenetically well positioned for advancing our understanding of vertebrate evolution. To understand chordate cell type evolution and examine hypotheses of vertebrate neural cell type origins, we developed and analyzed a single-cell RNA-sequencing dataset from seven amphioxus embryo stages. Numerous new amphioxus cell types were identified, including homologs to vertebrate aorta-gonads-mesonephros, hypothalamus, and neurohypophysis, rooting the evolutionary origin of these structures. Ancestor-descendant reconstruction of cell trajectories of amphioxus and other species then allowed us to deduce expression dynamics of transcription factor genes throughout embryogenesis, identifying three ancient developmental routes forming chordate neurons. This also served as a framework for characterizing cell specification at the mechanistic level and, exploiting this, we used gene knockout to examine the function of Lhx3/4, Msx1, FoxQ2a, Brachyury-1 and Brachyury-2, key transcription factors involved in neural specification. Our results demonstrate three developmental origins for the vertebrate nervous system: an anterior FoxQ2 dependent mechanism that is deeply conserved with invertebrates, a less-conserved route leading to more posterior neurons in the vertebrate spinal cord, and a mechanism for specifying neuromesoderm progenitors that is restricted to chordates. The evolution of neuromesoderm progenitors may have led to a dramatic shift in posterior neural and mesodermal cell fate decisions and the body elongation process in a stem chordate.
S24 - Unlocking the hidden dimensions of genomic diversity within species.
Evolutionary Constraints on Gene Expression in Humans: Implications for Drug Efficacy and Personalized Medicine

Alber Aqil

Alber Aqil, Omer Gokcumen
University at Buffalo (USA)

Previous studies have shown that highly expressed genes in humans tend to be evolutionarily constrained at the sequence level. However, the constraints on the regulation of genes with different magnitudes of expression remain understudied. Here, we use gene expression variability among individuals, measured by the Gini coefficient, as an inversely related proxy for constraint on gene regulation. Our study systematically quantifies the relationship between the expression magnitude (mean expression across individuals) and expression variability for 30,000 genes across 27 tissues using the Genotype-Tissue Expression (GTEx) database. We observe a consistent negative correlation (Spearman’s rho < -0.9; p < 2.2 x 10^-16 across tissues) between expression magnitude and variability. This indicates a stronger constraint on the regulatory architecture of genes with a higher magnitude of expression. Genes that are exceptions to this trend may be interesting in the context of both balancing selection and disease. In parallel, the analysis of these expression trends allowed us to test the hypothesis that the efficacy of drugs is negatively affected by the high expression variability of target genes, which hinders the determination of optimal drug dosage for individual patients. Indeed, genes targeted by FDA-approved drugs that were subsequently withdrawn exhibit greater expression variability compared to drugs still in use (p < 2.2 x 10^-16; Cohen’s d=0.5). These insights will inform drug discovery studies and advance the development of personalized medicine.
Micro-C reveals 3D genome conservation patterns in rice species
Amina Kurbidaeva

Amina Kurbidaeva, Sonal Gupta, Michael Purugganan
NYU (USA)

Chromatin structure plays a central role in evolution of genomic architecture. It plays an important role in speciation and domestication. Thus, it is of fundamental importance to determine the dynamics of 3D chromatin structure during evolution. The structure of chromosomes at different scales has been an object of intense study in many organisms. Yet, there hasn't been a study which compares and quantifies the overall chromosome structure of entire genomes, and correlates that with the evolutionary times since speciation events. Here, we performed a comprehensive analysis of 3D chromatin structure in five rice genomes, including three different species and three domesticated rice varieties. We focus on short evolutionary distances, spanning not more than 2.41 MYA. We show that even at such small evolutionary scale, there are clear differences in overall chromatin structure, and we can quantify those differences using CHESS. We show that these differences positively correlate with evolutionary times and with sequence similarity between the genomes. We also identify the genetic and epigenetic determinants of 3D genome similarities and differences. At high resolution, TADs are not evolutionary stable, and we identified signatures of conserved TADs and boundaries. We conclude that the overall 3D chromatin structure can be estimated and quantified with CHESS, and it positively correlates with evolutionary distances between genomes, and the individual TAD conservation between genomes depends on their genetic and epigenetic composition.
Studies examining the evolution of genomes have been mainly focused on sequence conservation. However, the inner working of a cell implies a tightly regulated crosstalk between complex gene networks, controlled by small dispersed regulatory elements of physically contacting DNA regions. How these different levels of chromatin organization crosstalk in different species underpins the potential for genome evolutionary plasticity. It is my intention to provide an overview on the evolution of chromatin organization across the Animal Tree of Life. I will discuss on general aspects of the mode and tempo of genome evolution to later explore the multiple layers of genome organization. We propose that both genome and chromosome size modulate patterns of chromatin folding and that chromatin interactions facilitate the formation of lineage-specific chromosomal reorganizations, especially in germ cells. Overall, analyzing the mechanistic forces involved in the maintenance of chromatin structure and function of germ line is critical for understanding genome evolution, maintenance, and inheritance.
Genome-wide analysis of transcription initiation in Paramecium

Berenice Jimenez-Marin, Jean-Francois Gout, Weibo Zheng, Taylor Raborn, Tia Swenty, Timothy Licknack, Michael Lynch
Arizona State University (USA)

The timing and intensity of transcription of different genes is fundamental to survival of the organism, and thus is orchestrated by sequences located upstream of the protein-coding portion of the gene itself. In the single-celled eukaryotes of genus Paramecium, the macronuclear genome is remarkably compact and streamlined for gene expression. Interestingly, member species of the Paramecium aurelia species complex have undergone two whole genome duplications, resulting in differential patterns of gene retention and loss which are influenced by their expression levels. As means to explore eukaryotic transcription in a streamlined genome and shed light on the role of sequence architecture on gene expression and loss, we analyzed the distribution and diversity of candidate transcription initiation sites in Paramecium seriaurelia and Paramecium tetraurelia. A total of 10,652 genes (30.5% of the P. seriaurelia genome) and 22,098 genes (55.9% of the P. tetraurelia genome) were queried for their transcription initiation site (TIS) diversity and conservation within and between species. Our analysis suggests that both Paramecium aurelia genes preferentially have narrow transcriptional initiation regions, and >90% of genes in both species have considerably short 5’ UTRs (<55 bp). The short region in which transcription initiation occurs most frequently and the short length of the Paramecium 5’ UTRs suggest differences in promoter regions between Paramecium species, in addition to divergence in 5’UTR architecture under relatively strict sequence-length constraints, influence gene retention and loss.
Population history and chromosomal speciation in gelada monkeys
Brooklynn R. Scott

Brooklynn R. Scott, Ferehiwot Ayele, Alemayehu Lemma, Jacinta C. Beehner, India A. Schneider-Crease, Amy Lu, Thore J. Bergman, Kenneth L. Chiou, Noah Snyder-Mackler
Addis Ababa University (Ethiopia), Arizona State University (USA), Stony Brook University (USA), University of Michigan (USA)

Genomic studies of the demographic history of extant populations have revealed a myriad of processes underlying genetic diversity, barriers to hybridization, and speciation. Geladas (Theropithecus gelada) are an especially interesting example of a long-lived primate that may be in the midst of an ongoing speciation event. We recently identified a novel, derived karyotype (2n=44) in one population ("northern geladas") that resulted from a fission on chromosome 7. This karyotype evolved recently in geladas, differing from the ancestral 2n=42 karyotype that other populations ("central geladas") share with geladas’ close relatives, including baboons (Papio spp.) and macaques (Macaca spp.). Here, we used population resequencing data from northern and central geladas to reconstruct their population history and identify when and how this fission event occurred. Using two complementary approaches, we estimate that the two populations diverged ~175 kya, with substantial changes in population size after the split. This suggests that the novel karyotype emerged in the northern population around the time of divergence. The limited migration estimated between the populations suggest that the fission represents a strong barrier to hybridization. The genetic divergence and karyotype difference between populations suggest emerging reproductive barriers, which carry substantial implications for gelada speciation and conservation efforts.
The distribution of fitness effects of mutations in enhancers, promoters, and conserved non-coding regions

Chenlu Di

Chenlu Di, Kirk Lohmueller
Los Angeles (USA), University of California

Functional non-coding regions are increasingly identified in the human genome by various methods. The challenge now is to understand the fitness effects of mutations in these non-coding regions. We developed a pipeline to infer the distribution of fitness effects (DFE) of non-coding mutations. Using polymorphism data from the 1KGP, we inferred DFEs of new mutations in enhancers, promoters, and conserved non-coding regions. We found that mutations in enhancers are frequently neutral, with a small proportion of weakly and strongly deleterious mutations. Promoters, in contrast to enhancers, show a lower proportion of neutral mutations and more weakly deleterious mutations. To investigate the temporal nature of the evolution of noncoding sequences and whether selection pressure changed over time, we estimated the fitness effects of mutations in regions of the genome showing different degrees of sequence constraint in mammals. We found the highest proportion of deleterious mutations in regions of the genome that are most conserved in both mammals and primates alone. However, regions of the genome conserved in primates alone showed a higher proportion of deleterious mutations than mutations in regions of the genome that were only conserved when considering all mammals. This pattern appears in putative enhancers, promoters, and when classifying conserved regions at two thresholds. Our results highlight the dynamic nature of the evolution of gene regulation, with changing selection coefficients over deep evolutionary time. More broadly, our work has implications for using comparative genomic approaches to detect non-neutrally evolving sequences in the human genome.
Gene expression variability broadly refers to a gene’s tendency to vary in expression due to stochastic fluctuations or differences in genetic, epigenetic, or environmental factors. Variability due to transcriptional noise and cell-to-cell heterogeneity has been well-studied in isogenic populations of bacteria and yeasts. However, for complex organisms with multiple cells, tissues, and organs sharing a genetic background, the interplay between variability, gene and organ function, and gene regulation remains an open question. We used 3’-end bulk RNA barcoding and sequencing (BRB-seq) to generate transcriptome profiles spanning at least nine organs in outbred individuals of three ray-finned fishes: zebrafish, northern pike, and spotted gar. Per organ, we measured inter-individual expression variation per gene independent of mean expression. We observed that lowly variable genes are enriched in cellular housekeeping functions whereas highly variable genes are enriched in stimulus-response functions. Furthermore, highly variable genes evolve under weaker purifying selection at the coding sequence, indicating that intra-species gene expression variability predicts inter-species protein sequence divergence. For genes with organ-biased expression, we inferred differences in selective pressure on gene regulation depending on their top organ. Genes that are most expressed in the brain show low expression variability across non-nervous organs, suggesting stabilizing selection on cis-regulatory evolution of brain-biased genes. By contrast, liver-biased genes have highly variable expression across other organs, implying weaker cis-regulatory constraints. Thus, investigating expression variability across organs can provide new insights into gene regulation in animals.
Dobzhansky-Muller incompatibilities are one proposed contributor to speciation, where incompatible alleles result in segregation distortion and thereby impact population-wide allele frequencies. In this study we detect and quantify segregation distortion loci in crosses produced from a diverse sampling of Arabidopsis lyrata and A. halleri populations. We show that both the frequency of occurrence and the effect size of distortion loci increase as the parents’ genetic distance from one another increases. We also show that distorter loci occur not only within interspecific hybrids, but also in intraspecific hybrids produced from isolated population crosses. Finally, we identify correlated distortion effects between loci that repeatedly appear on two different chromosomes in multiple F1 individuals. Our study demonstrates that pollen-acting segregation distortion may be ubiquitous, and shows how segregation distortion contributes not only to the ongoing reproductive isolation between A. halleri and A. lyrata, but also between very recently diverged populations of the same species. More generally, this work demonstrates that the rapid evolution of haploid-affecting incompatibilities are likely to disproportionately contribute to suppressing gene flow between diverging populations.
Relating FST to Information Theory gives multiple insights.
Daniel R Tabin

Daniel R Tabin, Nick J Patterson, David E Reich
Broad Institute (USA), Harvard (USA)

Wright's FST can be thought of in Information Theory terms. Specifically FST can be seen to be related to the ratio of Shannon entropy without subpopulation knowledge and the Shannon entropy with subpopulation knowledge, where the "subpopulation" can be any subdivision in the data meaningful or otherwise. Thinking of FST in this way offers a number of benefits. First and foremost, this makes the meaning of FST, as Wright himself describes it, much clearer than the traditional description of differentiation, which is easily confused with the related but non identical concept of difference. Likewise, we view thinking of FST as the ratio of information with and without subpopulation knowledge as more intuitive to students and the general public when compared with thinking of FST as variance. Many seemingly paradoxical facts about FST, such as the fact that Han and French have FSTs similar to those between Amazonian groups, and that African-Eurasian FST went down after Eurasians mixed with Neanderthals, become intuitive when thinking of FST in terms of information. Additionally, many insights can be gleaned from this view. For example, we show how Information Theory can be used to compare FST with other types of relatedness, such as genealogical relatedness, and we make an argument that the formula for FST should depend on the ploidy of the species, with haploid and tetraploid organisms requiring a different formula than the traditional diploid one.
Alternative splicing and environmental adaptation in wild house mice

David N. Manahan

David N. Manahan, Michael W. Nachman
Berkeley (USA), Museum of Vertebrate Zoology (USA), University of California

A major goal of evolutionary genetics is to understand the genetic and molecular mechanisms underlying adaptation. Previous work has established that changes in gene regulation may contribute to adaptive evolution, but most studies have focused on mRNA abundance and only a few studies have investigated the role of post-transcriptional processing. Here, we use a combination of exome sequences and short-read RNA-Seq data from wild house mice (Mus musculus domesticus) collected along a latitudinal transect in eastern North America to identify candidate genes for local adaptation through alternative splicing. First, we identified alternatively spliced transcripts that differ in frequency between mice from the northern-most and southern-most populations in this transect. We then identified the subset of these transcripts that exhibit clinal patterns of variation among all populations in the transect. Finally, we conducted association studies to identify cis-acting splicing quantitative trait loci (cis-sQTL), and we identified cis-sQTL that overlapped with previously ascertained targets of selection from genome scans. Together, these analyses identified a small set of alternatively spliced transcripts that may underlie environmental adaptation in house mice. Many of these genes have known phenotypes associated with body size, a trait that varies clinally in these populations. We observed no overlap between these genes and genes previously identified by changes in mRNA abundance, indicating that alternative splicing and changes in mRNA abundance may provide separate molecular mechanisms of adaptation.
If they don't have proteins, let them have RNA: evolutionary rescue of defective ribosome mutants by changes to transcription

Enea Franceschini

Enea Franceschini, Paul B Rainey
CBI, CNRS (France), Department of Microbial Population Biology, ESPCI Paris, Laboratory of Biophysics and Evolution, Max Planck Institute for Evolutionary Biology (Germany), Université PSL

The ribosome is one of the more conserved molecular machines present in all kingdoms of life. In the last couple of decades much progress has been made to describe it and determine its function. Despite this, there is still a lack of understanding of both the specific role performed by some of its protein components and of factors involved in the loss and acquisition of said components over evolutionary time. In our experiments we used an experimental evolution framework to test if and how bacteria can adapt to the loss of ribosomal proteins. Five different ribosomal proteins of varying ancestry were deleted from the chromosome of the model bacterium Pseudomonas fluorescens. As expected, these mutations caused a decrease in growth rate and translation speed. After more than 500 generations each evolving population showed partial or full recovery of fitness. Genome sequencing of derived populations allowed compensatory mutations to be identified, many of which targeted genes that couple transcription and translation. Further experiments showed stronger loss of fitness caused by the deletions we produced when there is lack of phosphate in the growth medium. Since phosphate is needed for nucleotide synthesis, this could point to a higher need of RNA transcription to maintain fitness. These results together suggest that ribosomal protein gene deletions not only affect translation speed, but also alter the required transcription rate needed to maintain growth. The need to optimize the balance between transcription and translation in a given environment could affect the evolution of the translation apparatus.
The evolution of transcriptome complexity using Oxford Nanopore long-read sequencing technology.
Gabriela Santos Rodriguez

Gabriela Santos-Rodriguez, Favour Oyelami, Agin Ravindran, Akanksha Srivastava, Chi Kin Ip, Pallavi Gupta, Jeannette Villanueva, Helen Elizabeth King, Abigail Grootveld, Ishaan Gupta, Grazielle Santos Vieira Helaine, Nikolay Shirokikh, Eduardo Eyras, Robert James Weatheritt
Australian National University (Australia), Department of Biochemical Engineering & Biotechnology, EMBL Australia (Australia), Garvan Institute of Medical Research (Australia), Indian Institute of Technology Delhi (India), John Curtin School of Medical Research (Australia), RNA Institute (Australia), School of Biotechnology and Biomolecular Sciences (Australia), University of New South Wales (Australia), Victor Chang Cardiac Research Institute (Australia)

The increase of biological complexity is partly due to the dynamic generation of unique cell-specific transcriptomes. Co- and post-transcriptional mechanisms including alternative splicing (AS), alternative promotor usage (APU); alternative polyadenylation (APA) have been identified as major drivers of transcriptome diversity and are known to drive lineage-specific phenotypic effects. However, this research is limited to the assessment of individual events (exons / 3'UTRs) due to the limitations of short-read sequencing To gain insight into the conservation of full-length transcript expression and the coordinated regulation of post-transcriptional mechanisms, we used Oxford Nanopore long-read sequencing technology to analyse isoform conservation across six tissues from five mammalian species and an outgroup. We found that multiple transcript isoforms have conserved tissue-specific expression, where >40% of the analysed genes in mammals had conserved transcripts with major isoform switching in a tissue-specific manner. Additionally, we found that conserved transcripts are enriched in alternative first exon splicing events. Furthermore, we studied the evolutionary conservation of coordinate-splicing events (the co-association of splicing events in the same transcript). We found that mammalian conserved events are depleted in mutually exclusive coordinate-splicing events and enriched in mutually associated events (present in the same transcript) and with a strong enrichment for tissue-specific expression. Together, our work uncovers extensive conservation of post-transcriptional regulation and provides a resource for understanding the transcriptome conservation across mammalian evolution.
The genetic basis of long-distance migration in Common swifts (Apus apus)
Heiman Ho

Heiman Ho, Leyi Su, Zhou Wu, Chenqing Zheng, Yang Liu
IMBIM, School of Ecology, Sun Yat-sen University (China), The Roslin Institute and Royal (Dick) School of Veterinary Studies R(D)SVS, University of Edinburgh (United Kingdom), Uppsala University (Sweden)

The genetic basis of long-distance migration has long been one of the interests in evolutionary biology. While integration of genetic, behavioural and physiological factors has complicated the decoding of general signs in migration adaptation, research on different systems is also essential to achieve this goal. Here, we applied WGS on 47 swifts, including 13 Apus pallidus, which migrate from southern Europe to Central Africa (~4500km), along with 17 of its sister species – A. apus apus – which have relatively longer migratory distance from Europe to South East Africa (~7000km), as well as 17 A. a. pekinensis moving between Asia and South Africa (~13000km). 11,405,577 SNPs were yield with at an average sequencing depth of ~20.5X. We adopted sliding windows approach with 50kb windows and estimated FST, dXY and XP-nSL to find the top 5% diverged regions which likely to be selected when comparing A. pallidus to A. a. apus and A. a. pekinensis separately. Correspondingly, we found a total of 128 and 75 non-overlapping genes, suggesting the 2 subspecies having different evolutionary routes. Candidate genes relating to mitochondrial functions such as IMMP2L and RYR2 proved the importance of mitochondria during long-distance migration. In addition, we found an overlapping gene – Interleukin 1 Receptor Accessory Protein Like 1(IL1RAPL1) – in both comparisons with signs of selection, which is likely to play a role in cerebellar development, cognitive functions and memory abilities. We suggest that these genes might be important in maintaining long-distance migratory behaviour.
Genetic and Epigenetic Data Reveal Complementary Climate Adaptation Pathways in an Endangered Sea Turtle

James D Gilbert

James D Gilbert, Eugenie C Yen, Samuel J Shrimpton, Stuart RB Negus, James O Bazely, Alice Balard, Kirsten Fairweather, Albert Taxonera, Christophe Eizaguirre

Project Biodiversity (Cape Verde), Queen Mary University of London (United Kingdom)

Genomic diversity underpins a population’s adaptive potential and resilience to environmental change. While variation in genetic sequence is often used to predict population persistence, genomes have other levels of diversity. DNA methylation is an epigenetic modification that regulates gene activity and can be both genetically controlled and environmentally induced. How genetic and epigenetic diversities determine a population’s ability to respond to climate change, particularly for wild, non-model organisms remains elusive. Here, we combined nanopore and bisulfite whole genome sequencing applied to endangered loggerhead sea turtles (Caretta caretta) nesting on different islands of the Cabo Verde archipelago, along a temperature cline. Sea turtles must acquire substantial energy reserves between breeding seasons, making them sensitive to temperature changes that affect energy storage and their ability to return to their natal place to reproduce. While we detected a large number of highly differentiated genetic regions among turtles from different islands, none overlapped with regions of high variation in DNA methylation. Many regions were also detected with low genetic differentiation but differential methylation. We also find island-specific methylation of genes related to fatty acid metabolism, suggesting variation in energy storage linked to migration along the temperature cline. These insights highlight how epigenetic mechanisms, operating independently of genetic diversity, may underpin resilience to climate change.
Evolution of Evolvability In Rapidly Evolving Populations (if accepted for talk); The Dynamics of Horizontal Gene Transfer In Rapidly Adapting Populations (if accepted for poster)

James Ferrare

James Ferrare, Benjamin Good
Stanford University (USA), none (USA)

(if accepted for talk): Mutations can alter the short-term fitness of an organism, as well as the rates and benefits of future mutations. While numerous examples of these evolvability modifiers have been observed in rapidly adapting microbial populations, existing theory struggles to predict when they will be favored by natural selection. Here, we develop a mathematical framework for predicting the fates of genetic variants that modify the rates and benefits of future mutations in linked genomic regions. We derive analytical expressions showing how the fixation probabilities of these variants depend on the size of the population and the diversity of competing mutations. We find that competition between linked mutations can dramatically enhance selection for modifiers that increase the benefits of future mutations, even when they impose a strong direct cost on fitness. However, we also find that modest direct benefits can be sufficient to drive evolutionary dead-ends to fixation. Our results suggest that subtle differences in evolvability could play an important role in shaping the long-term success of genetic variants in rapidly evolving microbial populations. (if accepted for poster): In rapidly adapting populations, horizontal gene transfer can facilitate adaptation by breaking down linkage disequilibrium and recombining beneficial mutations onto a single genetic background. When this process is sufficiently rare, basic aspects of evolutionary dynamics are poorly understood. Here, we derive analytical expressions showing when recombination events are favored by natural selection and how these events affect adaptation.
Beyond averages: transcriptional variability in outbred Drosophila melanogaster and its environmental and genetic dependence

James Phipps-Tan Sheng Yi

James Phipps-Tan, Julien F. Ayroles, Luisa F. Pallares
Department of Ecology and Evolutionary Biology, Friedrich Miescher Laboratory of the Max Planck Society (Germany), Lewis-Sigler Institute for Integrative Genomics, Princeton University (USA)

Why do different individuals of a population have different phenotypes? Most approaches to this age-old question compare average trait values of individuals with different genotypes or exposed to different environments. However, this ignores individual variation around averages (variability), which is increasingly recognized as a potentially adaptive property arising from developmental noise, bet-hedging and environmental robustness. To investigate the adaptive potential of gene expression variability, we collected the largest-known set (~2000) of matched head transcriptomes and genomes of outbred Drosophila melanogaster individuals raised for a generation on either a control or high sugar diet. This gave us unrivalled statistical power to test whether transcriptional variability is (1) gene-specific, (2) modified by environment and, for the first time in any animal system, (3) has an environmentally-dependent genetic basis. We show that a gene's transcriptional variability, measured by the outlier-robust median absolute deviation statistic (MAD), correlates positively with its nucleotide diversity and heritability, negatively with its essentiality, and non-linearly with its connectivity. Strikingly, high sugar increases transcriptional variability in 77% of genes and changes the identities of almost half of the 5% least variable genes. In stark contrast, the 5% most variable genes remain largely the same across diets. Finally, we mapped the genetic basis of transcriptional variability using a new MAD-based mixed modelling approach that explicitly tests for genotype-by-environment interactions. Taken together, our observations that transcriptional variability changes with environmental and genetic variation hint towards it being an adaptive and evolvable trait.
A population genetics-based approach to uncover 3D genomics variation among closely related individuals
Juan Antonio Rodríguez

Juan Antonio Rodríguez, Sarah S. T. Mak, Maria Marti-Marimon, Sofia Marcos, Claudia Fontserè, Antton Alberdi, Marc A. Marti-Renom, M Thomas P Gilbert
Applied Genomics and Bioinformatics, CNAG (Spain), Center for Evolutionary Hologenomics, Centre Nacional d’Anàlisi Genòmica, Faculty of Health and Medical Sciences, Globe Institute, University of Copenhagen (Denmark), University of the Basque Country (UPV/EHU) (Spain)

Characterizing genomic diversity is crucial to exploiting genetics in the basic and applied sciences. A key aspect that is gaining increased attention, is how the genome is not just a one-dimensional linear thread, but rather a three-dimensional (3D) folded polymer. Nuclear DNA fiber folds into non-random 3D structure, whose preservation is vital for correct cell homeostasis and gene-regulatory mechanisms. These mechanisms have recently gained considerable interest, thanks to chromosomal conformation capture techniques, that allow reconstruction of the genomic structure. However, 3D genomics analyses are yet to explore how 3D genome structure varies across individuals within a population. Neglecting to consider such variation at the population scale (should it exist), could result in sampling biases and/or yield exaggerated differences in 3D structure when comparing pairwise conditions. Using a cohort of 16 chickens from a single population raised in highly controlled conditions as a model, we evaluated the populational variability of 3D genome structure at several levels. We found that up to a 10% of the genome compartmentalization may show variability between individuals of a same population. While a significant fraction of 3D genomic variability for autosomes can be explained by sex and/or age, a large fraction of the remaining individual, populational variance stays unaccounted. As our chicken individuals derive from a controlled captive breeding population, we hypothesize that 3D genome variation in wild populations can be even larger. Hence, acknowledging the understudied standing 3D genomic variation of samples can be helpful to correctly frame and interpret discoveries.
Characterisation and quantification of deleterious genetic variants in non-model organisms: from present to extinct species.
Julia Höglund

Julia Höglund, Seyan Hu, Martijn F.L. Derks, Love Dalén, Mirte Bosse
Amsterdam Institute for Life and Environment (A-LIFE), Centre for Palaeogenetics, Dept. of Animal Breeding and Genomics, Dept. of Zoology, Section Ecology & Evolution, Stockholm University (Sweden), Vrije Universiteit Amsterdam (Netherlands), Wageningen University & Research (Netherlands)

Many animal populations are currently undergoing dramatic decline, to which human activities have significantly contributed. When a population becomes smaller, threats affecting the species’ genome become larger. A reduced population size increases the risk of inbreeding and loss of beneficial genetic variation, results in a worse ability to adapt to changes, and can potentially lead to extinction. Hence, there is a critical need to quantify damaging genetic variation to aid conservation efforts and understand past extinctions. One approach to characterise damaging genetic variants is to score them based on predicted deleteriousness. However, such models are currently species-specific and often unreliable beyond model organisms. To address this, we utilised the domesticated pig reference genome and estimated substitution and mutation rates from several ancestral nodes to compare the impact of evolutionary time on scoring and annotation. By using dense sequencing data, we will train a scoring model within domesticated pig, test it in wild boar and lastly extending it to endangered pig species. Preliminary results suggest similar distributions of variant effects across ancestral nodes. In contrast to previous similar scoring models, this model extends across similar species already in the initial score estimation. These scores will then be utilised in quantifying genetic load, to estimate how much genetic variation is contributing to animal extinction. By extending the model beyond model species, we enhance its applicability to several species, hopefully providing insights into the possible genetic factors underlying species decline.
Population genomics analyses are usually based on samples obtained from genetic variation in adult individuals. However, conducting genetic analyses using adult samples limits our ability to understand the full range of genomic variation such as during early development. Songbirds are ideal to study such variation because they possess an additional chromosome in the germline that is eliminated from somatic tissue. However, the precise mechanism of such chromosome elimination remains obscure. To address the intriguing question of developmental genomic variation, we obtained embryonic DNA from 14 great tit individuals isolated from eggs harvested in a local forest population in Bielefeld (Germany). We conducted a series of whole genome population genetic analyses by contrasting embryo samples to adult great tit individuals sampled across whole Europe. Interestingly, our embryo samples had a higher genetic diversity than the adult blood samples. We also found remarkable differences in genetic diversity between female and male embryos. Our observed patterns are hard to explain with standard Mendelian genetics or processes of selection like heterozygote advantage. Instead, they may suggest other forms of genomic variation, e.g. sex- and age-specific structural variation or a cryptic karyotype polymorphism. This illustrates that relying on standard pipelines for population genomic analysis may lead to false conclusions. Here, our analysis of wild embryos has made us able to detect life-stage dependent genetic variation, which would otherwise have been unnoticeable.
Exploring the evolutionary basis of gene expression trends in murine salivary glands

Luane Jandira Bueno Landau

Luane Jandira Bueno Landau, Achisha Saikia, Sarah Knox, Omer Gokcumen

UT Southwestern Medical Center (USA), University at Buffalo (USA), University of California San Francisco (USA)

The major salivary glands (parotid, submandibular and sublingual glands) show a high diversity in gene expression in mammals. More specifically, sex-bias in gene expression is present for the three glands, but variation across mouse strains was not investigated yet. Here, we use mouse as a model to understand the evolution of gene expression in the salivary glands. We generated sex- and tissue-specific transcriptome from the three major salivary glands of two mouse strains (CD1 and C57). We compared expression in each gland for strain and sex. Submandibular (SM) shows the higher number of differentially expressed genes (DEGs) across sexes. When performing a GO enrichment analysis for the sex-biased genes, we found genes highly expressed in females show an enrichment for transmembrane transport (p<10-7). This is consistent with sex differences in cell composition in the murine SM gland. As females present an increased number of acinar cells, these patterns could be related to differences in salivary fluid secretion. Alternatively, we found that the sublingual (SL) gland shows the higher number of DEGs between strains. Highly expressed genes in CD1 strain show an enrichment for developmental processes (p<10-2). This is consistent with reported differences in development across mouse strains for multiple tissues. Two highly expressed genes in CD1 strain (Galtn10 and St6GalNa) are related to glycosylation, which are known to have major phenotypic impacts on saliva biology. This dataset provides the necessary means for the analysis of the evolutionary bases of the salivary glands gene expression.
Unveiling the Diversity and Evolutionary Significance of Tandem Repeats
Luis Gerardo Fernandez Luna

Luis G Fernandez Luna, Rori Rohlf, Carolina De Lima Adam, Sebastián Iturbe, Diego Ortega-Del Vecchyo
International Laboratory for Human Genome Research (Mexico), National Autonomous University of Mexico (Mexico), University of Oregon (USA)

Repetitive sequences, known as tandem repeats (TRs), are highly polymorphic. Over 1 million identified loci constitute more than 3% of genomic DNA in humans. TRs exhibit significant diversity due to their complex underlying mutational processes, characterized by high mutation rates that depend on changes in repeat length, motif structure, and sequence. These fast-evolving regions serve as a significant source of genetic variability that may underlie adaptation processes within and between species. These regions have largely remained unexplored and ungenotyped due to their complexity. However, recent advances in long-read sequencing technologies and computational genotyping tools facilitate a more systematic study of TR variation. Now is feasible to address the roles of TRs in regulatory regions, their involvement in health and disease, and their contributions to complex traits such as height. Despite these advances, TRs remain an under-studied source of adaptive genetic variation due to the lack of a statistical framework supporting the comparison of TRs to test evolutionary hypotheses. To comprehend the evolutionary processes underlying the selection of specific TRs that play pivotal roles in the dynamic nature of the genome, we propose the development of a tool named TR evolutionary variance analysis (TREVA). TREVA aims to assess the variance of TRs within and between species, providing a framework to test evolutionary hypotheses regarding TR variation. To validate and draw meaningful inferences from TREVA results, we are simulating comparative TR data under several mutational and evolutionary models. These data will provide a test for TREVA and other similar methods.
Investigating BUSCO databases and performance in the light of a new generation of chromosome-level molluscan genomes

Marcela Uliano da Silva

Marcela Uliano-Silva, Mark Blaxter
Wellcome Sanger Institute (United Kingdom)

The Benchmarking of universal single copy orthologs (BUSCO) is a widely used tool for evaluation of genome assembly quality and completeness. Mollusca is a phylum known to recover low BUSCO scores, and a comprehensive analysis of the possible causes is lacking. We have run BUSCO v5.6.1 with metaeuk and mollusca_db10 for 60 highly contiguous chromosome-level genomes produced by the Darwin Tree of Life Project and others, plus the more fragmented genomes of the 7 species included in mollusca_db10. Our analyses included genomes from main molluscan groups (Scaphopoda, Gastropoda, Cephalophoda, Bivalvia). Out of the 5295 mollusca_db10 BUSCOs, only 41 were found as Complete in all species, and no common BUSCO genes were found to be Missing, Fragmented or Duplicated in all species analysed. A presence/absence analysis of Complete BUSCOs showed clearly that completeness BUSCO recovery has a strong bias towards the species included in mollusca_db10 and its close-relatives, and recovery decreases as further away species are in the phylogeny. We are further investigating if the mostly common missed genes are (i) difficult to predict by homology, (ii) if they are affected by the predictor used (metaeuk/augustus) and the mode run (genome/proteome), or (ii) if these missed genes may not represent universally conserved elements across all molluscan taxa. This investigation aims to better inform our use and understanding of genome quality assessment tools in the light of new chromosome-level genomes for multiple never-before-sequenced species across the main phylogenies.
Unraveling the evolutionary history of polyploid Arabidopsis kamchatica
Maria Vasilarou

Maria Vasilarou, Uliana Kolesnikova, Alison D Scott, Polina Novikova
Max Plank for Plant Breeding research (Germany)

Polyploidy is an enduring force in plant evolution and speciation. The study of allopolyploid species presents a fascinating avenue for exploring the dynamics of polyploid evolution. In allotetraploids, the fusion of two divergent genomes in a single nucleus can lead to dynamic interactions and genomic changes. For example, homoeologous exchange between the two subgenomes may occur, though these are more often detected in synthetic or very young polyploids. Established allopolyploids may instead demonstrate genomic stability and absence of large homeologous exchanges. Arabidopsis kamchatica is an established allotetraploid with a complex evolutionary history, involving multiple hybridization events between A. lyrata and A. halleri. Using de novo genome assemblies and population level resequencing data, we explore the evolutionary trajectory of the two subgenomes within A. kamchatica. In particular, we investigate how the subgenomes interact since their fusion into the allopolyploid state. To that end, we generated a reference genome that allowed us to detect evidence for large genomic rearrangements between the two subgenomes of our A. kamchatica accession. Ultimately, the study of A. kamchatica will enhance our comprehension of polyploid genome evolution while also underscoring the necessity of a population-level analysis for resolving the complex evolutionary origins of allopolyploidization.
Chromatin 3D conformation of a silent genome: the case of nucleated erythrocytes
Mayra Furlan-Magaril

Presented by self
Institute of Cellular Physiology -UNAM (México)

Chicken erythrocytes are nucleated cells often considered to be transcriptionally inactive, although the epigenetic changes and chromatin remodeling that would mediate transcriptional repression and the extent of gene silencing during avian terminal erythroid differentiation are not fully understood. We characterized the changes in gene expression, chromatin accessibility, genome organization and chromatin nuclear disposition during the terminal stages of erythropoiesis. We observe a robust decrease in transcription in erythrocytes, but a set of genes maintains their expression, including genes involved in RNA polymerase II (Pol II) promoter-proximal pausing. Erythrocytes exhibit a reoriented nuclear architecture, with accessible chromatin positioned towards the nuclear periphery together with the paused RNA Pol II. In erythrocytes, chromatin domains are partially lost genome-wide, except at minidomains retained around paused promoters. Our results suggest that promoter-proximal pausing of RNA Pol II contributes to the transcriptional regulation of the erythroid genome and highlight the role of RNA polymerase in the maintenance of local chromatin organization.
Antarctic microbial genomics: insights into cold adaptation and secondary metabolite biosynthesis
Michele Giovannini

Antarctica, one of the most extreme environments on Earth, hosts diverse microbial communities. These microbes have evolved and adapted to survive in these hostile conditions, but knowledge on the molecular mechanisms underlying this process remains limited. The Italian Collection of Antarctic Bacteria [Collezione Italiana Batteri Antartici (CIBAN)], managed by the University of Messina, represents a valuable repository of cold-adapted bacterial strains isolated from various Antarctic environments. In this study, we analysed the genomes of 62 ?-Proteobacteria strains from the CIBAN collection, isolated during Italian expeditions from 1990 to 2005. Genomic DNA was extracted, sequenced using Illumina technology, and de novo assembled. By employing average nucleotide identity (ANI) analysis, we confirmed the genomic relatedness between strains. This enabled us to discern four distinct clusters belonging to different genera: Pseudomonas, Pseudoalteromonas, Shewanella and Psychrobacter. Genome annotation and comparative analyses revealed insights into the functional potential of these strains, including the presence of secondary metabolite biosynthetic gene clusters and antibiotic resistance genes. In addition, phylogenomic analyses provided a framework for understanding the evolutionary context. Whereas assessment of cold-shock protein presence shed light on adaptation mechanisms. Therefore, our findings underscore the importance of the culture collection as a resource for understanding Antarctic microbial life and its biotechnological potential. The genomic data open new horizons for understanding bacterial life in Antarctica and offer practical solutions for preserving this delicate ecosystem. This research was funded by the Italian National Program of research in Antarctica "Programma Nazionale di Ricerca in Antartide (PNRA)", grant number PNRA18_00075.
Genome size variations and putative length-variable sex chromosomes within surface, subterranean and F1 hybrid populations of Asellus aquaticus isopods.
Mohammed M. Tawfeeq
Mohammed M. Tawfeeq, Meredith Protas, Serban M. Sarbu, Jean-François Flot
California State University (USA), Dominican University of California (USA), Université Libre de Bruxelles (Belgium), “Emil Racoviță” Institute of Speleology (Romania)

Genome size is a fundamental biological trait that varies a lot among eukaryotic species. However, research on its intraspecific diversity remains limited. Previous studies comparing surface and subterranean isopod species found that cave species had larger genomes than surface ones (Lefébure et al. 2017), hence we wondered whether the same pattern would be observed within species. To fill this gap, we examined intraspecific variations in genome sizes in Asellus aquaticus, a freshwater-isopod crustacean with surface and cave-dwelling populations. A total of 39 specimens were collected from surface and subsurface populations in Slovenia and Romania, as well as seven individuals from a hybrid F1 population obtained by crossing hypogean males with epigean females (both from Romania). Using a new, refined Feulgen Image Analysis Densitometry protocol, we generated precise genome size estimates with ranges 0.979-1.98pg (surface), 1.98-2.86pg (cave) and 1.75-2.91pg (F1) for Romania vs. 1.95-2.43 (surface) and 2.12-4.11 (cave) for Slovenia. Intraspecific 1.25 to 2-fold genome size variations were detected in every population. For surface populations, variation was similar among males as among females, but for cave populations the variation was 1.7-2 times larger in males than in females, suggesting strong variations in Y chromosome length in cave populations. Cave populations also had larger average genome sizes than surface ones, with the F1 population having a genome size intermediary between those of its parent populations. The observed variations raise the question of the underlying genomic features, which future whole-genome sequencing will uncover.
On mutational neighbourhoods and their influence on evolutionary processes
Paco Majic Bergara

Paco Majic
EMBL (Germany)

The inheritance of genetic information and the adaptive consequences of variation in that information are concepts that are at the heart of Evolutionary Biology. However, another concept that is typically overlooked is the inheritance of the mutational potential that is associated to each genetic sequence. During reproduction, what is passed on to the descent is not only the genotype that informs phenotypic development, but also that genotype’s mutational neighborhood - that is, the likelihood that each genotype has of mutating into each of every other genotype. I will discuss two ways in which such hidden mutational potential can impact the evolutionary trajectories of populations and entire lineages. Firstly, at a microevolutionary scale, I will focus on how the inheritance of mutational neighborhoods can bias selection and facilitate adaptation by affecting phenotypic development. Secondly, at a macroevolutionary scale, I will describe how mutational neighbourhoods might explain patterns of phylogenetic diversification that have traditionally been ascribed to other evolutionary forces. Summoning cases from developmental biology and palaeontology, my discussion will orbit the framework of genotype spaces, genotype-phenotype maps and adaptive landscapes. Overall, the view of evolution I will present is a view that does not consider species as standalone products of genomic information, but which considers the evolutionary fate of species to be influenced by the material and historical context of that information.
A Comparative Multi-Omic Analysis of Plethodontid Salamander Mucus

Paul Anthony Nicolosi

Paul A Nicolosi
The Ohio State University (USA)

Amphibian cutaneous secretions are critical for diverse and rapidly evolving physiological functions, such as territorial signaling, courtship, and immune responses. As their mucus mediates both interspecific and intraspecific interactions, its utility as a general mechanism of communication mechanism suggests potential coevolution of its components to those of other species. Despite the large body of work demonstrating the diverse functionality of amphibian cutaneous secretions, little is known about their biochemical composition. In this study, we combine proteomic and transcriptomic methods to build high quality “mucosomes” for two lungless salamander species (family Plethodontidae), Plethodon shermani and Desmognathus ocoee. Plethodon salamanders, known as slimy salamanders, produce substantial amounts of mucus relative to their size. The quantity of mucus along with past works demonstrating its involvement in communication makes them ideal models for studying the composition of cutaneous secretions. The distantly related D. ocoee is found in the same habitats and produce similarly complex but biochemically distinct mucus. Using high performance liquid chromatography (HPLC) in tandem with mass spectronomy (MS), we characterized the protein contents of these excretions, with transcriptomic analyses of skin samples collected from matched animals used to construct search databases for proteomics. Our results demonstrate the diverse functionality of Plethodontid mucus, illuminating aspects of their biology. The use of cutaneous secretions as a system of information exchange between salamanders and their environment provides a starting point for understanding how these interactions have shaped their life histories.
Maize and wild relatives show distinct patterns of genome downsizing following polyploidy

Samantha J Snodgrass

Samantha J Snodgrass, Margaret R. Woodhouse, Arun S. Seetharam, Michelle C. Stitzer, Matthew B. Hufford
Cornell University (USA), Department of Ecology, Department of Ecology Evolution and Organismal Biology, Department of Genetics, Development, Evolution, Institute for Genomic Diversity, Iowa State University (USA), UC Davis (USA), USDA-ARS Corn Insects and Crop Genetics Research Unit (USA), and Cellular Biology, and Organismal Biology

Polyploidy and fractionation—the processes by which genomes grow and shrink in size—characterize plant genome evolution. Work on single ancient polyploid species and species descended from separate whole genome duplication (WGD) events has shown fractionation happens quickly and likely before subsequent speciation events. New, chromosome level assemblies for most members of the Tripsacinae subtribe represent multiple descendant lineages from the same WGD event ~5-12 MYA. The Tripsacinae includes the wild genera Tripsacum, Zea, and the culturally and economically important domesticate maize, and spans diverse geographical and ecological ranges centered in Central and North America. We describe shared and segregating fractionation at the exon level across 35 Tripsacinae genomes. Tripsacum (haploid chromosome number n=18) shows less fractionation than any Zea species (haploid chromosome number n=10), with more exons retained in pairs. Using parsimony to estimate the relative timing of fractionation, most exons from either ancestral subgenome (~80%) are either completely fractionated or completely retained across all genomes, suggesting the majority of fractionation occurred soon after the inciting WGD event in the ancestral lineage. However, ~5-6% of exons fractionated after divergence of the genera, but before later speciation events. The higher amount of fractionation observed in Zea may be connected to extensive chromosomal rearrangements observed in this genus while Tripsacum has remained relatively stable. While most fractionation patterns of exon pairs are shared across all genomes (~65.8%), thousands of exon pairs segregate for fractionation pattern. These results suggest that fractionation is a much more dynamic and ongoing process than previously hypothesized.
All living cells are continuously challenged by exogenous, endogenous, and spontaneous DNA damage. While conserved DNA damage response keeps organisms safe broadly, natural variants in these mechanisms can carry profound consequences for cell biology, health, and evolutionary genomics. In this study we use wild nematode strains to study the DNA repair and mutagen tolerance variants that exist in natural populations, as well as the cell biological and mutagenic consequences of those variants. We have developed a highly sensitive assay to identify which wild strains are naturally more or less tolerant to different DNA damaging agents during chronic exposure over multiple generations. We have applied this assay to 35 strains across two species: C. elegans strains that were previously collected from around the world, and O. tipulae that we collected from the Chernobyl Exclusion Zone, a radioactive landscape that we hypothesized may have already selected for mutagen tolerance. We have tested these strains for their diverging sensitivities to a range of DNA damaging chemicals, and have identified phenotypic outlier strains. We have performed bulk segregant analyses with hybrid populations of these outlier strains, identifying alleles of interest for mutagen tolerance phenotypes.
The Genome of the Blueberry Stem Gall Wasp
Stephanie O Castro-Marquez

Stephanie O Castro-Marquez, Glen R Hood, Mike Sergeant, Carlos A Farkas-Pool, Walter Wolfsberger, Ellen O Martinson, Taras K Oleksyk
Oakland University (USA), Universidad Católica de la Santísima Concepción (Chile), University of New Mexico (USA), Wayne State University (USA)

The blueberry stem-gall wasp (Hemadas nubilipennis) is a small insect pest native to eastern North America that infests highbush and lowbush blueberry cultivation. This infestation significantly impacts the yield and quality of both wild and predominantly commercial crops. Although multiple studies have been published regarding managing and characterizing of this species and the gall, the exact mechanism for the gall formation has yet to be discovered. The first genome size estimate for this species (~547.91 Mb) was established from the short-read data using GenomeScope. The quality of our assembly was underscored by a BUSCO completeness score of 95% and average coverage of 96.49x, with sequence distributed across 20,813 contigs with an N50 of 40.8 Kbps. This study is the first look at the efforts towards the complete genome assembly and annotation of this impactful agricultural species. The blueberry stem-gall wasp genome’s high-quality assembly will provide the foundation for gaining new insight into gall formation mechanisms, co-evolution, host-specific behaviors, and genetic markers for targeted pest control management strategies.
DNA replication initiators have rapidly evolving chromatin-binding elements for adaptive specification of replication start sites

Tasia Bos

Tasia Bos, Matthew W Parker
The University of Texas Southwestern Medical Center (USA)

The first step of eukaryotic DNA replication is the so-called "licensing" step where the Mcm2-7 helicase is loaded onto chromosomal loci known as origins. Three proteins are required for licensing, including the origin recognition complex (ORC), Cdt1, and Cdc6. These licensing factors are conserved across eukaryotes. ORC is composed of six subunits, Orc1-Orc6, and is responsible for recognizing and binding the origin sites where Mcm2-7 will be loaded. The Orc1 subunit is primarily responsible for origin selection in metazoans and, consistently, possesses two chromatin-tethering appendages in addition to its core AAA+ domain: an intrinsically disordered region (IDR) that binds DNA non-specifically and a bromo-adjacent homology (BAH) domain that binds a specific epigenetic mark. To determine whether origins of replication are conserved, we have assessed the molecular evolution of ORC and its chromatin-binding elements across the metazoan lineage. Despite its central role in origin binding, we find that Orc1 is the fastest evolving ORC subunit. In fact, the chromatin binding elements themselves (i.e., IDR and BAH) are the most rapidly evolving domains with alterations that suggest dramatic changes in origin specificity. Ongoing work is investigating how these changes alter the specificity of DNA and chromatin binding in vitro and in a cellular setting. Altogether, these studies demonstrate that ORC’s essential chromatin binding elements are rapidly evolving. Our future work will investigate whether the observed plasticity is essential to adapt this fundamental cellular process to the varied cellular and environmental conditions of metazoan organisms.
Novel topological units characterise a distinct regulatory architecture underlying the evolution of complex traits in coleiod cephalopods

Thea F. Rogers

Thea F. Rogers, Oleg Simakov
University of Vienna (Austria)

Our understanding of how genomic changes translate into organismal novelties is often confounded by complex and multi-layered genome architecture. Coleoid cephalopods (squid, octopus and cuttlefish) have large, uniquely structured brains, as well as many species-specific innovations, such as camouflaging ability, complex behaviours and novel organs. These have recently been identified to be associated with an interplay of various of genomic mechanisms, including large-scale genome reorganisation in the ancestor of all coleoid cephalopods, repetitive element accumulation and tandem gene duplications resulting in significant genome expansion. However, regulatory aspects of this genomic architecture remain understudied. We use Micro-C, ATAC-seq and transcriptomic data from several tissues and developmental stages in coleoid cephalopods to assess how a distinct evolutionary history has impacted gene regulation and led to the emergence of cephalopod innovations. We classify locally conserved interacting gene pairs as coleoid-specific ‘spatiosyntenies’ and uncover distinct patterns of regulatory features associated with their emergence, including chromatin remodelling, alternative splicing, RNA editing, noncoding RNAs and repetitive elements. We find that these interactions often occupy pronounced ‘bi-TAD’ arrangements in coleoid genomes, and suggest that relaxed regulatory constraints on cephalopod genomic regions led them to undergo rearrangement and expansion, resulting in their topological linkage and such novel regulation. Taken together, our results describe a distinct ‘unit’ of regulatory evolution for coleiod cephalopods.
Gene expansion and multiple combinations confer retromer functional diversity in traffic in Entamoeba histolytica

Tomoyoshi Nozaki

Natsuki Watanabe, Kumiko Nakada-Tsukui, Tomoyoshi Nozaki
National Institute of Infectious Diseases (Japan), The University of Tokyo (Japan)

The protease transport is essential for various cellular functions such as protein degradation, maturation, activation, and extracellular matrix remodeling, migration, and intercellular communication. It is elaborately regulated in both antero- and retrograde directions, the latter of which is mainly mediated by the retromer complex. The retromer core complex, cargo-recognition complex (CRC), is composed of Vps26, Vps29, and Vps35, and accessory proteins including sorting nexin and Rab7, which interact with CRC to control its localization and function. It has been demonstrated that the retromer regulates recycling of plasma membrane- and endosome-localized receptors. The human intestinal protozoan parasite Entamoeba histolytica causes amoebic dysentery and its parasitism and pathogenesis largely depend upon proteases and their transport mechanisms. In this study, we investigated gene repertoire expansion and functions of highly diversified heterogeneous retromer CRCs in E. histolytica. Immunoprecipitation of individual Vps26 and Vps35 isotypes led to identification of multiple heterogeneous CRCs, which form multimeric super complex. Vps35-1 and Vps35-2, two predominant Vps35 isotypes, are in the multiple forms of CRCs, selectively interact with specific effectors, and are engaged with distinct functions. Vps35-1 controls the formation of super complex and recruitment of CRC to the trogosome, while Vps35-2 regulates CP transport and secretion. This study provides that unprecedented diversification of the repertoire of CRC components and heterogeneous combination and oligomer formation of CRCs allow functional diversity in membrane traffic.
Biochemical interactions in sperm-egg fusion can impose strong selection on functionally important residues between binding proteins. Sperm lysin and egg VERL in the marine mollusk abalone are a classic model of sexual coevolution whose species-specific interactions restrict hybridization of natural populations. VERL is both glycosylated and disulfide-bonded such that recombination expression requires a eukaryotic host. The methylotrophic yeast Komagataella pastoris – better known as Pichia – is an extremely cost-effective and efficient expression system for such proteins, where available genetically engineered strains allow for precise control over post-translational modifications. However, evolutionary studies using high throughput mutagenesis screens have been limited in this system due to a lack of efficient cloning methods. By combining a suite of recently characterized genetic tools, we have developed a new yeast display vector to explore the genetics and biophysics of lysin-VERL coevolution. Only recently were universal yeast origins of replication discovered to enable stable plasmid maintenance in Pichia for efficient transformation of many mutants. The expression cassette includes a GPI-anchor for cell wall attachment with engineered fluorescence and protease susceptibility for multiple layers of surface control. Pichia expression allows us to display mutagenized libraries of proteins with controlled PTMs and screen binding with simple fluorescence measurements. Lysin and VERL coevolve rapidly such that many sites vary between the orthologs of any two abalone species. Our Pichia display system is a robust and accessible tool for understanding the structural and functional consequences of mutations in this complex evolutionary system.
Untargeted nanopore-based genomics enable phylogenetic analysis of highly repetitive genes required for species-specific fertilization in marine abalone.

Trevor Gregory Stepanyan

Trevor G Stepanyan, Damien B Wilburn
The Ohio State University (USA)

Repetitive sequences are common throughout animal genomes but are notoriously difficult to uniquely assign to loci through short-read sequencing. This creates challenges for phylogenomic analyses between different species. Marine abalone are gastropods found along the Pacific coast of the United States where 7 species exist with overlapping habitat ranges. Natural hybrids are rare despite external fertilization with common breeding seasons and lab-bred hybrids being viable. Abalone eggs contain a glycoprotein known as VERL that comprises ~30% of their egg extracellular matrices. Interactions between egg protein VERL and sperm protein lysin are critical for species-specific fertilization in nature. VERL is a highly repetitive molecule with many tandem repeats of a lysin-binding domain which experience concerted evolution. This repetitive domain structure has created enormous experimental challenges for enriching and sequencing the locus via traditional molecular biological approaches. Here we demonstrate that shotgun whole-genome nanopore sequencing with ultra-high molecular weight genomic DNA provides an untargeted, cost-effective, and bioinformatically simple strategy for performing phylogenetic analyses of complex, difficult-to-sequence loci. Optimizations were performed for DNA extraction methods, tissue type, and library conditions to produce >10X coverage of multiple abalone genomes, including previously un-sequenced genomes, with median read lengths >60 kb. Phylogenetic comparisons after extracting full-length VERL using simple BLAST searches has enabled identifying signatures of molecular evolution in a critical fertilization protein. This method provides a roadmap for affordable characterization of formerly difficult-to-sequence genomic regions within evolutionary critical non-model systems.
Comparative study of population TSS maps in yeast unravels genetic basis underlying divergence of transcription initiation
Zhenguo Lin
Zhenguo Lin, Zhenbin Hu, Junhao Chen, Yu Zhan, William Park
Saint Louis University (USA)

Genetically mediated changes in gene expression play a significant role in phenotypic diversification and adaptation. As the first step of gene expression, transcription initiation integrates regulatory signals, determining transcription start sites (TSS) and transcript abundance. TSS turnovers during evolution often change gene expression and gene structure, and dysregulation of TSS usage is associated with many human diseases. Yet, its genetic basis is poorly understood. The great genetic diversity and rich population genomic data of the budding yeast Saccharomyces cerevisiae offer an ideal system for such studies. Here, we generated and characterized quantitative TSS maps for 24 S. cerevisiae strains by CAGE sequencing. Comparative studies of these TSS maps reveal a highly dynamic transcription initiation landscape among yeast populations, outpacing their sequence substitution rates. Integrative interrogations of these TSS maps and SNP data showed that transversion substitutions near a TSS (i.e., -8, -1, and +1 sites) profoundly changed TSS usage within a core promoter, but had limited impact on its overall expression level. Major changes in TSS locations and activities are linked to mutations in upstream binding sites of gene-specific transcription factors and trans-acting factors. Comparative studies of allele-specific TSS usage in an artificial hybrid strain with its parental strains support the major role of trans-acting factors in TSS turnovers. Our study provides a detailed picture depicting the evolutionary dynamic of transcription initiation landscape and its genetic basis, contributing to a better understanding of the genetic mechanisms underlying the diversification of gene expression.
Comparative genome microsynteny illuminates the fast evolution of nuclear mitochondrial segments (NUMTs) in mammals

Zixia Huang

Zixia Huang, Marek Uvizi, Sebastien Puechmaille, Sarahjane Power, Martin Pippel, Samuel Carthy, Wilfried Haerty, Eugene Myers, Emma Teeling
Charles University (Czech Republic), Earlham Institute (United Kingdom), Max Planck Institute of Molecular Cell Biology and Genetics (Germany), University College Dublin (Ireland), University of Montpellier (France)

The escape of DNA from mitochondria into the nuclear genome (nuclear mitochondrial DNA, NUMT) is an ongoing process. Although pervasively observed in eukaryotic genomes, their evolutionary trajectories in a mammal-wide context are poorly understood. The main challenge lies in the orthology assignment of NUMTs across species due to their fast evolution and chromosomal rearrangements over the past 200 million years. To address this issue, we systematically investigated the characteristics of NUMT insertions in 45 mammalian genomes and established a novel, synteny-based method to accurately predict orthologous NUMTs and ascertain their evolution across mammals. With a series of comparative analyses across taxa, we revealed that NUMTs may originate from nonrandom regions in mtDNA, are likely found in transposon-rich and intergenic regions, and unlikely code for functional proteins. Using our synteny-based approach, we leveraged 630 pairwise comparisons of genome-wide microsynteny and predicted the NUMT orthology relationships across 36 mammals. With the phylogenetic patterns of NUMT presence-and-absence across taxa, we constructed the ancestral state of NUMTs given the mammal tree using a coalescent method. We found support on the ancestral node of Fereuungulata within Laurasiatheria, whose subordinal relationships are still controversial. This study broadens our knowledge on NUMT insertion and evolution in mammalian genomes and highlights the merit of NUMTs as alternative genetic markers in phylogenetic inference.
S25 - Mitochondria: from powerhouse to processor and from marker to meaning.
Patterns of gene loss in endosymbionts at the origin of eukaryotes
Aidan Pierce

Aidan Pierce, Nick Lane, Andrew Pomiankowski
University College London (United Kingdom)

Understanding the formation of the eukaryotic cell remains a major challenge for evolutionary biology. A central question relates to the transition of endosymbiont to organelle and why modern cases of endosymbiosis don't follow the same patterns of gene loss and genome size reduction as seen in mitochondria. Using computer simulations we provide insights into patterns of endosymbiont gene loss in the context of multi-level selection (on the eukaryote itself and competition between its constituent organelles). In doing so, we identify how endosymbiont population size is a key factor that can determine the types of genes retained by endosymbionts, and highlight how epistasis within and between endosymbionts impacts gene retention. These results provide a context for organelle gene loss at the origin of eukaryotes.
Clues on the biogenesis of mitochondrial short non-coding RNAs from in silico and in vitro approaches

Alessandro Formaggioni

Alessandro Formaggioni, Federico Plazzi, Marco Passamonti
University di Bologna (Italy), University of Bologna (Italy)

The network of regulatory systems that has arisen from the genomic interactions between the nuclear and mitochondrial genomes is, for the most part, poorly known. A new class of short non-coding RNA was recently discovered in four phylogenetically distant species (i.e., Mus musculus, Drosophila melanogaster, Caenorhabditis elegans and Ruditapes philippinarum). These small RNAs are transcribed in the mitochondrial genome and, according to in silico and in vivo investigations, appear to target nuclear transcripts. However, how these Small Mitochondrial Highly Transcribed RNAs (smithRNAs) can regulate different cellular processes is not clear yet. In the present work we identified a possible smithRNA maturation pathway by analyzing which proteins interact with this class of small RNAs. Analyzing publicly available RNA immunoprecipitation (RIP) and cross-linking immunoprecipitation (CLIP) libraries revealed that in Homo sapiens, proteins involved in miRNA maturation (Ago2, Drosha, and DGCR8) interact with mitochondrial tRNAs, where the majority of smithRNAs are located. Furthermore, the protein-mitochondrial tRNA interaction was also observed in D. melanogaster and M. musculus Ago2, and C. elegans ERGO1, all of which are members of the Argonaute family and typically interact with nuclear short non-coding RNAs. According to our results, it is likely that proteins involved in the maturation of nuclear small RNAs have been coopted for the smithRNA maturation pathway. Moreover, we conducted in vitro experiments in which biotinylated smithRNA sequences were immunoprecipitated alongside the interacting proteins. These interacting proteins were identified and characterized through mass spectrometry, aiming to depict a possible pathway of smithRNA biogenesis.
Random mutations display biases in their relative frequencies with important consequences for genome structure and composition. While laboratory studies have provided insights into the spectrum of spontaneous mutations, the laboratory environment of optimal growth conditions may produce its own biases. Hence, further investigation is necessary to determine if laboratory results are truly representative of natural populations. We analyzed the mitochondrial (mtDNA) genomes of 1,524 Caenorhabditis elegans natural isolates comprising 550 unique mtDNA haplotypes to investigate patterns of mtDNA evolution in the wild. Ancestral reconstruction was used to polarize 3,321 mutations (92 indels, 3,229 SNPs) and the results were compared to mutations identified in experimentally evolved laboratory populations under minimized selection. The frequency of mutations is associated with both codon position and site degeneracy in a manner consistent with strong purifying selection on mtDNA protein-coding sequences. There is significant variation in the per site nonsynonymous and synonymous substitutions between genes and nd1, nd2, and nd4 in ETC complex I are particularly enriched in nonsynonymous substitutions. Whereas laboratory evolved lines exhibit a uniform distribution of heteroplasmic frequencies of mtDNA variants, natural isolates possess a bimodal distribution comprising (i) a majority of variants having reached fixation and (ii) possibly deleterious heteroplasmic variants in lower frequencies. Although some of the patterns of mtDNA mutational bias are similar between the laboratory and wild populations, we also observe significant differences. In particular, transversions typically associated with oxidative damage are less common at four-fold degenerate sites in natural populations than in the laboratory.
The genetic and phenotypic correlates of mtDNA copy number in the Mexican Biobank

Amara Shaukat

Amara Shaukat, Maria Jose Palma, Lourdes Garcia Garcia, Andrés Moreno Estrada, Arslan Zaidi, Mashaal Sohail
CINVESTAV (Mexico), Centro de Ciencias Genomics, Department of Genetics, Institute of Health Informatics, Instituto Nacional de Salud Pública (Mexico), Laboratorio Nacional de Genómica para la Biodiversidad (UGA-LANGEBIO), UNAM (Mexico), University of Minnesota (USA), and Development, cell biology

Mitochondria are important components of the cell and mitochondrial dysfunction can be involved in complex disease through alteration in mitochondrial copy number. We analyzed ~4400 individuals from the Mexican Biobank (MXB) to investigate the genetic and phenotypic correlates of mtDNA copy number (mtCN). We computed mtCN using array data and investigated the role of blood cell type variation in mtCN variation. We computed polygenic scores for 11 blood cell types using trans-ethnic GWAS summary statistics. Cumulatively, they explained a small fraction of the mtCN variation (adj R² = 0.0029) with only Eosinophil (WB-cell type) significantly associated with mtCN, which we used as a covariate for all subsequent analyses. We performed a genome-wide association study on mtCN and identified three variants rs533353962, rs372129830 and rs867763743 significantly associated with mtCN, which are low-frequency variants in MXB (MAF ~0.01) and are rare or absent elsewhere in the world. These fall in SNAPIN shown to regulate mitochondrial homeostasis at synapses, INF2 shown to be involved in stimulating mitochondrial division through mechanisms leading to mitochondrial membrane constriction, and PKD1, mutations in which have been implicated in polycystic kidney disease likely through mitochondrial abnormalities. Further, we performed a pheWAS for mtCN with 20 complex traits and observed mtCN to be significantly associated with Creatinine, Triglycerides, Cholesterol, HDL and LDL levels, Diastolic blood pressure and Rheumatoid arthritis (FDR < 0.05). Overall, these results indicate a significant role of the mitochondrial genome through copy number influencing various complex diseases in a population-specific manner.
Investigating the impact of naturally occurring heteroplasmy on mitochondrial functions in gills of the blue mussel (Mytilus edulis)

Ariane Pouliot-Drouin

Ariane Pouliot-Drouin, Stefano Bettinazzi, Sophie Breton
University College London (United Kingdom), Université de Montréal (Canada)

The process of oxidative phosphorylation (OXPHOS) in mitochondria is essential for energy production, relying on genetic contributions from both the mitochondrial and nuclear genomes. Mitochondrial DNA, usually maternally inherited, undergoes strong selection pressures to maintain homoplasmy and efficient coevolution, since mitonuclear incompatibilities due to mitochondrial heteroplasmy can severely affect bioenergetic efficiency and predispose cells to diseases. However, natural heteroplasmy has been reported in about a hundred bivalve mollusk species where paternal mitochondrial DNA (M-mtDNA) is passed to male descendants and is exclusively kept in male gametes. While most tissues strictly contain maternal mitochondrial DNA (F-mtDNA), occasional presence of M-mtDNA has been observed in somatic tissues of males and, on rarer occasions, in females. This research project aimed to explore the impact of this heteroplasmy on mitochondrial functions in gills of the blue mussel Mytilus edulis by studying enzyme activities associated with energy production. Heteroplasmy levels were assessed via qPCR in both male and female gill samples and heteroplasmy was found to be much more frequent than anticipated. Spectrophotometric analysis evaluated the activity of key mitochondrial enzymes coded by the mitochondrial and nuclear genomes. A negative response to heteroplasmy was observed for all enzymes tested but only in females. Once normalized by mitochondrial content, trends of higher activities following the increase in heteroplasmy were observed. Silenced mitochondria containing M-mtDNA in females could explain our results as well as a general downregulation of OXPHOS and Krebs cycle. Further experiences will confirm theses hypotheses.
Mitonuclear coevolution is a phenomenon involving mitochondrial and nuclear genomes where changes in one of the genomes drive complementary changes in the other genome. This preserves functionality of oxidative phosphorylation as the nuclear genome encodes most of the proteins involved in these pathways and must work in conjunction with proteins encoded by the faster evolving mitochondrial genome. With a wide range of lifestyles of varying energetic needs, birds provide an interesting group in which to examine how mitonuclear coevolution plays out. More specifically, here we aimed to determine how evolutionary rate correlations (ERCs) between nuclear and mitochondrial genes vary in birds with different means of locomotion and metabolic needs, i.e., in different metabolic backgrounds. To do so, we gathered the nuclear and mitochondrial genomes of 105 birds with representatives from every order, making sure to include birds with terrestrial, aquatic, and flighted lifestyles. We analyzed these genomes using a combination of Orthofinder and ERCnet to obtain ERCs for every gene-by-gene comparison within each bird. We will present on patterns that may come up in our ERC comparisons and discuss them in relation to differences in lifestyle. To clarify, we will present on the general hypothesis that ERCs should be strongest between mitochondrial genes and nuclear genes encoding products with mitochondrial functions. We will also examine whether this pattern is stronger in birds with more energetically demanding lifestyles, typically meaning flighted lifestyles.
Mitonuclear genotypes as integrators of epistasis and GxE for complex traits in Drosophila

David M Rand

Proper cellular function requires coordinated expression of the 37 mitochondrial-encoded and >1200 nuclear-encoded genes that have been interacting for more than 1 billion years. Disruption of these complex interactions can cause a variety of diseases and influence adaptive evolution in heterogeneous environments. To advance our understanding of how these complex genetic interactions influence organismal fitness we have constructed a panel of 90 mitonuclear genotypes built from 22 different mtDNAs (10 Zimbabwe, 10 Beijing, from D. melanogaster plus D. simulans and D. yakuba mtDNAs) placed onto two nuclear chromosomal backgrounds (OreR and DGRP375) in duplicate. Each genotype was cultured on standard Drosophila diet and a diet containing 25 mM rotenone, a natural pesticide that inhibits the activity of mitochondrial complex I, NADH dehydrogenase. Climbing speed, flight ability, development time and body weight were measured in adult males and females from each diet. This design allows for partitioning of mtDNA effects, nuclear genotype effects, diet environment, and their interactions in models that test for GxGxE effects. There were strong mitonuclear epistatic effects across the phenotypes, with further modification of epistasis by the rotenone treatment. The mtDNA, nuclear and environmental effects were different for each phenotype. The outgroup mtDNAs of D. simulans and D. yakuba had climbing and flight performances comparable to D. melanogaster mtDNAs despite 100s of substitutions in these mtDNA genomes. Overall, the analyses identify specific mtDNA and nuclear genome pairs that show sensitivity or resistance to rotenone, amenable to further genetic mapping of epistatic and environmental interactions.
Evolution of mitochondrial mRNA editing and linear multipartite mt-genome architecture in calcaronean sponges

Dennis Lavrov

Presented by self
Iowa State University (USA)

Transcripts of mitochondrial genes in most calcaronean sponges undergo an unusual process of insertional mRNA editing in which 1-3 uridylates are inserted at specific sequence motifs. Although, phenotypically, the editing found in calcaronean sponges is similar to that in some other eukaryotes, its mechanism appears to be unique as both the occurrence and the pattern of editing (the number of inserted Us) can be predicted from the primary DNA sequence. In a sense, editing sites in calcaronean mtDNA are “genetic abbreviations” that are expanded to the full sequence at the RNA level. Surprisingly, we found that editing in the group persists despite high rate of mt-sequence evolution and that the editing sites are gained and lost frequently both within and between species even though such events should lead to a frameshift in the coding sequence. Here we provide evidence that evolution of editing sites in calcaronean sponges is enabled by highly unusual mt-genome organization in these species, where most of the genes are located on separate linear chromosomes and multiple copies of each chromosome are transmitted from generation to generation. At the same time, there appear to be selection against nucleotide composition that promotes appearance of new editing sites. Overall, calcaronean sponges provides an interesting model for studying co-evolution between DNA- and RNA-level processes in animal mtDNA.
Evolution of mitochondrial mRNA editing and linear multipartite mt-genome architecture in calcaronean sponges (invited talk)
Dennis V. Lavrov

Dennis V. Lavrov, Noah Gettle
Iowa State University (USA), Wellcome Sanger Institute (United Kingdom)

Transcripts of mitochondrial genes in most calcaronean sponges undergo an unusual process of insertional mRNA editing in which 1-3 uridylates are inserted at specific sequence motifs. Although, phenotypically, the editing found in calcaronean sponges is similar to that in some other eukaryotes, its mechanism appears to be unique as both the occurrence and the pattern of editing (the number of inserted Us) can be predicted from the primary DNA sequence. In a sense, editing sites in calcaronean mtDNA are “genetic abbreviations” that are expanded to the full sequence at the RNA level. Surprisingly, we found that editing in the group persists despite high rate of mt-sequence evolution and that the editing sites are gained and lost frequently both within and between species even though such events should lead to a frameshift in the coding sequence. Here we provide evidence that evolution of editing sites in calcaronean sponges is enabled by highly unusual mt-genome organization in these species, where most of the genes are located on separate linear chromosomes and multiple copies of each chromosome are transmitted from generation to generation. At the same time, there appear to be selection against nucleotide composition that promotes appearance of new editing sites. Overall, calcaronean sponges provides an interesting model for studying co-evolution between DNA- and RNA-level processes in animal mtDNA.
Mitochondrial mutation spectrum in Chordates: damage versus replication signatures, causes, and dynamics

Dmitrii Iliushchenko

Dmitrii Iliushchenko, Bogdan Efimenko, Konstantin Gunbin, Konstantin Popadin

(Russian Federation), Immanuel Kant Baltic Federal University (Russian Federation)

Detecting mutational signatures in diverse biological samples, particularly in oncology, requires the analysis of mutational spectra. It has resulted in the identification of approximately 60 unique mutational signatures linked to various mutagens. However, the mutational processes within mitochondrial DNA (mtDNA) are not as well characterised. This project aims to bridge this knowledge gap by analysing mitochondrial mutagenesis through a comprehensive 192-component mutational spectrum, utilising 86,149 polymorphic synonymous mutations from the mitochondrial CytB gene across 967 chordate species. Our analysis has revealed insightful patterns in mtDNA repair and mutation processes, which can be classified into three predominant signatures: (i) symmetrical mutations distributed uniformly across both DNA strands, primarily induced by gamma DNA polymerase, which account for about 50% of all mutations; (ii) asymmetrical, heavy-strand-specific C to T mutations, accounting for approximately 30%; and (iii) asymmetrical, heavy-strand-specific A to G mutations, comprising about 20%, which appear to be influenced by metabolic rates and age-related factors. It is hypothesised that these asymmetrical signatures result from strand-specific damage in conjunction with suboptimal base excision repair mechanisms on the mtDNA\'s lagging (heavy) strand. Understanding these intricate mutational mechanisms is critical for developing solutions aimed at preventing somatic mtDNA mutations, which could potentially reduce the progression of age-related ailments.

Supported by RSF21-145-20143.
Evolutionary Dynamics of Mitochondrial Sequences Present in Nuclear Telomere-to-Telomere (T2T) Genomes of Apes

Edmundo Torres-González, Saswat K Mohanty, Jessica M Storer, Kateryna D Makova, T2T Primates Consortium
Center for Medical Genomics, Department of Biology, Department of Genome Sciences, Genome Informatics Section, Huck Institutes of the Life Sciences, National Human Genome Research Institute (USA), The Pennsylvania State University (USA), University of Connecticut (USA), University of Washington School of Medicine (USA)

Nuclear sequences of mitochondrial origin (NUMTs) are remnants of the endosymbiotic origin of mitochondria—where mitochondrial DNA (mtDNA) are incorporated into the nuclear genome. Novel NUMT insertions occur 1/10,000 births in humans and are often indistinguishable from mtDNA, confounding studies of low-level heteroplasmy as well as genome assemblies and alignments. Despite the importance of NUMTs, they have been critically under-analyzed due to the lack of complete genome assemblies. To fill this gap, we used telomere-to-telomere (T2T) genomes for seven ape species (bonobo, chimpanzee, human, gorilla, Sumatran and Bornean orangutans, and siamang), to perform a comprehensive analysis of NUMT insertion and evolution. The number of NUMTs increased, in T2T vs. previous assemblies, across all species. The highest NUMT content—total number of bp annotated as NUMTs—was observed in bonobo and chimpanzee, consistent with prior studies, but also in Bornean orangutan. Interestingly, we found significantly higher NUMT content in Bornean orangutan as compared to its sister species, Sumatran orangutan, suggesting different rates of NUMT insertion over 1 MY of divergence. Siamang, a gibbon species, shows a dramatically lower NUMT content compared to great apes. Next, we studied the dynamics of NUMT insertions using multi-species alignments and identified shared and lineage-specific NUMTs. Furthermore, we analyzed the sequence context of NUMTs and their proximity to transposable elements (TEs) to glean potential insertion and propagation mechanisms, respectively. This study provides a complete picture of NUMT evolutionary dynamics in the available T2T ape genomes, which should be augmented by intra-species comparisons.
Physiologists and classical geneticists would likely disagree on the adaptive potential of the small, but important, mitochondrial genome (mtDNA). The long-standing assumption in evolutionary and population genetics that positive selection rarely acts on mitochondrial genomes is, however, being slowly eroded away by increasing volumes of new evidence from model organisms and natural populations alike. In contrast, evidence of local adaptation in human mitochondrial DNA is still debated, with much of the discussion occurring in the early 2000s. Over a decade later, the quality and size of genomic and environmental datasets has dramatically improved, as have our methods to detect signatures of selection within genomes. Consequently, we can now address the question of local adaptation in human mitochondrial genomes with much increased power. Using publicly available genomic and remote sensing climate databases (NCBI assemblies and WorldClim) we have curated a collection of 23,791 complete human mitochondrial genomes from 80 countries paired with associated bioclimatic metadata, including temperature metrics and precipitation levels. This dataset was used in an established generalised linear model approach that identified sites where the distribution of genotypes was better explained by models which included bioclimatic variables. This would be unexpected under neutral evolution. We further inspected identified candidates of local adaptation in the context of mitochondrially relevant stressors (diet, pathogen load, and thermoregulation), and examined changes to protein structure and mitochondrial physiology. Together, our work establishes the role of environmentally mediated natural selection on human mitochondrial evolution, highlighting the contribution of mitochondrial genomes to adaptive process.
SmithHunter: a novel, unified workflow for the identification of candidate smithRNAs and their targets
Giovanni Marturano

Giovanni Marturano, Diego Carli, Claudio Cucini, Antonio Carapelli, Federico Plazzi, Francesco Frati, Marco Passamonti, Francesco Nardi
Department of Biological, Department of Life Sciences, Geological and Environmental Sciences, NBFC, National Biodiversity Future Center (Italy), University of Bologna (Italy), University of Siena (Italy)

SmithRNAs (Small MITochondrial Highly-transcribed RNAs) are a novel class of small RNA molecules that are encoded in the mitochondrial genome and regulate the expression of nuclear transcripts. Initial evidence for their existence came from the Manila clam Ruditapes philippinarum, where they have been described and whose activity has been validated through RNA injection experiments. Current evidence on the existence of these RNAs in other species is scanty and based on small RNA sequencing only. In order to evaluate the potential of this phenomenon as a novel signaling pathway in the broader context of mito-nuclear cross talking, a comprehensive investigation across Metazoa is imperative. We therefore present SmithHunter: a novel workflow specifically designed for smithRNAs. Sequence data (from small RNA sequencing) uniquely mapping to the mitochondrial genome are clustered into putative smithRNAs and pre-filtered based on abundance, presence in replicate libraries and end-conservation. Surviving sequences are subsequently matched to the untranslated regions of nuclear transcripts to identify possible targets. To display SmithHunter structure and functionalities, as well as the effect of parameter customization, the original Manila clam data has been re-analyzed. Furthermore, preliminary results are presented on the presence of smithRNAs in a selection of metazoan species. With this systematic approach we wish to contribute to our understanding of the functional implications of retrograde mitochondrial RNAi across animal lineages. Are smithRNAs an ‘odd feature of an odd system’ or a novel signalling pathway of general interest, with implications as far as the origin of the eukaryotic cell?
The stick insect genus Bacillus and the role of small mitochondrial highly transcribed RNAs (smithRNAs) in hybridization and speciation.

Iuri Icaro

Iuri Icaro, Diego Carli, Alessandro Formaggioni, Federico Plazzi, Marco Passamonti
B.I.G.E.A. Department, University of Bologna (Italy)

The stick insect genus Bacillus displays complex reproductive dynamics, with species such as B. grandii, B. rossius, and B. atticus engaging in various modes of reproduction, including bisexual and facultative parthenogenetic forms. Hybridization among these species has given rise to novel hybrids like B. whitei and B. lynceorum. Notably, mitochondrial DNA analysis has revealed asymmetrical hybridization events, with B. rossius consistently acting as the maternal parent. Recent research focused on a class of small non-coding RNAs called smithRNAs, encoded by the mitochondrial genome and implicated in regulating nuclear gene expression. The distribution and characteristics of smithRNAs were examined across parental Bacillus species. Results indicated a significant presence of smithRNA in B. grandii and B. atticus, while B. rossius, the mitochondrial donor species for hybrids, exhibited minimal quantities of smithRNA. These findings suggest a potential role for smithRNAs in cyto-nuclear interactions and the evolution of isolating mechanisms driving speciation within the Bacillus genus. The study underscores the importance of understanding the molecular mechanisms underlying hybridization and speciation processes, particularly in taxa exhibiting reticulate evolution. Bacillus stick insects serve as an experimental model for exploring the intricate relationships between sexual and clonal taxa, shedding light on the evolution of reproductive strategies and species boundaries.
Do the sex-specific mitochondrial genes play a role in sex determination in DUI bivalves?

Julie Brémaud

Julie Brémaud, Alizée Debelli, Annie Angers, Bernard Angers, Sophie Breton
Département de Sciences Biologiques, Groupement de Recherche Interuniversitaire en Limnologie (Canada), Université de Montréal (Canada)

While mitochondria are maternally inherited in almost all animals, there is a group of about a hundred bivalve species that have a biparental transmission of mitochondria named DUI (Doubly Uniparental Inheritance). In these species, offspring inherit their mitochondria from both parents, but only male embryos will retain the paternal mitochondria in their germ cells. Even more intriguing, the paternal (M-mtDNA) and maternal (F-mtDNA) mitochondrial genomes have additional genes absent in other animal species (orphan genes or ORFans): M-orf (Male-specific ORF) in M-mtDNA and F-orf (Female-specific ORF) in F-mtDNA. It has been hypothesized that the existence and maintenance of such genes could imply a function in sex determination, but this remains to be demonstrated. Immunoblotting and immunocytochemistry methods were performed to determine the tissue and temporal expression patterns of the M-ORF and F-ORF proteins and their sub-cellular localization. Co-immunoprecipitations allowed us to find their interaction partners. Our results show that F-ORF is expressed in all tissues of both sexes in the freshwater mussel Venustaconcha ellipsiformis and marine mussel Mytilus edulis, while M-ORF is only expressed in the gonadal tissues of males sexually mature. The two proteins co-localize with the mitochondria in the gametes (F-ORF in eggs and M-ORF in sperm), and M-ORF is also found in the sperm acrosome. Both proteins possess interaction partners with functions in gamete-gamete interaction. These preliminary results suggest non-conventional functions for mtDNA-encoded proteins.
Decoding the Puzzle of Human Somatic mtDNA Mutagenesis: Bridging Species Life-History Traits and Organ-Specific Cellular Properties
Konstantin Popadin

Konstantin Popadin, Bogdan Efimenko, Alina Mikhailova
Immanuel Kant Baltic Federal University (Russian Federation)

Mitochondrial DNA (mtDNA) mutagenesis in human cancers is still enigmatic: external mutagens like UV light or tobacco smoke, known to impact nuclear DNA, don’t seem to affect mtDNA. This discrepancy suggests a dominant, mitochondria-specific internal mutagen. To understand this, we analyzed mtDNA mutational spectra across a wide range of vertebrates, uncovering that some mutations (A>G on the heavy chain) are linked to internal, age-related chemical damage, likely oxidative (DOI: 10.1093/nar/gkac779). Others (C>T on the heavy chain) stem from internal replication processes (DOI: 10.1101/2023.12.08.570826). Leveraging our insights, we hypothesized that in somatic mtDNA mutagenesis, slow-dividing cells would exhibit more damage-related mutations, while fast-dividing cells would show a predominance of replication-associated mutations, i.e. we posited a variable damage-to-replication ratio (A>G/C>T) across tissues with different replication rates. To test this, we reanalyzed the two most comprehensive human mtDNA datasets: the TCGA, with 7,611 mutations across 37 cancer types and 21 tissues, and the GTEx, with 2,762 mutations in 25 healthy tissues. In both datasets, a higher damage-to-replication ratio was found in organs with slow-dividing cells (neurons, ovary, heart, skeletal muscles) and a lower ratio in those with rapid cell division (colon, intestine, stomach, vagina). Furthermore, a decrease in this ratio in late-stage somatic mutations across all cancer types (TCGA data) aligns with the anticipated increase in cell turnover during carcinogenesis. Our research shows that mtDNA mutations can reveal cell traits, unveiling their potential importance in biomedical applications.
The contribution of within- and between-species mitochondrial variation to adaptation in experimental Drosophila populations

Leah Darwin

Leah Darwin, Yevgeniy Raynes, Faye Lemieux, David Rand
Brown University (USA)

The oxidative phosphorylation system (OXPHOS) relies on both the mitochondrial and nuclear genomes to function. Due to its importance to cellular function, it has been hypothesized that strong purifying selection prevents functional genetic variation from accumulating in the mitochondria. This theory has been challenged by evidence of the contribution of mitochondrial variation to fitness differences. We use an evolve and resequence approach to quantify the role of within-species and between-species mitochondrial variation in adaptation to a novel environment. We constructed 24 outcrossed experimental populations (12 control, 12 experimental) from a founder population of Drosophila melanogaster with introgressed mitochondria (mt) DNA from 24 different strains of D. melanogaster from Zimbabwe or Beijing, 3 strains of D. simulans, and 1 strain of D. yakuba. The design of this study provided a replicate source of extensive nuclear and mtDNA variation. We subjected experimental populations to selection on rotenone, an insecticide that inhibits mitochondrial complex I in the OXPHOS system to specifically target a mitochondrial response. Phenotypic assays were performed on the founder lines and demonstrated fitness differences between mitochondrial haplotypes in the rotenone environment. After 20 generations of selection, pooled sequencing analyses of the evolved populations were used to quantify haplotype frequency dynamics of nuclear and mtDNA loci. We find the evolutionary trajectories to be contingent on the history of the experimental population. We report the mitochondrial and nuclear variants that have contributed to the differentiation between experimental populations.
The alternative mystery of the mitochondrial genome
Ludovic Nadeau-Lachance

Ludovic Nadeau-Lachance, Thierry Choquette, Hajar Hosseini Khorami, Annie Angers, Sophie Breton
Université de Montréal (Canada)

Mitochondria possess a circular genome distinct from the nuclear genome. In animal species, the mitogenome is generally described as being limited to 37 genes, thirteen of which produce proteins involved in energy production. However, this "mitochondrial dogma" has recently fallen following the discovery of genes inside canonical genes in the human mtDNA. Among these recent discoveries, we have demonstrated the existence of a gene in an alternative reading frame inside the mitochondrial protein-coding gene nd4: mitochondrial alternative nd4 or mtaltnd4. We found that the protein MTALTND4 impacts cell and mitochondrial physiology, possibly by interacting with C1QBP, a protein that is known to regulate cellular energy metabolism through modulation of mitochondrial translation or through protein-protein interactions. We followed up on these discoveries related to MTALTND4 by studying its effects on gene expression. Our results suggest that the peptide downregulates the expression of several genes involved in metabolic or biosynthetic processes. The discovery of MTALTND4 sheds light on the intricate nature of the mitochondrial genome and enhances our understanding of mitochondrial biology.
The emerging role of mitochondrially encoded small interfering RNAs to regulate nuclear gene expression.
Marco Passamonti

Marco Passamonti, Diego Carli, Alessandro Formaggioni, Iuri Icaro, Youn Le Cras, Giovanni Marturano, Giorgio Muneretto, Francesco Nardi, Oscar Wallnoefer, Federico Plazzi
Institut Pasteur (France), University of Bologna (Italy), University of Siena (Italy)

Mitochondria seem to be connected to many more cell functions than ATP production. The possibility that mitochondrial DNA (mtDNA) can act on nuclear gene expression has been suggested only recently: mtDNAs have been found to produce small noncoding RNAs (sncRNAs), long non-coding RNAs (lncRNAs) and peptides, all of them suggested or demonstrated to interact via different pathways with the nucleus. Our research aims at characterizing the retrograde mitochondrial-to-nucleus signaling by mitochondrially encoded small RNAs (called small mitochondrial highly transcribed RNAs, smithRNAs), in a comparative way. We have already predicted the presence of many and proved functionality of two smithRNAs in Ruditapes philippinarum, in which they are likely involved in gonad formation and sex determination. While we have developed a bioinformatic pipeline to detect and analyze smithRNAs in metazoans, we present new data on:
1. their phylogenetic distribution among Metazoa; 2. their role in mito-nuclear incompatibilities and speciation; 3. their evolvability in the context of the polycistronic maturation of mtDNA; 4. their possible ways to escape mitochondria to deliver their function in the cytoplasm; and 5. their maturation processes, to understand whether they can be ascribed to the known classes of small interfering RNAs (i.e. miRNA, siRNA or piRNA), or to a new unknown one. Overall, smithRNAs are emerging as a fast-evolving generalized new form of retrograde signaling and, potentially, a common feature among metazoans, adding a brand-new functionality level of mitochondria in the eukaryotic cell.
The Importance of Mito-nuclear Epistasis in Evolutionary Genetics

Michael Garvin

Presented by self
Oak Ridge National Laboratory (USA)

There is great interest in understanding the genetic basis of adaptation, especially in the face of ongoing climate change. We and others have identified positive selection in the mitochondrial genomes of diverse taxa in the subunits that comprise the main proton pump, complex I. We hypothesized that the trait under selection was the bioenergetics of reproduction as it relates to food availability. The emergence of AI and large public data sets provides opportunities to understand these evolutionary processes in greater detail. A re-analysis of our previous study using a structure-based alignment with predicted protein structures from AlphaFold rather than simple sequence alignment identify two hotspots of evolution. One site creates or removes a di-sulfide bond with the nuclear-encoded subunit of mitochondrial complex I, NDUFB8, emphasizing the importance of mito-nuclear interactions. We employed a computationally intensive algorithm called CCC in the program BlocBuster to test for allele haplotypes across the nuclear genome that co-evolve with positively selected sites in the mitochondrial genome and identify regulatory blocs in genes that are directly related to mitochondrial function. These approaches combined with more refined phenotypes will provide greater insight into adaptive and maladaptive mito-nuclear epistatic networks in relation to variable resources as the climate continues to rapidly change.
Evidence of convergent evolution in the nuclear and mitochondrial OXPHOS genes across Squamata deep lineages

Oscar Wallnoefer

Oscar Wallnoefer, Alessandro Formaggioni, Federico Plazzi, Marco Passamonti
University of Bologna (Italy)

The oxidative phosphorylation system (OXPHOS) is composed of five complexes, with subunits encoded by both the mitochondrial genome (mtOXPHOS) and the nuclear genome (nucOXPHOS). mtOXPHOS and nucOXPHOS subunits tightly interact; indeed, signs of coevolution between these two sets of genes have been observed in Metazoa. Squamata is particularly intriguing from a mitochondrial point of view: the phylogenetic signal inferred from mitochondrial markers is radically discordant compared to nuclear markers. According to mitochondrial topology, lizards of the family Agamidae (Acrodonta) are in sister relationship with the crown-group of snakes (Serpentes), rejecting the monophyly of Iguania (Acrodonta + Pleurodonta), which is supported by most phylogenetic studies on Squamata. It has been proposed that the convergence between the early lineages of snakes and agamids is due to a common selective pressure on OXPHOS. In this study, we aim to investigate whether convergent evolution between these two lineages (i.e., Serpentes and Agamidae) is also detectable in nucOXPHOS. We annotated 88 nuclear OXPHOS genes from 56 squamatan genomes encompassing 24 families, along with the 13 mitochondrial OXPHOS genes. Phylogenetic trees were inferred from both datasets, and we evaluated signals of convergent evolution between Agamidae and Serpentes. The evolutionary rate correlation analysis revealed a clear coevolutionary signal between nucOXPHOS and mtOXPHOS, which is stronger when considering only subunits in close contact within the complexes. Moreover, some nucOXPHOS genes resulted under positive selection along the Serpentes and Acrodonta lineages, with clues of convergent substitutions observed between the two.
Does mitonuclear coevolution affect genome-wide signatures of speciation?

Shady A Kuster

Shady A Kuster, Molly Schumer, Justin C Havird, Daniel B Sloan
Colorado State University (USA), Stanford University (USA), University of Texas at Austin (USA)

Coevolution between mitochondrial and nuclear-encoded mitochondrial (n mt) proteins is said to contribute disproportionately to speciation processes through the formation of genetic incompatibilities. During secondary contact, incompatibilities can emerge when mitochondrial haplotypes are placed into cellular contexts without their coevolved n mt proteins. However, the extent to which mitonuclear coevolution affects nuclear genome evolution and speciation is unknown. An increasingly common approach used to identify signatures of mitonuclear coevolution is to divide the genome into gene products that function in mitochondria and those that do not. The efficacy of this approach has not been extensively tested, but a few studies have found evidence of mitonuclear coevolution contributing to reproductive isolation in this fashion. Here, we leverage a swordtail fish (Xiphophorus) species hybrid system in which a lethal mitonuclear genetic incompatibility has been identified to test the hypothesis that large numbers of n-mt loci are involved in mitonuclear incompatibilities. We partition the nuclear genome into loci that form complexes with mitochondrial gene products, loci that are targeted to mitochondria (but not in mitonuclear complexes), and loci that have no known function involving mitochondria. Linkage disequilibrium analyses show no statistical difference between the three groups, despite selection against the mitonuclear incompatibility. We additionally compare our findings to previous analyses in hybrid systems that similarly partition nuclear loci to determine when mitonuclear coevolution produces detectable genome-wide signatures. It may be that these genome-wide scans are insufficient when used on such broad functional groups, and their efficacy may be situational.
Phylonumtomics uncovers diverse evolutionary trajectories of mitogenomic fossils buried in mammalian and avian genomes

Toni Gossmann

Sporadically genetic material that originates from an organelle genome integrates into the nuclear genome. However it is unclear what processes maintain such an integration over longer evolutionary time. Recently it was shown that nuclear DNA of mitochondrial origin (NUMTs) may harbour genes with intact mitochondrial reading frames despite the fact that they are highly divergent to the host's mitochondrial genome. Two major hypotheses have been put forward to explain the existence of such mitocoding nuclear genes: (A) recent introgression from another species and (B) long-term selection. To address whether these intriguing possibilities do play a role we scanned the genomes of more than 1,000 avian and mammalian species for NUMTs. We indeed identified that the subclass of divergent NUMTs harbouring mitogenes with intact reading frames are widespread across mammals and birds. We can show that for these NUMTs signatures of cross-species introgression are widespread in birds, but not mammals with the exception of ungulates. We can also show that a substantial fraction of deeply divergent NUMTs are maintained by selection. For a small number of NUMT genes we identify an evolutionary signature that is consistent with adaptive evolution including one human NUMT that is shared among seven ape species. This highlights the intriguing possibility that NUMT insertions occasionally may be functional.
No passive sidekick: a substantial role for mitochondrial compensatory evolution during adaptation of ETC-deficient strains of Caenorhabditis elegans
Vaishali Katju

Vaishali Katju, Alberto Corral-López, Zachary P. Dietz, Suzanne Estes, Ulfar Bergthorsson
Portland State University (USA), Uppsala University (Sweden)

The intergenomic coadaptation of the mitochondrial (mtDNA) and nuclear (nDNA) genomes is reliant on the dual genetic control of both nDNA- and mtDNA-encoded protein subunits comprising the mitochondrial electron transport chain (ETC). The molecular basis of these mitonuclear epistatic interactions requisite for maintaining key metabolic function remain poorly understood. We conducted a laboratory evolution experiment to investigate the functional and genomic patterns of mitonuclear adaptation in independent populations of Caenorhabditis elegans bearing four focal ETC mutants. Although a large number of both nDNA and mtDNA mutations were identified following 60 generations of laboratory adaptation, mutations in genes with assigned mitochondrial function were restricted to the mtDNA genome. A total of 77 mtDNA variants with intra-individual frequencies ranging from 5-100% were identified across 124 adaptation lines. In direct contrast to the patterns observed in C. elegans mutation accumulation lines under minimal selection, the adapted lines were enriched in base substitutions relative to small indels. The vast majority of base substitutions in protein-coding genes were nonsynonymous changes (82.5%) with significantly higher heteroplasmic frequencies, suggesting potential targets for positive selection. Notably, novel mutations in the same ETC complex as the original background mutation reached significantly higher frequency than mutations in a different complex, suggestive of compensatory evolution. While the nuclear genome has been hypothesized as the primary driver of mitonuclear coevolution given its larger mutational target, our results argue for a disproportionately outsized role for mitochondria in mitonuclear adaptation despite their diminutive genome size and protein-coding capacity.
Originate from an endosymbiotic alphaproteobacterium during the eukaryotic evolution over two billion years ago, mitochondria play an important role in cellular physiology, such as oxidative phosphorylation. Based on comparative genomics studies, it has been inferred that the mitochondrial genome of Last Eukaryotic Common Ancestor (LECA) encoded proteins related to electron transport chain complexes, ribosomes, protein translocation and heme maturation. After LECA, the gene contents in mitochondrial genomes varies much across different eukaryotic taxa. While some cases of mitochondria-to-nucleus transposition were identified, many of these genes remain in the mitochondrial genomes in many lineages. Why there are very mixed outcomes of retention versus reduction in mitochondrial genome evolution remains controversial. To study this question, we ask what kind of population-genetic environment facilitates or constrain the mitochondria-to-nucleus transposition. Specifically, we used SLiM to simulate a population of haploid cells with intercompartmental duplication in mitochondrial and nuclear genomes and tracked which copy is lost earlier. The simulation results show the transposition can be facilitated by beneficial effects. Interestingly, even when transposition is deleterious, it can still be preferred with a much higher mutation rate in the mitochondrial genome compared to the nuclear genome. These findings have implications for our understanding of the mode and tempo of endosymbiont gene transfer.
S26 - Deciphering the functional and adaptive effects of genomic structural variation.
Imperfect plagiarism: the evolutionary fate of laterally acquired structural variants in grasses

Catherine F Collins

Catherine F Collins, Luke T Dunning
University of Sheffield (United Kingdom)

Lateral gene transfer (LGT) is widespread in eukaryotes, including in grasses, wherein gene acquisition is reported in crops and wild relatives alike. These transfers involve the insertion of large multigene DNA fragments, but it is presently unclear whether all the acquired genes are expressed and retained, or if non-beneficial linked genes are purged from the recipients’ genome. We use RNA-seq to determine the proportion of laterally acquired genes that remain expressed. We then compare the relative expression levels of the laterally acquired genes to their orthologous vertically inherited copy in the recipient. Finally we determined the expression level of the laterally acquired genes prior to the transfer event by sequencing the transcriptome of the donor species. We see a range of patterns in relative expression among the genes, including laterally acquired genes that are silenced, expressed more similarly to the copy in the donor, and those that align with the vertical copy in the recipient. These patterns shed light on the fate of a laterally acquired gene, suggesting that neutral and deleterious genes tend to be silenced or pseudogenised. In contrast, advantageous genes are retained and result in subsequent pseudogenisation of the vertical copy. This research provides an insight into the integration of laterally acquired genes and the ripple effect this acquisition causes to the genome.
Spontaneous mutation rate for and strength of purifying selection against structural variants in the C. elegans genome

Charles F Baer

Ayush S Saxena, Charles F Baer
Dept. of Biology (USA), University of Florida

Structural variants (SVs) are the most understudied part of the mutation spectrum, and much less is known about their mutational properties than those of single-nucleotide variants (SNVs) and short indels. We estimated the SV mutation rate from six mutation accumulation (MA) lines of C. elegans using long-read (PacBio) sequencing. We identified 32 SV mutations in four N2 strain lines and 17 in two PB306 lines. We previously identified an average of 56 SNV and 14 short indel mutations per line. The total number of bases affected is ~20kb in N2, with a net gain of ~3kb across 4 lines (15 insertions and 17 deletions). For PB306, ~16kb were affected, with a net gain of 7kb (6 insertions and 11 deletions). Comparison of the SNV/SV ratio in MA lines to that in three natural isolates provides an estimate of the relative strength of selection against SVs, which is ~7X greater for SVs. Selection against SVs in intergenic regions is ~4X greater than against SNPs in exons. The correlation between the pairwise divergence at SNVs and SVs in the wild isolates was >0.99. False positives were identified by visual inspection of putative mutants in IGV, resulting in a signal (inferred mutant) to noise (false positive) ratio in the MA lines of ~0.7. False negative rates were determined by simulating variants in the reference genome, and observing 'recall'. Recall rate ranges from >90% for short (~100-bp) indels and declines with increasing SV length, with deletions having higher recall rates than insertions.
Repetitive sequences drive genomic variation and vary regulatory landscapes

Christine Beck

Presented by self

The Jackson Laboratory of Genomic Medicine and University of Connecticut Health Center (USA)

Repetitive sequences comprise more than half of both human and mouse genomes, and these repeats continue to mobilize and diversify modern DNA. From transposable elements (TEs) to segmental duplications and tandem repeats, these repetitive sequences change at a more rapid rate than other loci, and this change can lead to differences in gene regulation, splicing, and downstream functional consequences. We are using long read DNA sequences of diverse mice and humans to examine variation caused by mobile elements and segmental duplications. Across 64 human haplotypes, we have now identified more than 10,000 TE insertion variants and over 2,000 rearrangements catalyzed by TEs at the junctions. Across fourteen diverse mouse genomes, we have identified over 150,000 TE insertion variants and a preliminary analysis has found more than 5,000 deletions mediated by homologous TEs. Further, with orthogonal optical mapping data, we have begun to characterize segmental duplication-mediated variation across eight of these strains, and have found more than 150 rearrangements mediated by segmental duplications, some leading to deletion or duplication of genes. Moreover, we have paired mouse genomic variation data with chromatin accessibility and RNA sequencing data and have clear instances where insertions and deletions of mobile elements and variation within these sequences can additionally alter chromatin accessibility and splicing across short evolutionary timespans. Overall, our data clearly show that paired genomic, epigenetic and transcriptomic diversity can illuminate how structural variation affects closely related individuals within a species.
The role of structural variation in clownfish adaptive radiation

Diego A. Hartasánchez

Anna Marcionetti, Diego A. Hartasánchez, Nicolas Salamin

University of Lausanne (Switzerland)

Clownfishes are a complex of 28 species that rapidly diversified following the acquisition of mutualism with sea anemones around 15 Mya. While the genomic signatures associated with this adaptive radiation have started to emerge, little attention has been paid to structural genomic variation in clownfish. We know, however, that structural variation can disrupt gene function and regulation and alter gene dosage, and that it has been demonstrated to play a central role in adaptive evolution and species diversification in multiple taxa. Hence, we have investigated the number, type and role of structural variation across the clownfish phylogeny, paying particular attention to chromosome 18, which has recently been suggested to harbor two large inversions. Using long sequencing reads, we obtained high-quality genomes for several clownfish species spanning the main lineages identified in the genus. We found a large number of insertions, deletions, duplications, and inversions that occurred during clownfish diversification. We were able to reconstruct and date their evolutionary history, confirming the presence of large inversions on chromosome 18. We show that hybridization events during clownfish diversification might have contributed to generating the striking pattern of presence/absence of one of these large inversions across the genus, and suggest a possible link between this inversion and clownfish phenotype, in turn associated with anemone choice. This study represents the first evidence of considerable structural variation in clownfish genomes. It also provides new, high-quality, valuable genomic resources for pursuing research on this emerging model system and on its adaptation to sea anemones.
Molecular evolution of a maize hybrid barrier over 12 million years suggests epistatic silencing

Elli Cryan

Elli Cryan, Garnet A Phinney, Arun S Seetharam, Matthew M S Evans, Daniel J Kliebenstein, Jeffrey Ross-Ibarra
Carnegie Institution for Science (USA), Development and Cell Biology (USA), Evolution and Organismal Biology (USA), Iowa State University Ecology, Iowa State University Genetics, UC Davis Evolution & Ecology (USA), UC Davis Plant Sciences (USA)

Maize and its wild relatives teosinte can all readily hybridize, yet they remain distinct. Mating incompatibility loci provide one mechanism that can allow populations to maintain reproductive isolation. In Zea mays, three complex mating incompatibility loci encode genes that disrupt directional pollen tube growth down the silk. When a maternal plant receives pollen with incompatible alleles, fertilization is impeded, creating a prezygotic reproductive barrier. However, this barrier is not complete. Infrequent fertilization of incompatible gametes facilitates introgression of incompatibility genes into other populations. Previous modeling shows that each factor should only undergo brief periods of strong selection, on a timescale shorter than time to speciation in this clade. Against this expectation of transient benefit, these loci are hypothesized to have aided in speciation. All of the loci display presence absence variation and copy number variation. We used public and new genome sequences to reconstruct the evolutionary history of these complex loci. We classified haplotype diversity at all three loci in over 30 Zea mays genomes, identified syntenic loci in related species, constructed gene trees of known and candidate functional genes, and analyzed rates of molecular evolution. We find evidence of potentially functional reproductive barrier loci and genes in lineages estimated to be twelve million years diverged. Over millions of years, these genes have undergone selection and evolved to be distinct barriers, which is reflected in predicted protein structures. We also find evidence of loci driving epigenetic silencing of interacting loci. These loci may have played a role in Zea diversification.
Genomic and Transcriptomic Signatures of Alternative Reproductive Tactics in a Mexican Poeciliid

AC (Mexico), Instituto de Ecología, Ohio University (USA), Stanford University (USA), Texas State University San Marcos (USA), Xiphophorus Genetic Stock Center (USA)

Intense sexual selection can drive the emergence of alternative reproductive tactics (ARTs), or strategies that increase reproductive success through mechanisms other than courtship. The evolutionary history of ARTs is puzzling, as multiple reproductive morphs can exist together within a population for long periods of time, hinting at fitness tradeoffs between reproductive strategies that maintain multiple phenotypes. The swordtail fish, Xiphophorus multilineatus, exhibits two opposing genetically-determined male size morphs. Large males exhibit sexual ornamentation and are behaviorally fixed for courtship while small males lack sexual ornamentation and are plastic for courtship and coercive mating tactics. Variation in the number of Y-linked non-functional mc4r copies has been shown to account for some of the differences between morphs. However, crosses have pointed to the existence of autosomal factors that contribute to some of the variation between morphs as well. Here, we take a whole-genome approach to dissect the variants contributing to differences between morphs. We perform a genome-wide association study (GWAS) for male morph and demonstrate a polygenic basis for male size while capturing the Y-linked mc4r signal demonstrated by previous studies. We generate PacBio assemblies for each morph to identify differences in structural variation and confirm the variation in Y-linked mc4r copy number. Finally, we explore the transcriptomic profiles of relevant tissues between morphs to refine our GWAS results and further explore physiological differences between them. These results provide a comprehensive characterization of a balanced polymorphism under strong selection in natural populations.
Evolution of the peroxiredoxin gene family: an antioxidant enzyme in the context of hypoxia-induced by aquatic mammals dives.

Giovanna Selleghin Veiga

Giovanna Selleghin Veiga, Mariana Nery
UC Berkeley (USA), University of Campinas (Brazil)

Peroxiredoxins are a group of antioxidant enzymes that play a crucial role in eliminating hydrogen peroxide (H2O2) produced during cellular metabolism. In conditions such as tissue ischemia/reperfusion, H2O2 production is heightened. This scenario is recurrent in aquatic mammals, such as cetaceans and pinnipeds, when they perform breath-hold dives that lead to depletion of oxygen stores and cell hypoxia due to prolonged submersion. In such circumstances, peroxiredoxins are essential for regulating metabolism and preventing oxidative damage. Thus, our objective was to investigate evolutionary patterns of the peroxiredoxin gene family among aquatic mammalian species. Accordingly, we identified the number of ortholog and paralog genes among mammals, examining events of gene duplication, patterns of chromosomal distribution, and adaptive selection. We found that most subtypes of peroxiredoxins are characterized by single-copy genes with minimal missing data, with PRDX1, the most highly expressed subtype found in both the nucleus and cytoplasm, exhibiting an increase in gene copies in aquatic mammals, particularly in pinnipeds. The distribution of the copies was dispersed along the genome and the majority were retrocopies of the main coding sequence with little divergence among them. These results suggest that the peroxiredoxin gene family has a role in the evolution of aquatic mammals by reducing oxidative damage intrinsic to diving behavior in these species.
Some of the most striking polymorphisms in nature are seen in the maintenance of multiple mating types. One example in angiosperms is heterodichogamy - dimorphism for the order of male vs. female flowering in hermaphrodites. This mating system is common throughout Juglandaceae, including globally important nut crops - walnuts (Juglans), as well as pecan and other hickories (Carya). In both genera, heterodichogamy is controlled by a single dominant allele. We fine-map the locus in each genus, and find two ancient (>50 Mya) structural variants involving different genes that both segregate as genus-wide trans-species polymorphisms. The Juglans locus maps to a 20 kb structural variant adjacent to gene involved in a flowering time signaling pathway (TPPD-1). Across species, the dominant haplotype contains a tandem array of 8-12 duplicated sequence motifs, each of which contains an inverted copy of the 3′ UTR. In developing male flowers, this acts as a template for the transcription of numerous small RNAs, corresponding with a relative increase in expression of the cis copy of the flowering gene and delayed male flowering of female-first morphs. The Carya locus maps to a cluster of 20 tightly-linked variants across 20 genes, likely acting as a supergene. The dominant haplotype rarely recombines and shows both reduced diversity and increased transposable element accumulation, mirroring patterns seen in heterogametic sex-chromosomes. We did not detect either genetic system in other heterodichogamous genera within Juglandaceae, suggesting that additional genetic systems for heterodichogamy remain undiscovered.
DNA replication errors are a major source of gene amplification in adaptive evolution

Julie Chuong

Julie N Chuong, Nadav Ben Nun, Ina Suresh, Julia Matthews, Titir De, Grace Avecilla, Farah Abdul-Rahman, Yoav Ram, David Gresham
Center for Genomics & Systems Biology (USA), Dept. of Biology (USA), New York University, New York University (USA), Tel Aviv University (Israel)

Copy number variants (CNVs) are an important source of genetic variation underlying rapid adaptation and genome evolution. However, little is known about factors contributing to the structure, formation rate, and fitness effects of CNVs. Local genome elements are likely important determinants. Previously, we found the GAP1 gene in Saccharomyces cerevisiae undergoes repeated amplification and selection under glutamine-limited selection. This gene has a unique genomic architecture consisting of two flanking long terminal repeats (LTRs) and a proximate origin of DNA replication (autonomously replicating sequence (ARS)), which contribute to GAP1 CNV formation. To test the role of these elements on CNV-mediated adaptation, we experimentally evolved in glutamine-limited chemostats strains lacking either the LTRs, ARS, or all elements. Using a CNV reporter and neural network simulation based inference we quantified the rate and fitness effect of CNVs for each genotype. Removal of local DNA elements has minimal impact on the rate or fitness effect of CNVs. We defined the mechanisms of CNV formation for ~40 CNV lineages from each genotype. Regardless of genotype, 49% of CNVs are mediated by the DNA replication based mechanism Origin-Dependent Inverted-Repeat Amplification (ODIRA). In the absence of the local ARS, distal ones can mediate ODIRA CNV formation. Homologous recombination can mediate gene amplification in the absence of LTRs in part initiated by de novo insertion of retrotransposons at the locus. Our study reveals the remarkable plasticity of the genome and demonstrates that DNA replication errors are a predominant source of adaptive CNVs.
The new complete human genome assembly provides new insights in Copy Number Variations across humans
Lucia Bazan Williamson

Lucia Bazan Williamson, Emily Jane McTavish
Merced (USA), University of California

Copy number variations (CNVs) represents a crucial aspect of human genomic diversity, influencing phenotypic variation, disease susceptibility, and population genomics. This study examines CNV patterns across human populations utilizing the latest complete human genome assembly (T2T-CHM13) and contrasts the findings with the previous GRCh37 assembly. Discrepancies were found in CNV frequencies between different assembly, with notable differences in the number of duplications and deletions. The T2T-CHM13 assembly unveils substantial differences in copy numbers, including many novel duplications, particularly in telomeric regions, reflecting improvements in genome assembly quality. Population structure analyses highlighted contrasting patterns of population differentiation between assemblies, impacting interpretations of human evolutionary history. The improved T2T-CHM13 assembly enabled discovery of highly differentiated CNVs, known introgressed regions, and gene candidates previously undetected. Overall, these findings underscore the critical importance of leveraging updated, high-quality genome assemblies for comprehensive CNV landscape characterization and accurate population genetic inferences, while highlighting potential biases introduced by previous assemblies.
Anopheles gambiae and An. coluzzii mosquitoes are major human malaria vectors in Africa, accounting for most of the transmission. While urban environments were until recently considered to be unfit for Anopheles larvae development, these mosquito species have rapidly adapted to polluted habitats, posing challenges for malaria control. Therefore, understanding the genetic factors driving this adaptability is crucial. In this work, we have analyzed 375 An. gambiae and An. coluzzii WGS samples from urban and rural areas in six Central African countries. Taking advantage of recent high-quality long-read assemblies for both species, our analysis focused on identifying genetic variants, from SNPs to structural variants including transposable elements (TEs) and inversions. We have created the first manually curated TE library for An. gambiae, containing 295 consensus sequences and including 53 new TE families. By combining three TE annotation tools, PoPoolation2, TEMP2 and TEFLoN, we identified 5,462 and 4,773 euchromatic TE insertions present at high frequencies in An. gambiae and An. coluzzii populations, respectively, that could be potentially involved in adaptation. We also found that the frequencies of the cosmopolitan 2La and 2Rb inversions, previously linked to ecological adaptation in these species, varied from low in urban locations to high in rural locations (6%-86% and 0%-100% for 2La, 0%-78% and 0%-28% for 2Rb, in An. gambiae and An. coluzzii, respectively). Overall, our work shows that beyond SNPs, structural variants, including TEs and inversions, contribute to the ecological adaptation of malaria vectors.
Gene expansions contributing to human brain evolution
Megan Y Dennis

Daniela C Soto, José M Uribe-Salazar, Gulhan Kaya, Aarthi Sekar, Alfie Gleeson, Nicholas K Haghani, Keiko Hino, Gabriana N La, Natasha A Mariano, Cole Ingamells, Aidan E Baraban, Li-En Jao, Sergi Simó, Gerald Quon, Aida M Andrés, Megan Y Dennis
Davis (USA), University College London (United Kingdom), University of California

Genomic drivers of human-specific traits remain largely undiscovered. Duplicated genes expanded uniquely in the human lineage likely contributed to brain evolution, as demonstrated by SRGAP2C and its association with improved sensory-task performance when introduced in mouse models. Challenges exist in discovering duplicated genes due to the paralog similarity making them prone to sequence-assembly errors. Mitigating this issue, we used a complete telomere-to-telomere human genome sequence to identify 491 duplicated gene families comprising likely human-specific paralogs (>98% identity). Using genome sequence and fetal-brain transcriptomic datasets, we narrowed in on 741 brain-expressed paralogs universal across modern humans. This included 42 paralogs co-expressed in modules enriched for autism-associated genes, connecting them with human language and cognition. Further study of ten duplicate gene families using long-read DNA sequencing allowed us to characterize variation across ~200 modern humans of diverse ancestries, showing that a subset of paralogs retain conservation at the amino-acid level and are likely still functional, evolving under strong purifying selection. To understand their roles in brain development, we generated zebrafish CRISPR “knockout” models of orthologs and transiently introduced mRNA-encoding paralogs, effectively “humanizing” the larvae. Morphometric, behavioral, and single-cell RNA-seq screening highlighted a possible new mechanism for SRGAP2C function in synaptogenesis and, for the first time, a role for GPR89B in dosage-mediated brain expansion. Our holistic approach provides new insights and a useful resource to the community for studying gene expansion drivers of human brain evolution.
After 1960s, the Green Revolution led to a broad introduction of a limited number of elite cultivars. Generally, it led to a move away locally adapted landraces and diminishment of the gene pool. However, the genetic erosion in allelic diversity and fraction of dispensable genes has not been well quantified. In Vietnam, most of the elite varieties and landraces were previously classified into two indica rice subgroups, I1 and I2. We believe that this reflects introduction of elite varieties and narrow genetic diversity. An additional subgroup, I5, was found to be unique to Vietnam and different from all defined sub-populations in 3000 Rice Genetic Project (3K-RGP), a widely used dataset believed to cover global rice genetic variation. We assume that I5 was maintained owing to the local environment or culinary preferences and could be valuable genetic resources. To validate our hypotheses, we generated a de novo assembly of an I5 representative with Pac-bio. We set out to map the population data onto the genome to measure the genetic diversity and identify structural variation and introgression.
DNA satellite evolution triggers a cross-species incompatibility

Mia Levine

Presented by self
University of Pennsylvania (USA)

The recent explosion of telomere-to-telomere assemblies has radically altered our grasp of the identity, organization, and evolution of satellite DNA. However, the functional and evolutionary impacts of this so-called "junk DNA" remain poorly characterized. Here we exploit a large, Drosophila melanogaster-specific DNA satellite array, called 359bp, to investigate the functional significance of fast-evolving DNA satellites. Previous work demonstrated that 359bp colocalizes with a DNA repair protein called Spartan/Maternal Haploid. Intriguingly, Spartan/Maternal Haploid has evolved adaptively along the D. melanogaster lineage. Upon swapping into D. melanogaster a closely related species version of Spartan/Maternal Haploid, we discovered an interspecies incompatibility that manifests in the female germline. This heterologous protein poisoned both oogenesis and early embryogenesis through a DNA damage pathway. Deletion of 359bp completely restored genome integrity and fertility, consistent with a history of antagonistic coevolution between these two fast-evolving loci. To evaluate the universality of this antagonistic coevolution, we explored the evolutionary history of the Spartan gene family across the Drosophila phylogeny. Our phylogenomic and molecular evolution analysis uncovered pervasive positive selection and gene turnover. Notably, one ovary-enriched Spartan family gene recurrently births functionally diverged, testis-enriched daughter genes. These data suggest that DNA satellite divergence not only between species but also between males and females drives Spartan family evolution. Integrating these data with published literature on human Spartan, we develop a new mechanistic model of antagonistic coevolution between satellite DNA and satellite-interacting proteins. Our discoveries offer rare insight into the functional and evolutionary impacts of DNA satellites.
Investigating inversions in Lake Malawi cichlids for a role in speciation and phenotypic variation

Nikesh M Kumar

Nikesh M Kumar, Jeffrey T Streelman, Patrick T McGrath
Georgia Institute of Technology (USA)

Large inversions and other more complicated structural variants have the remarkable property of suppressing recombination between inverted and non-inverted haplotypes, allowing for the accumulation of many functional alleles that are inherited as a single unit. These large variants have been proposed to play an important role in adaptation to local environments and speciation in the presence of gene flow. Here, we use optical genome mapping to explore the presence of inversions in Lake Malawi cichlids, a collection of over 800 species representing one of the largest vertebrate evolutionary radiations known to man. We identify eight new inversions or tandem inversions, between about 0.5 - 22 Mbp in length. Using short-read sequencing of over 141 species, containing members from all seven ecogroups in the lake, we infer the genotypes of these inversions across the lake. Some inversions appear fixed within several of the ecogroups, while others segregate within individual ecogroups in a manner consistent with their spread through hybridization. Their distribution and evolutionary history are supportive for a role in speciation, either in separating the major ecogroups into specific habitats or in subsequent speciation into distinct ecological niches. We will discuss their potential role in controlling sex-determination and prezygotic mating behaviors.
NLRs on the move: Transposition drives the diversification of fungal immune receptors
S. Lorena Ament-Velásquez
S. Lorena Ament-Velásquez, Hanna Johannesson, Aaron A. Vogan
Stockholm University (Sweden), University of Bordeaux (France), Uppsala University (Sweden)

NOD-like receptors (NLRs) are intracellular proteins that independently evolved as key components of the innate immune system of plants and animals. NLRs also exist in fungi, where they can be highly diverse and fast-evolving. The few fungal NLRs of known function control vegetative incompatibility, a process akin to immunity that prevents the spread of mycoviruses. Using high-quality genome assemblies of the Podospora anserina species complex, we show a rapid turnover of NLR gene copies. Indeed, NLRs are often associated with structural variants within and between species, some of which exhibit signatures of balancing selection and/or introgression. Importantly, these polymorphic NLRs are consistently flanked by specific sequences and 5bp-long short direct repeats typical of DNA transposons. We further show that one Podospora lineage that lost a genome defense mechanism against transposons also carries the expansion of two unrelated NLR families. We speculate that these NLRs act as non-autonomous elements mobilized by an unknown transposon, boosting the diversity of variants with potential immunity functions, while simultaneously promoting divergence in genome architecture between species.
Convergent Structural Variation of Antifreeze Proteins in Polar and Deep Sea Fishes

Samuel N Bogan

Aquaculture et Conservation, Australian National Insect Collection, Biodôme de Montréal (Canada), Brigham Young University (USA), CSIRO (Australia), Centre for Deep Sea Research, Conservation et Recherche, Division des Collections Vivantes, Laboratoire de Physiologie, Mississippi State University (USA), Moss Landing Marine Laboratories (USA), San Jose State University, Santa Cruz (USA), Senckenberg LOEWE Center for Translational Biodiversity Genomics (Germany), University of Bergen (Norway), University of California, University of Oregon (USA), Utah State University (USA), Washington State University (USA)

Zoarcoidei is the only suborder of fishes adapted to both the Arctic and Southern Oceans and is uniquely valuable for studying convergent environmental adaptation. A key innovation permitting zoarcoid polar invasions was the evolution of type III antifreeze protein (AFP III). There is evidence that zoarcoids adapted to colder environments possess more AFP copies. It is unknown (i) whether AFP copy number variation (CNV) has convergently evolved during adaptation to both poles and (ii) what genomic mechanisms influence AFP CNV. Addressing these questions requires long read sequencing to obtain accurate measures of AFP copy number, synteny, and structural variation. We sequenced and assembled high quality, long read genomes for nine zoarcoids including deep-water, shallow, temperate, and polar species. Leveraging these assemblies and five other zoarcoid long read genomes, we integrated oceanographic measures of temperature, salinity, and pressure across depth-informed species ranges to model AFP gene family evolution as a function of temperature and pressure, which jointly affect freezing point. AFP copy number convergently increased in species from colder habitats and decreased in deep-water species under high pressure. High AFP copy number in cold, shallow water species was driven by expansions of AFPs independently translocated out of an AFP tandem array in which ancestral AFP III arose. AFP CNV positively correlated with transposable element (TE) density in AFP tandem arrays, and TE density was greater in arrays of translocated AFPs. Our findings suggest roles for translocation and AFP CNV facilitated by TEs during convergent adaptation to extreme cold.
The evolutionary history of mammals reflects a continuous process of genetic innovation, giving rise to a diversity of adaptations. To explore the mechanisms underlying this, substantial efforts have been made to investigate the evolution of protein-coding genes. However, genomic analyses have revealed a significant degree of conservation in these genes, inconsistent with the variation observed across mammals. This suggests that protein-coding genes alone cannot fully explain the diversity of adaptations within mammals, and that regulation of non-coding genes likely plays a crucial role. MicroRNAs are short non-coding molecules that play a key role in regulating genes post-transcriptionally. Although there is substantial evidence to implicate microRNAs in organismal complexity, their biological relevance in mammalian evolution remains largely unclear due to their poor characterisation within the clade. The current landscape of curated microRNAs is limited to only a few model species, and their annotation is heavily reliant on RNA-seq availability. To this end, this study aims to unravel the evolutionary dynamics of microRNAs across mammals using a mapping-based approach. Leveraging known microRNA sequences, their orthologues will be identified in various mammalian genomes. By examining their evolutionary trajectories, this research will ascertain the molecular changes of microRNAs across the current mammal phylogeny, such as copy number variation and nucleotide substitution. This research will provide a comprehensive investigation into microRNA evolution in mammals, building our knowledge beyond the coding genome and helping to bridge the gap between genome and phenotype in the complex class Mammalia.
Evolution of G-Quadruplex Motifs Across Ape Telomere-To-Telomere Genomes
Saswat Kumar Mohanty

Saswat K Mohanty, Kaivan Kamali, Kateryna D Makova, The T2T Primates Consortium
49 Convent Dr (USA), Center for Medical Genomics, Department of Biochemistry and Molecular Biology, Department of Biology, Department of Genome Sciences, Genome Informatics Section, Huck Institutes of the Life Sciences, National Human Genome Research Institute, Penn State University (USA), The Pennsylvania State University (USA), University of Washington School of Medicine (USA)

G-quadruplexes (G4s) are non-canonical (i.e. different from the right-handed double-helix) DNA structures that can form at approximately 1.5% of the human genome. G4s can fold at guanine-rich motifs due to hydrogen bonds among nonconsecutive guanines, and their formation occurs in living cells. G4 structures facilitate double-strand breaks and genomic instability, and thus induce structural variation and point mutations. Additionally, G4 structures have recently emerged as critical regulators of replication, transcription, and telomere maintenance, and, consequently, have been demonstrated to evolve under purifying selection. Despite their significance for mutagenesis and cellular functions, the evolution of G4s remains understudied. To fill this gap, we conducted a comprehensive analysis of G4 motifs in the recently-released complete, telomere-to-telomere (T2T) genomes of apes—bonobo, chimpanzee, gorilla, Bornean and Sumatran orangutans, and siamang gibbon. Using genome-wide multi-species alignments, we found that in general, G4 motifs accumulate at a rate that is expected based on the phylogeny among the studied species. Whereas many G4 motifs were shared among species, such sharing was not uniform across the genome and was particularly strong at enhancers and promoters, highlighting the role of G4 structures in regulation of transcription. At the same time, we discovered thousands of species- and genus-specific G4 motifs, suggesting their rapid emergence/decay and potential role in evolution of individual ape taxa. Our results open avenues for subsequent studies of individual G4 motifs and structures and of their role in ape evolution.
**Intergenic structural variation and ancient gene duplication underpin pigmentation diversification in swordtail fish**

Tristram Dodge

Tristram O Dodge, Daniel L Powell, John J Baczenas, Theresa R Gunn, Manfred Schartl, Molly Schumer
Centro de Investigaciones Científicas de las Huastecas "Aguazarca" (Mexico), Howard Hughes Medical Institute (USA), Stanford University (USA), Texas State University (USA), University of Wuerzburg (Germany)

Pigment patterns provide striking examples of phenotypic diversification and are a tractable system to study the genetic basis and evolution of adaptive traits. While melanic patterns in mammals are best studied, teleost fish present an exciting system due to their high levels of pattern variation and increased pigment cell diversity. Swordtail fish in the genus Xiphophorus are polymorphic for dozens of pigmentation traits under balancing selection, providing excellent opportunities to characterize both genetic architecture and selection. Here, we study four melanic patterns in two diverged Xiphophorus clades. Using GWAS and QTL mapping in five species, we find three distinct patterns on the body and tail map to intergenic regions upstream of kitlga, a gene involved in melanocyte migration. RNaseq and allele-specific expression data show increased kitlga expression in these specific tissues, and the nonoverlapping GWAS peaks hint that distinct regulatory elements may drive each pattern. Surprisingly, we find tailspot phenotypes in a second Xiphophorus clade map to kitlgb, a gene that arose by duplication 350 million years ago. Using long-read sequencing, we identify and resolve intergenic structural variants perfectly associated with each trait. Pairwise alignment and pan genomic approaches show these structural variants are complex and often include multiple duplications. Haplotype comparisons across species reveal unexpected patterns of allele sharing and divergence, suggesting prominent roles of introgression and allelic turnover. Our results suggest that ancient duplications can set the stage for recent phenotypic diversification, and intergenic structural variants may play an under-appreciated role in gene regulation and adaptation.
"Characterization of Inversions in 1000 Individuals across the Human Population using Single-Cell Pooled and Long-Read Data."

Vasiliki Tsapalou

Vasiliki Tsapalou, Wolfram Hoeps, Tobias Rausch, Thomas Weber, Siegfried Schloissnig, Eva Benito-Garagorri, Tobias Marschall, Jan Korbel

Department of Human Genetics, European Molecular Biology Laboratory (EMBL), Genome Biology Unit (Germany), Heinrich Heine University Duesseldorf (Germany), Institute for Medical Biometry and Bioinformatics, Medical Faculty, Radboud University Medical Center (Netherlands), Research Institute of Molecular Pathology (IMP), Vienna Biocenter (Austria)

Structural variants (SVs) are key contributors to human genomic diversity. Despite technological advances in SV detection, knowledge gaps remain, notably for inversions. These copy-number neutral variations, often flanked by identical repeat sequences, are notoriously challenging to detect. Using data from 1000 individuals in the 1000 Genomes Project, we aim to comprehensively identify inversions, understand their formation mechanisms, and explore their phenotypic impact and evolutionary relevance across populations. Employing single-cell sequencing (Strand-Seq) and Oxford Nanopore Technologies (ONT) data from the 1000 Genomes IMP/MARVL study, we aim to identify a broad spectrum of previously undetected inversions. Strand-Seq\'s ability to reveal read directionality is crucial for detecting large inversions, particularly near complex regions. In this regard, we pursued an innovative single-cell data generation method, coupling cell line pooling with strand-specific sequencing. This approach enables efficient sequencing of diverse human samples, optimizing cell numbers for precise inversion identification and genotyping. Concurrently, we leverage ONT data and refine existing computational approaches to capture smaller to mid-sized genomic rearrangements, with a focus on detecting small inversions up to 1kb. Our research has identified 638 novel polymorphic inversions, establishing the most comprehensive inversion dataset to date. We will discuss these events, emphasizing their recurrence, functional consequences, and population distribution, including rare haplotypes with diverse phenotypic impacts. This endeavor is poised to advance large-scale population-genetics studies, providing a unique opportunity to decipher natural and disease-causing structural variations within various population groups, and their role in adaptation and genome evolution.
A genomic perspective of variation in a highly diverse Dipteran family

Wenjun Zhou

Wenjun Zhou, Chris D Jiggins
Department of Zoology, University of Cambridge (United Kingdom)

Structural variation and transposable elements have been shown to be major factors influencing species diversity and adaptation. Here, we explore the structural variation and repetitive landscape of six high-quality chromosome level reference genomes of the soldier fly family, Stratiomyidae, a highly diverse taxa with differences in distribution, behaviours and morphology. Whole genome synteny showed chromosomal rearrangements across the phylogeny. A total of 86,085 genes were assigned to 14,181 orthogroups, with 1,142 of which are species-specific. Result showed a large number of gene duplications throughout their speciation history, with most of them happening at the species nodes in the phylogenetic tree, indicating recent and species-specific duplication events related to their adaptation. All species used in the study showed a high proportion of repetitive elements, varying from 63.73% to 73.18% of the whole genome. Among these species, the black soldier fly (Hermetia illucens) has the highest proportion of long interspersed nuclear elements (LINEs), with 45.14% of its genome consisted of this type of transposable elements. Our result reveals rarely-explored divergence in a wide-spread Dipteran family, and suggests potential link between genomic structural variation and functional adaptation of their genes.
Transcriptomic and proteomic effects of gene deletion are evolutionarily unconserved

Yang Li

Yang Li, Jianzhi Zhang
Department of Ecology and Evolutionary Biology, University of Michigan (USA)

The function of a gene is the effect for which the gene was selected and/or by which it is maintained. Although gene function is commonly inferred from the phenotypic effects of deleting the gene, not all phenotypic effects reflect the gene’s selected-effect function. To evaluate the degree to which the phenotypic effects of gene deletion inform gene function, we compare the transcriptomic and proteomic effects of systematic gene deletions in budding yeast (Saccharomyces cerevisiae) with those effects in fission yeast (Schizosaccharomyces pombe). Surprisingly, despite evidence for functional conservation of orthologous genes, their deletions result in no more sharing of transcriptomic or proteomic effects than that from deleting non-orthologous genes. Because the wild-type mRNA and protein levels of orthologous genes are significantly correlated between the two yeasts, our observation cannot be explained by rapid evolution or large measurement error of gene expression. Analysis of transcriptomic and proteomic effects of gene deletions in multiple S. cerevisiae strains reveals a high sensitivity of these effects to the genetic background. Together, these observations suggest that most transcriptomic and proteomic effects of gene deletion do not inform selected-effect function. This finding has important implications for assessing and/or understanding gene function, pleiotropy, and biological complexity.
S27 - Transposable elements in the population genomics era: towards a better understanding of their contribution to evolution.
Two centuries of transposable element invasions in Drosophila melanogaster

Almorò Scarpa

Almorò Scarpa, Riccardo Pianezza, Sarah Signor, Robert Kofler
North Dakota State University (USA), University of Veterinary Medicine, Vienna (Austria), Vienna Graduate School of Population Genetics (Austria)

Transposable elements (TEs) significantly influence genome evolution and phenotype. However, the rate of TE horizontal transfer remains a key unresolved question. Previous studies have demonstrated four TE families invading D. melanogaster in the 20th century. Expanding on this, we examined historical specimens spanning an additional 100 years, revealing three LTR retrotransposons (Blood, Opus, and 412) spreading in the 19th century. Furthermore, we identified a novel TE, named Spink, invading D. melanogaster post-1990s. Employing a novel methodology, we detected three more TEs, the latest appearing globally less than ten years ago. In total, eleven TE families have invaded D. melanogaster over 200 years, expanding the genome by approximately 1 Mbp. This rapid horizontal transfer rate, coupled with the origin of some TEs from American species, suggests human activity may have facilitated inter-species contact, contributing to TE spread.
Transposon-mediated evolution of bat immune signaling receptors
Andrea D Ordoñez

Andrea D Ordoñez, Conor J Kelly, Edward Chuong
Boulder (USA), University of Colorado, University of Washington (USA)

Bats are reservoirs for zoonotic viruses that pose significant threats to human health. There is growing evidence that bats evolved unique immune adaptations, allowing them to coexist with viruses. Amongst these adaptations is a subdued inflammatory response, shown by a decrease of activity or loss of important antiviral sensors and inflammasome proteins in bats. By leveraging long-read RNA sequencing from cell lines and primary tissues of the Jamaican fruit bat, Artibeus jamaicensis, we detected a truncated isoform of the interleukin 18 receptor (IL18R1), referred to as IL18R1-short, that lacks most of the intracellular domain. IL18 is a pro-inflammatory cytokine important for innate and adaptive immunity and is highly elevated in autoimmune conditions. We predict that IL18R1-short is a decoy receptor and downregulates IL18 induced inflammatory responses. Interestingly, this truncation is mediated by a LINE2 transposable element (L2-TE) that introduces an early polyA signal. IL18R1-short (and the L2) is conserved in humans but very lowly expressed in a few tissues and immune cells. Comparatively, bat IL18R1-short is highly detected in all Artibeus tissues and cell-lines tested thus far. This suggests a bat-specific adaptation increases the relative expression of this truncated isoform. We are over-expressing Il18R1-short in various human and bat cells to test how it modulates IL18 signaling. To investigate the mechanism resulting in the differential expression of IL18R1-short in bats and humans, we are studying the relative strengths of the polyA signal from the L2 element and splicing regulation for the terminal exons of the human and bat IL18R1 genes.
Transposable element copy variation in cultivated apples, and their wild apple relatives

Anthony Venon

Anthony Venon, Xilong Chen, Aurélie Mesnil, Karine Alix, Amandine Cornille
AgroParisTech (France), CNRS (France), INRAE (France)

Transposable elements (TEs) are repetitive sequences of the genome that can play critical roles in major agronomic traits. In addition, TEs can be involved in the adaptation of populations to new environments. The role of TEs in the domestication of perennial crop, including fruit trees, still needs to be studied. Exploiting short-read sequencing data, we annotated TEs and estimated their copy numbers among populations of the cultivated apple tree (Malus domestica) and its wild relatives, representing a total of 230 accessions. To do so, first, we built an original bioinformatic pipeline combining short-reads and TE predictors (RepeatExplorer and REPdenovo) to obtain an exhaustive TEs database for the genus Malus. Subsequently, we annotated this database with different bioinformatic tools (PASTEC and TEsor). This database contains a total of 158k consensus sequences with a predominance of LTR-retrotranposons. We then mapped the raw short read sequences of the 21 accessions onto this TE database to get the copy number of each TE consensus. Statistical analyses revealed significant TE copy number variations between cultivated and wild apple trees and among the cultivated apple trees depending on their uses. Further analyses are being done to identify which specific TEs mainly explain the observed variation. Our study represents a significant advance in understanding the role of TEs in the domestication of fruit trees and their role in genome evolution and adaptation.
Transposable elements (TEs) are genomic elements that are ubiquitous across the tree of life and play a crucial role in genome evolution and function. Advances in long-read sequencing have allowed highly accurate TE detection, though at a higher cost than short-read sequencing. Recent long-read studies have shown that existing short-read TE detection methods perform inadequately when applied to real data. In this study, we use a machine learning approach (called TEforest) to discover and genotype TE insertions and deletions with short-read data by leveraging true-positive annotations from long-read genome assemblies. Our method first uses a highly sensitive algorithm to discover potential TE insertion or deletion sites in the genome, extracting relevant features from short-read alignments. To discriminate between true and false TE insertions, we train a random forest model with a labeled ground-truth dataset for which we have measured the same set of short-read features. We conduct a comprehensive benchmark of TEforest and traditional TE detection methods using real data, finding that TEforest identifies more true positives and fewer false positives across datasets with different read lengths and coverages, while also accurately inferring genotypes and the precise breakpoints of insertions. By learning short-read signatures of TEs previously only discoverable using long reads, our approach bridges the gap between large-scale population genetic studies and the accuracy of long-read assemblies. This work provides a user-friendly tool to study the prevalence and phenotypic effects of TEs across species.
Transposable elements rapidly induce chromosomal rearrangements in locally adapted island Drosophila

Brandon Turner

Brandon Turner, Theresa Erlenbach, Nicholas Stewart, Cathy Moore, Robert Reid, Rebekah Rogers, Rebekah Rogers
Southern Oregon University (USA), University of Georgia (USA), University of North Carolina - Charlotte (USA)

By evaluating rearrangements in island Drosophila and comparing to mainland Africa Drosophila we can utilize population differentiation, population genetic statistics and differential expression data to quantify the magnitude of these various genetic mechanisms and their role on evolution. By evaluating how chromosomal rearrangements result in phenotypic changes, we can gain insight on how rapid evolution can reshape populations. We hypothesize that chromosomal rearrangements can be a rare source of innovation that potentially produces cellular changes and, while generally detrimental or neutral, can potentially result in adaptive variation during habitat shifts. We wish to discern whether these "hopeful monsters" among genes can lead to adaptive success in new environments while firmly separating out their effects from that of random genetic drift.
Evolution of transposable elements as the source of extracellular RNA

Cei Abreu-Goodger

Isaac Martinez-Ugalde, Kyriaki Neophytou, Yenetzi Villagrana-Pacheco, Xiaochen Du, Mike Evans, Lewis Stevens, Mark Blaxter, Amy Pedersen, Amy H Buck, Cei Abreu-Goodger

Institute of Ecology and Evolution, Institute of Immunology and Infection Research, Tree of Life, University of Edinburgh (United Kingdom), Wellcome Sanger Institute (United Kingdom)

Beyond self-replicating, transposable elements (TEs) can be a source for new functions. Small RNAs (sRNAs) and Argonaute proteins are important for silencing TEs, however, sRNAs can also be effector molecules for some parasites. The parasitic nematode Heligmosomoides bakeri can modulate the immune response of mice by secreting a variety of molecules including proteins, lipids, sRNAs and the extracellular worm Argonaute (exWAGO). Importantly, secreted sRNAs are mostly derived from TEs, suggesting a relationship between TE silencing and parasitism. H. bakeri has one of the largest, TE-rich, nematode genomes. Nevertheless, little is known about the evolution of these TEs, or their functions in parasitism. Using immunoprecipitation (IP) of exWAGO in H. bakeri, and sequencing bound sRNAs, we found that they map to all major classes of TEs, with a preference for particular families. LTR/Pao and LTR/Gypsy retrotransposons seem particularly relevant, producing ~8-fold more sRNAs than expected. IP experiments in H. polygyrus, Nippostrongylus brasiliensis, Teladorsagia circumcincta, and Ancylostoma ceylanicum confirm that, although their genomes vary in size and TE composition, exWAGO preferentially loads sRNAs derived from LTRs. Furthermore, we found that specific TE families are relevant for secreted exWAGO in H. bakeri, and are conserved across parasites and free-living nematodes. We propose exWAGO as an ancestral regulator of retrotransposon activity. Parasitic nematodes have since gained the capacity to secrete exWAGO and sRNAs during infection, alongside an expanded and diverged repertoire of transposable elements in their genomes.
Insights into PIF transposase-derived gene functionality in germline development in Drosophila
Chathuri Devmika Wickramasinghe

Chathuri Devmika Wickramasinghe, María del Pilar Castellanos, Diwash Jangam, Fatema B. Ruma, Dragomira N. Markova, Cédric Feschotte, Esther Betrán
Cornell University (USA), University of Texas at Arlington (USA)

Many studies have demonstrated the contribution of transposable elements (TE) derived proteins to the emergence of new host proteins. Across Drosophila genus, there are seven PIF TE-derived genes known as Drosophila PIF Like Genes (DPLG1-7), domesticated from PIF/Harbinger transposase superfamily. We are studying potential functions of the four DPLGs (DPLG1-4) present in D. melanogaster, which are highly conserved and were likely independently domesticated. DPLG1-4 are highly expressed in gonads and nervous system. Here, we focus on DPLG1 and DPLG4 female germline function. We have studied localization of DPLG1-HA and DPLG4-HA tagged proteins in ovaries. RNA-Seq analysis of ovaries from CRISPR-Cas9 generated knock-out flies revealed that the absence of these genes causes differential expression of a set of overlapping genes in ovaries, with positively correlated changes in expression. Several genes in piRNA pathway were found among DE genes in DPLG1-KO ovaries. Further, several TE families were differentially expressed in both DPLG1-KO or DPLG4-KO ovaries. Most upregulated TEs belonged to Gypsy LTR retrotransposon superfamily, whereas downregulated TEs were telomeric non-LTR retrotransposons. Interestingly, preliminary small RNA-Seq analysis revealed an increase in piRNAs targeting these telomeric elements, which may explain the reduction in telomeric elements. These findings strongly suggest that DPLG1 and DPLG4 may have potential regulatory functions in Drosophila telomere elongation through piRNA pathway. Collectively, our studies imply complex interactions among transposase-derived genes repeatedly co-opted from the same TE superfamily, to function within the same cellular pathways.
How transposable element shapes genome evolution through epigenetic mechanisms?
Grace Yuh Chwen Lee

Presented by self
Irvine (USA), University of California

Transposable elements (TEs) are powerful engines of genome evolution, as illustrated by their implication in genome size variation and the creation of new cellular functions. Short-term consequences of TE mobilization can also be particularly dramatic given that TE insertions are a unique source of large effect mutations and that transposition can be sensitive to the environment. Despite these considerations, there is a little knowledge about the role of ongoing transposition to within-species variation. I will present our efforts to characterize the rate and landscape of natural TE mobilization using the model plant species Arabidopsis and to assess the role of the environment both as a selective force and a trigger of TE-driven variation.
Endogenous retroviruses (ERVs) are a class of transposable elements which comprise 10-15% of vertebrate genomes and are known to be a major source of genomic structural variation, contributing to vertebrate genome evolution. While most ERV research has focused on mammals, we are interested in ERV presence and distribution in Anuran species (frogs and toads). We have designed a bioinformatic pipeline to identify and characterise ERV insertions, using iterative in-silico screening to refine detection and assess insertion loci and structure. This pipeline was used to screen for alpha- and gamma- ERVs. While gamma- ERVs are ubiquitous in vertebrate genomes, alpha-like ERVs appear avian-specific, with detection in Amphibians only in the sister order to Anura, Gymnophiona. Here we processed publicly available Anuran genomes (n=48) and have identified multiple loci of alpha-like and gamma-like ERV elements, some of which are shared across Anuran species. Additionally, through phylogenetic and motif analysis, the genome organisation of some full-length insertions appeared variable, suggesting recombination amongst ERVs in Anura. We propose that the rarity of previous reports of amphibian alpha-like ERVs is likely explained by the limited number of studies using methods that can detect them, as well as their possible recombined states in the genome. These new discoveries suggest that alpha-like ERVs are more widely distributed in vertebrates than originally proposed. These initial steps in characterising ERVs in Anura can provide insight into whether ERV impact on genome evolution is consistent across vertebrates.
Independent transitions to sociality in Stegodyphus spiders are associated with transposable element expansions

Jilong Ma

Jilong Ma, Anna Igolkina, Jesper Bechsgaard, Palle Villesen, Trine Bilde, Magnus Nordborg, Mikkel Heide Schierup
Aarhus (Denmark), Aarhus University (Denmark), Gregor Mendel Institute of Molecular Plant Biology (Austria)

Transposable elements (TEs) can insert into functional genomic regions and induce ectopic recombinations in the genome. Due to these larger effect mutations, transposable elements are a possible threat to the genome integrity in evolution. The threat is expected to be larger when nature selection is less efficient in inhibiting TE proliferation. Social evolution in the spider genus Stegodyphus is a case where the development of a life history trait reduces selection efficiency. The evolution of sociality is repeatable in Stegodyphus across three independent transitions within the past hundred-thousands years. All social species appear to be evolutionary short-lived in the phylogeny. We wish to test whether TE could be a reason for the hypothesized evolutionary dead end, exploiting de novo haplotype-phased long-read genomes of 7 species in Stegodyphus. We used polymorphisms of structural variants from alignments of haploid genomes to identify all mobile TE families across Stegodyphus. We then constructed and contrasted time-resolved TE proliferation rate per TE family in closely related pairs of social and subsocial species. We find that more TE families expand, and each to a larger extent, in social species. The more TE proliferation in social lineages correlates with the establishment of sociality over time. The recently acquired TE loci of these TE families also show a higher expression level in social species. We conclude that the social evolution in spiders, representing the development of a high-inbreeding mating system, leads to the expansion of TEs.
Humans and chimpanzees differ in anatomy and physiology despite their genomes being approximately 98% identical. Differential gene regulation is the major driver of the phenotypic differences between these closely related species. Transposable Elements (TEs) comprise fifty percent of primate genomes and can contribute to gene regulation within and between species. However, their relative contribution appears to differ between pluripotent and somatic cell types. To study the evolutionary dynamics of TE regulation during development, we utilize an in vitro system to differentiate iPSCs of humans and chimpanzees into cardiomyocytes. We collected cells at four stages of differentiation, as determined by 50% of the cell population expressing stage-specific markers by flow cytometry. Stages include pluripotent cells (OCT3/4+), mesoderm (Brachyury+), cardiac mesoderm (ISL1+), and terminal cardiomyocytes (TNNT2+). To study TE regulatory dynamics, we profiled the enrichment of H3K27ac, a marker of active regulatory regions, by CUT&Tag followed by sequencing. We detected tens of thousands of H3K27ac regions in orthologous regions in both species, with the majority being variable across differentiation stages. Stage-variant regions in humans are less likely to be conserved across species than stage-invariant regions. Overall, conserved H3K27ac regions are enriched for orthologous SINE elements amongst all annotated orthologous TEs, suggesting involvement in gene regulation. Conserved stage-variant regions are more likely to overlap an annotated orthologous TE than conserved stage-invariant regions, suggesting TEs contribution to stage-specificity. We are currently collecting data from additional individuals to investigate the stage- and species-specificity of TE family regulation.
Evolution of transposable elements in arbuscular mycorrhizal fungi: insights into genome variability and gene regulation

Jordana I.N. Oliveira

Jordana I.N. Oliveira, Catrina Lane, Ariane M. Nicol, Nicolas Corradi
Queens University (Canada), University of Ottawa (Canada)

Arbuscular mycorrhizal fungi (AMF) are plant symbionts that provide nutrients from the soil to their hosts in exchange for carbohydrates and lipids. AMF exhibit low morphologic variability and no sign of plant specificity, raising questions about how these organisms have adapted to a wide spectrum of host species. Genomic analysis has revealed that AMF species harbor a significant portion of transposable elements (TEs), which play a crucial role in promoting genome variability by generating new copies and insertions. However, on average 60% of TEs in AMF genomes remain unknown. We hypothesize that investigating the TE content can unveil AMF diversity and detect the impact of these sequences in promoting genetic variability. To achieve this, we constructed a TE library using 26 AMF species, identifying the presence of transposition domains in consensus sequences. The annotation of TEs revealed that the major groups of AMF share patterns of transposition, with ancestral Paraglomerales exhibiting lower TE content and smaller genome sizes compared to more derived groups Archeosporales, Glomerales, and Diversisporales, which show clade-specific expansions. Furthermore, a significant proportion of TE insertions are located in exons, suggesting an impact of these repetitive sequences on protein diversity through exonization processes. Expression analysis in the model species Rhizophagus irregularis showed a fourfold increase in upregulated TEs during symbiosis compared to extraradical mycelium, indicating a role of TEs in communicating with the host. We propose that TE variability in AMF may contribute to the evolution and diversification of genes, influencing their successful symbiotic relationships with plants.
Transposable elements (TEs) are powerful engines of genome evolution, as illustrated by their implication in genome size variation and the creation of new cellular functions. Short-term consequences of TE mobilization can also be particularly dramatic given that TE insertions are a unique source of large effect mutations and that transposition can be sensitive to the environment. Despite these considerations, there is a little knowledge about the role of ongoing transposition to within-species variation. I will present our efforts to characterize the rate and landscape of natural TE mobilization using the model plant species Arabidopsis and to assess the role of the environment both as a selective force and a trigger of TE-driven variation.
Functional study of DPLG3, a PIF transposable element domesticated protein, in Drosophila melanogaster
María del Pilar Castellanos

María del Pilar Castellanos, Chathuri Devmika Wickramasinghe, Fatema Ruma, Esther Betrán
Coriell Institute for Medical Research (USA), The University of Texas at Arlington (USA)

Molecular domestication of transposable element (TE) proteins can be viewed as the process of recycling the TE selfish protein/s for cellular functions. This process is gaining attention in evolutionary biology due to increased genome sequencing and functional analyses. Here, we analyze a PIF/Harbinger family of TE proteins domesticated gene in Drosophila melanogaster, Drosophila PIF-like gene 3 (DPLG3). We tagged DPLG3 with HA and observed its localization and co-localization with DNA in ovaries. Knock-out flies generated via CRISPR-Cas9 displayed abnormal gonadal phenotypes that are partial penetrant and maternally inherited, with 10-15% of females and fewer males affected. RNA-Seq analysis of non-rudimentary DPLG3-KO ovaries revealed differentially expressed genes, TEs and long non-coding RNAs suggesting a potential regulatory role for DPLG3. Our hypothesis posits that DPLG3, because of its nuclear localization, has retained its DNA binding domain and likely functions as a transcription regulator. Further studies aim to elucidate the role of this domesticated transposase in Drosophila germline development.
DNA Transposons favour de novo transcript emergence through enrichment of transcription factor binding motifs

Marie Lebherz

Marie Lebherz, Bertrand Fouks, Julian Schmidt, Erich Bornberg-Bauer, Anna Grandchamp
Department of Protein Evolution, Institute for Evolution and Biodiversity (Germany), Max Planck Institute for Biology Tübingen (Germany)

Transposable elements (TE) are known to play an essential role in driving evolvability and thus facilitating fast acting molecular adaptive processes at the DNA level. We report here on how TEs can affect the recently detected and meanwhile well confirmed process of de novo gene emergence. Specifically, we focused on the gain of new transcription events. In many species a continuum between absent and very prominent transcription has been reported for essentially all regions of the genome. In this study we searched for de novo transcripts by using newly assembled genomes and transcriptomes of seven inbred lines of Drosophila melanogaster, originating from seven geographically diverse populations. This setup allowed us to detect line specific de novo transcripts, and compare them to their homologous non-transcribed regions in other lines, as well as genic and intergenic control sequences. We studied the association with TE and the enrichment of transcription factor motifs upstream of de novo emerged transcripts. We found that de novo transcripts overlap with TEs more often than expected by chance. The emergence of new transcripts correlates with epigenetic marks islands and regions of TEs activity. Moreover, upstream regions of de novo transcripts are highly enriched with regulatory motifs. Such motifs abound in new transcripts overlapping with TEs, particularly DNA TEs, and are more conserved upstream de novo transcripts than upstream their non-transcribed homologs. Overall, our study demonstrates that TE insertions play an important role in new transcript emergence, partly by introducing new regulatory motifs from DNA TE families.
Epistasis analysis reveals adaptive separation-of-function mutants in experimentally evolved Saccharomyces cerevisiae

Michelle Hays

Michelle Hays, Katja Schwartz, Gavin Sherlock
Stanford University (USA)

Previously, we showed that several mutational mechanisms give rise to adaptation when Saccharomyces cerevisiae are experimentally evolved under fluctuating nitrogen starvation. Adaptive mutations arise from SNPs, but also from other mutational mechanisms such as increased retrotransposon activation and microhomology mediated recombination leading to structural variation and gene conversion. Different adaptive alleles differ in their fitness benefit. The notable differences in adaptive alleles at the same loci, both in fitness and putative gene products, led us to pursue epistatic analysis of several beneficial mutations that arose in the experimental evolution at recurrently mutated genomic loci (such as MEP1, GAT1, and PAR32). Here, through double mutant analysis, we see that negative epistasis is most common, with double-mutant fitnesses most often being similar to the fitness increase of the higher of the two single mutants. These data largely best fit a model of diminishing returns, but with a few notable exceptions. In specific examples we observe unexpected combinatorial fitness effects, suggesting that some adaptive alleles may in fact be separation of function or gain of function mutants. For example, beneficial missense mutations at the nitrogen transporter MEP1 exhibit different epistatic interactions and overall fitness changes than novel 3’ retrotransposon insertions at the same genomic locus. Nitrogen catabolism regulation in S. cerevisiae is complex, with several feedback loops: these evolved mutations provide means for disentangling adaptation in a complex regulatory network and exploring the mechanism of specific subfunctions in gene products most likely to impact fitness and organism adaptation.
Double trouble: two retrotransposons triggered a cascade of horizontal transfers in Drosophila species within the last 50 years

Riccardo Pianezza

Riccardo Pianezza, Almorò Scarpa, Sarah Signor, Robert Kofler
North Dakota State University (USA), Veterinary Medicine University (Austria), Vienna Graduate School of Population Genetics (Austria)

Horizontal transfer of genetic material in eukaryotes has rarely been documented in short evolutionary timescales. We show that two retrotransposons, Shellder and Spoink, invaded the genomes of multiple species of the D. melanogaster subgroup in the past 50 years. Through horizontal transfer, Spoink spread in D. melanogaster during the 1980s, while both Shellder and Spoink invaded D. simulans in the 1990s. Likely following hybridization, D. simulans infected the island endemic species D. mauritiana (Mauritius) and D. sechellia (Seychelles) with both TEs after 1995. Shellder additionally invaded D. teissieri, a species confined to sub-Saharan Africa, in approximately the same time-frame. The donors of Shellder and Spoink are likely American Drosophila species from the willistoni, cardini, and repleta groups. Thus, the described cascade of TE invasions only became feasible after D. melanogaster and D. simulans extended their habitats into the Americas 200 years ago, likely aided by human activity. Our work reveals that cascades of TE invasions, likely initiated by human-mediated habitat expansions, could have a profound impact on the genomic and phenotypic evolution of many geographically dispersed species. Within a few decades, TEs can invade many species, including island endemics, with habitats very distant from the initial donor of the TE.
Mammalian genomes are rich in transposable elements (TEs), which have been shown to contribute to phenomena as varied as the concerted regulation of genes to specific phenotypic changes, all of which can impact an organism’s capacity to adapt to its environment. Nonetheless, TEs are often omitted from genomic studies. Here, we explore the transcriptional landscape of TEs throughout the development of seven organs (cerebrum, cerebellum, heart, kidney, liver, ovary and testis) from early organogenesis to adulthood for human, rhesus macaque, mouse, rat, rabbit, opossum, and chicken as an outgroup. We identified considerable differences in TE expression among organs and developmental stages within species and between species. Taken together with each species’ TE repertoire, with this work, we seek to uncover niches of transcriptional permissibility, some of which could functionally impact mammalian organ development.
"Un-trapping": Does the size of piRNA clusters predict the abundance of transposable element insertions?

Sarah Saadain

Sarah Saadain, Riccardo Pianezza, Robert Kofler
Institute of Population Genetics, University of Veterinary Medicine Vienna (Austria), Vienna Graduate School of Population Genetics (Austria)

Identifying the factors that determine the abundance of individual transposable element (TE) families remains a major open question in the field. Under the trap model, it is assumed that an invasion of a TE is stopped when a member of the invading family jumps into a piRNA cluster and thereby triggers the production of piRNAs, which in turn silences the TE. Simulation studies suggest that the total size of all piRNA clusters (as a percentage of the genome) should be the most crucial aspect, where species with a smaller proportion of clusters ought to accumulate many TE insertions and the inverse being true for those with larger proportions of piRNA clusters. By estimating the size of piRNA clusters and the abundance of individual TE families in 15 Drosophila species, we tested whether the cluster size can predict the observed abundance of the individual TE families. To exclude negative selection against TEs as a potential bias, we focused on young TE families, i.e. families with a low pairwise divergence among the insertions. Contrary to expectations under the trap model, we find no correlation between the size of piRNA clusters and the abundance of TEs. Our work suggests that other mechanisms, such as siRNAs, may largely be responsible for triggering host defense mechanisms to combat invading TEs.
S28 - Human genetic variability in the Pang genomic era
Living with your dynamic genome: lessons learned from T2T genomes
Arang Rhie

Presented by self
National Human Genome Research Institute (USA)

The completion of the first telomere-to-telomere (T2T) human genome elucidated sequences that were not accessible before. The majority of the new sequences reside in centromeric, satellite repeats or segmental duplications, which were difficult to assemble and map despite its important role. Among the large, newly accessible regions, several reside in the acrocentric chromosomal p-arms. The draft human pangenome sequences revealed loci of high sequence homology and recombination rates, which we termed "pseudo homologous regions (PHR)". Frequent recombination occurs especially among three chromosomes: Chrs 13, 14 and 21, with any translocation with Chr 14 leading to the classical Robertsonian translocations, resulting in the loss of the distal junctions of the rDNAs and a dicentric chromosome. To confirm the frequency of Robertsonian carriers in healthy individuals, we built a pipeline to accommodate these recombinations for short-read mapping, variant calling, and performed functional annotation of the variants. We confirmed the pipeline successfully identified losses of the distal junctions in patients diagnosed as Robertsonian translocation carriers, matching results from chromosome painting. Using the pipeline on ~4,000 healthy new-borns and its family cohort, we found 4 additional potential Robertsonian carriers, matching the previous report at about >1 in every 800 individuals. Linking the distal junctions of the ribosomal DNA to the pairing q-arm was a long standing challenge in genome assembly for the last couple years. However, using the latest assembly methods and the pangenome, it is now possible to obtain the complete sequences of the most dynamic region, finally giving access to the biology behind it.
Tandem tales: comparative analysis of tandem repeats in Ape genomes

Carolina de Lima Adam

Carolina de Lima Adam, Joana L Rocha, Peter Sudmant, Rori Rohlf
University of California Berkeley (USA), University of Oregon (USA)

Tandem repeats (TRs) are important contributors to genetic variation in eukaryotes. Their high mutation rate makes them strong candidates for surveys of recent evolution across species. Also, recent studies highlight the association of TRs with gene expression and complex traits. However, commonly used short-read sequencing technologies have failed to sequence long, repetitive TRs and their multiallelic nature is a challenge for genotyping tools designed for biallelic markers. Our investigation uses long-read, telomere-to-telomere high-fidelity genome data, and TR-specific genotyping software to identify TR loci in seven ape species. Our results revealed that distributions of TR total length and repeat unit size are conserved across all examined species, with a high frequency of primate-specific transposable elements, including Alu elements. To obtain homologous TRs between humans and the remaining species we performed liftover analysis with the human TR catalog as the target. The resulting orthologous regions were then intersected with the TR catalog of each species. In the chimp x human comparison, ~90.7% of the TR-containing regions are orthologous. By employing a phylogenetic framework across all seven species, we can access signs of past selection. TR phylogenies with exceptional species-specific long branches suggest directional selection, while TR loci subjected to balancing selection will exhibit deeper genealogies due to the maintenance of ancient polymorphisms. This comprehensive investigation will advance our understanding of TR evolution and provide valuable insights into the potential role of TRs in complex traits and rapid adaptation in primates.
Implicit pangenomics
Erik Garrison

Presented by self
University of Tennessee Health Science Center (USA)

We explore a new paradigm in pangenome analysis that allows us to scale base-level pangenome analyses to the entire vertebrate clade in parallel. Instead of building pangenome graphs directly, we represent them implicitly as a set of alignments between sequences. We have developed a tool, impg (IMplicit Pan Genome), which allows us to work on the variation graph implied by whole genome alignments without ever rendering it. On its most basic level, impg provides a liftover from any genome to any other genome in the dataset based on the set of alignments provided. While this lifting function may seem trivial, we show that it lies at the heart of modern pangenomic approaches, which focus on aligning genomes and constructing graphical models or relationships from them. To showcase the potential of this approach, we present analyses based on telomere-to-telomere assemblies of primate genomes. By compressing whole genome alignments against each reference genome in turn, we establish metrics of conservation, divergence, and incomplete lineage sorting based purely on pairwise alignments between all genomes. These metrics are established in all frames of reference, allowing us to explore different perspectives when considering them. Our future work will expand on this paradigm, using impg to subdivide complex graph building problems for efficient completion. This approach has the potential to revolutionize comparative genomics by enabling efficient, dynamic, large-scale pangenome analyses across diverse species, ultimately deepening our understanding of evolutionary relationships and genome function.
Maximum-likelihood inferences from ancient environmental DNA using panmitogenomes

Gabriel Renaud

Gabriel Renaud, Nicola Alexandra Vogel, Joshua Daniel Rubin
Technical University of Denmark (Denmark)

Due to its small size and high abundance relative to nuclear DNA, mitochondrial DNA continues to play a central role in ancient DNA and ancient environmental DNA studies. Some special challenges plaguing such studies include high divergence to all known modern references and high rates of ancient DNA damage. Especially for ancient environmental DNA, the scarcity of data becomes problematic when one seeks to make inferences based on such scant data. We present two different applications of panmitogenomics, i.e. pangenomics applied to mitochondrial DNA. These applications target ancient DNA from a single source and ancient environmental DNA. The first involves the robust inference of human mitochondrial haplogroups from ancient DNA. We show that our method outperforms existing tools by offering greater resilience against low coverage or incomplete consensus sequences via the mitigation of reference-related biases. Additionally, it generates Bayesian confidence scores informed by phylogeny that are impartial to any specific population. The second application concerns the sensitive detection of animal DNA traces in ancient environmental DNA samples. The method can perform taxonomic assignment as well as quantify the respective abundance of detected taxa. In addition, we developed a method that can use our previous taxa assessment and perform their respective species identification. Both of these applications show that panmitogenomics is a groundbreaking novel framework to improve phylogenetic inferences for ancient DNA as well as ancient and modern environmental DNA.
A Pan-pangenome captures the full spectrum of genetic variation in humans, chimpanzees and bonobos

Joana L. Rocha

Joana L Rocha, Runyang Nicolas Lou, Juan M Vasquez, Alison Killilea, Stacy Li, Andrea Guarracino, Kendra Hoekzema, Evan E Eichler, Erik Garison, Matt W Mitchell, Peter H Sudmant

(USA), Coriell Institute for Medical Research (USA), Department of Genetics, Department of Genome Sciences (USA), Department of Integrative Biology (USA), Department of Molecular and Cell Biology (USA), Genomics, Howard Hughes Medical Institute (USA), Human Technopole (Italy), Memphis, UC Berkeley, USA (USA), University of California Berkeley (USA), University of Tennessee Health Science Center, University of Washington, University of Washington School of Medicine, and Informatics

Chimpanzees and bonobos (genus Pan) are our species closest living relatives. The genetic background of these species provides evolutionary context for capturing and representing the extensive structural diversity of human haplotypes. However, while long-read sequencing consortium projects such as T2T and HPRC are providing an unprecedented opportunity to decipher some of the most intractable and complex regions of the human genome, we have yet to fully understand the extent of genome structural diversity in the Pan genus. Here, we have assembled 60 haplotype-resolved genomes from a comprehensive sampling of a diverse set of chimpanzees and bonobos throughout their geographic range in Africa, to create a graph-based, telomere-to-telomere representation of Pan genomic diversity. This Pan-pangenome effort comprises some of the most contiguous assemblies to-date for non-human primates at a population scale. Our work complements and extends the recent human focused efforts to obtain a near-complete representation of the full spectrum of global genetic variation. Using these resources, we are characterizing the structure, composition, function, and evolutionary trajectories of genomic variation that is shared between humans, chimpanzees, and bonobos. We found that chimpanzees and bonobos have a higher length distribution of SVA retrotransposons and show extensive structural variation in regions associated with resistance to malaria resistance and sleeping sickness in humans. This will allow us to interrogate and test hypotheses on the adaptive significance of complex genomic structures that give way to shared genetic diversity across the primate lineage.
Pangenome assemblies reveal the evolution and recent in vivo activity of human LINE-1 retrotransposons

Rick McLaughlin

Lei Yang, Amanda Norseen, Rick McLaughlin
Pacific Northwest Research Institute (USA), University of Washington (USA)

LINE-1 retrotransposons dominate the human genome. As the most abundant and only active autonomous transposable elements in extant human genomes, LINE-1s affect human physiology and disease risk. Because LINE-1s are long and highly repetitive, our understanding of LINE-1 variation and recent evolution within diverse humans has been limited by the scarcity of haploid-resolved and long read-based assemblies. Here, we provide the first analysis of LINE-1 insertions, sequence variation, and recent activity using diverse, long read-resolved human genomes from the Human Pangenome Reference Consortium (HPRC). We found 13,045 LINE-1s with both intact open reading frames required for retrotransposition located at 671 unique insertion sites, and we infer ~3,000 such insertion sites in the human population. We constructed networks of these sequences which show evolutionary transitions from ancient, active groups of LINE-1s to many younger, currently active LINE-1s in modern humans. Using sequence-distances computed between these 671 LINE-1s and the 3’ transduction between LINE-1 insertion sites, we quantified the recent in vivo activity of each LINE-1, revealing the specific LINE-1s that have replicated at high rates in humans. Finally, our evolutionary analysis shows the emergence of new adaptive changes which we test in vitro to identify putative drivers of LINE-1 evolution in humans. These data constitute a novel analysis of the diversity and activity of the youngest human LINE-1s, made possible by an increasing number of diverse, long read-based human genome assemblies.
Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans

Runyang Nicolas Lou

Davide Bolognini, Alma Halgren, Runyang Nicolas Lou, Alessandro Raveane, Joana L. Rocha, Andrea Guarracino, Nicole Soranzo, Jason Chin, Erik Garrison, Peter H. Sudmant

Center for Computational Biology, Department of Genetics, Department of Integrative Biology, Foundation for Biological Data Science (USA), Genomics, Human Technopole (Italy), Human Technopole (USA), University of California Berkeley (USA), University of Tennessee Health Science Center (USA), and Informatics

The adoption of agriculture, first documented ~12,000 years ago in the Fertile Crescent, triggered a rapid shift toward starch-rich diets in human populations. Amylase genes facilitate starch digestion and increased salivary amylase copy number has been observed in some modern human populations with high starch intake, though evidence of recent selection is lacking. Here, using 52 long-read diploid assemblies and short read data from ~5,600 contemporary and ancient humans, we resolve the diversity, evolutionary history, and selective impact of structural variation at the amylase locus. We find that both salivary and pancreatic amylase genes have higher copy numbers in populations with agricultural subsistence compared to fishing, hunting, and pastoral groups. We identify 28 distinct amylase structural architectures and demonstrate that identical structures have arisen independently multiple times throughout recent human history. Using a pangenome graph-based approach to infer structural haplotypes across thousands of humans, we identify extensively duplicated haplotypes present at higher frequencies in modern day populations with traditionally agricultural diets. Leveraging 534 ancient human genomes we find that duplication-containing haplotypes (with 5 or more AMY copies in contrast to the ancestral 3 copies) have increased in frequency more than seven-fold over the last 12,000 years providing evidence for recent selection in Europeans at this locus comparable in magnitude to that at lactase. Together, our study highlights the strong impact of the agricultural revolution on human genomes and the importance of long-read sequencing in identifying signatures of selection at structurally complex loci.
300 Whole Genomes of People from the border of Ukraine and Romania
Taras K Oleksyk

Taras K Oleksyk, Walter W Wolfsberger, Khrystyna K Shchubelka
Oakland University - Biology (USA), Oakland University - Biology (Ukraine), Uzhhorod National University (Ukraine)

We present the results of the Whole Genome Sequencing of two human populations in the Carpathian Mountains region in Eastern Europe, specifically Ukraine’s Transcarpathia and Romania’s Satu Mare and Baia Mare provinces, areas previously underexplored in population genomics. The database contains the raw and annotated files of the whole genome sequences from 300 individuals from these regions, including annotations of common and unique genetic variants following a sampling protocol designed to capture the genetic diversity of Ukrainians and Romanians, including minority groups like Wallachians and Roma. The data is hosted on a dedicated web resource. We provide information on how to access to results of primary and secondary analysis of the data, including comparative analysis with previously published populations from Ukraine, and populations from the International Genome Sample Resource and Human Genome Diversity Project. The free research access to this database is contributing to a growing understanding of human genetic diversity in Central Europe. This effort emphasizes the potential for reuse of the generated data, advocating for open access to support future research in genomics, bioinformatics, and personalized medicine.
S29 - Advances in Machine Learning for Evolutionary Genomics.
A Deep Learning approach for HLA typing on ancient DNA data.

Alan Vladimir Godinez Plascencia

Alan Vladimir Godinez Plascencia, Andres Rivera Ramirez, Adriana Michel Avila Garcia, Carlos-Francisco Mendez-Cruz, Federico Sanchez Quinto
Center for Genomic Sciences - UNAM (México), Faculty of Sciences - UNAM (México), International Laboratory for Human Genome Research - UNAM (Mexico)

The allelic characterization of the Human Leukocyte Antigen system (HLA typing) has proven crucial to deepen our understanding of immunological function, population diversity and susceptibility to disease. Recently, computational methods have surged as a cost-effective strategy to carry out HLA typing for clinical and research purposes. Furthermore, novel advancements in ancient DNA (aDNA) recovery and sequencing techniques have uncovered important questions regarding the evolution of the HLA system, allowing researchers to potentially take a glimpse into the immunogenic makeup of individuals from the past. Nevertheless, the application of computational HLA typing on ancient samples still poses a significant challenge due to intrinsic properties of aDNA; such as being highly fragmented, the presence of cytosine deamination at the ends of sequencing reads, and low levels of endogenous DNA present. To address these difficulties, we present a Deep Learning architecture tailored for conducting HLA typing on aDNA data, which aims to predict database-derived labels corresponding to HLA haplotypes directly from query sequences in a mapping-agnostic manner. This strategy takes advantage of the vast amount of HLA sequence information that is broadly accessible, as well as the ability to incorporate damage patterns akin to those present in aDNA for the generation of a comprehensive training dataset. Our approximation fills the need of a fully-automated way to perform HLA typing that takes into consideration aDNA and its caveats. Overall, this work represents a significant step towards enhancing the suitability of computational methods to paleogenomics, facilitating greater insights into human evolutionary history.
Enhancers are cis-regulatory regions that orchestrate the precise control of spatiotemporal gene expression by mediating transcription factor (TF) interactions. Deciphering their sequence properties is challenging due to rapid evolution, low conservation, and binding complexity (1,2). We developed "Bag-of-Motifs" (BOM), a tree-ensemble–based machine-learning framework designed to identify cell type-specific enhancers using TF binding motif composition. BOM is an adaptable framework for dissecting the key TF binding motifs within regulatory sequences between specific cell types, states, and conditions. We use Shapley values, a concept rooted in game theory, to elucidate the individual contribution of binding motifs to the classification tasks (3). BOM excels in classifying candidate enhancers from multiple cell types and states across various species and developmental stages and outperforms deep learning models in the same classification task (auROC = 0.98). Shapley values also allow us to correlate motifs with transcription factor expression to gain a deeper insight into the motifs crucial for enhancer function and gene regulation. In conclusion, our 'Bag-of-Motifs' framework not only excels in classifying candidate enhancers across diverse biological contexts but also sheds light on the critical motifs governing enhancer function and gene regulation. 1. Smith, G. D., Ching, W. H., Cornejo-Páramo, P., & Wong, E. S.(2023). Decoding enhancer complexity with machine learning and high-throughput discovery. Genome Biology, 24(1),1-25. 2. Spitz, F., & Furlong, E. E.(2012). Transcription factors: from enhancer binding to developmental control. Nature reviews genetics, 13(9),613-626. 3. Shapley, L. S.(1953). A value for n-person games.
Ecolocator: A supervised machine learning model for location and climate-of-origin prediction

Jordan Rodriguez

Jordan Rodriguez, Richard Cronn, Andrew Kern
United States Forest Service (USA), University of Oregon (USA)

As the Earth’s climate changes, nearly every organism on the planet will shift its geographical range in response. For ecologically and economically important species, it is vital that scientists and managers be able to make informed decisions about future suitable environments for populations. In previous work, we have shown that supervised machine learning methods can predict origin from genomic variation using georeferenced training data. However, the spatial genetic variation signal that we are capturing may be the result of both neutral processes and local adaptation. Reasoning that the connection between genotype and environment mediated by local adaptation might be discernable from neutral variation, we introduce Ecolocator, a deep learning method that jointly predicts location of origin as well as environmental variables of that location. Ecolocator trains on a dataset of genotypes with known geographic location and associated climate variables such that it can predict the climate-of-origin and geographic location based on the genotype of an unlabeled sample. To validate our approach, we applied our method to one of the most important timber species in the world, Pseudotsuga menziesii var. menziesii, which has been shown to have strong local adaptation. Using a dataset of georeferenced genotypes and associated climate variables for those locations, we are able to perform accurate joint prediction of location and climate. Our open-source tool clarifies the biological implications of the connection between environmental variation and genetic variation so that informed conclusions regarding the conservation, adaptation, and resiliency of this species can be made.
Novel approaches for detecting ghost admixture
Martin Kuhlwilm

Xin Huang, Josef Hackl, Konstantina Cheshmedzhieva, Aigerim Rymbekova, Martin Kuhlwilm
University of Vienna (Austria)

Many populations experienced a complex history of admixture. In present-day humans, the availability of thousands of genomes allows to detect introgressed fragments with high confidence. However, gene flow has been suggested from lineages other than Neanderthals and Denisovans into humans, and ghost admixture has been found in our closest living relatives, the great apes. A comparative analysis has not been possible yet due to limits of the detection methods, which have often been tailored to the admixture history of modern humans, or even specific datasets. However, demographic histories and available datasets differ vastly between populations and species, introducing the need for more comprehensive tools. Here, we present a framework to detect introgressed fragments from unknown source populations based on previously applied strategies, using summary statistics or densities of private variants, as well as supervised machine learning algorithms. Furthermore, we implement different machine learning architectures, improving the performance specifically in datasets which substantially differ from present-day humans from Eurasia. In particular, this includes smaller sample sizes (on the order of tens of individuals), unphased data, and approximate instead of fine-tuned demographic models. A modularized framework allows versatile treatment of input data, and flexible handling of different datasets. Under this framework, it becomes possible to choose the most appropriate strategy for a given dataset, evaluate the reliability of the results based on built-in simulation functions, and generate fine-grained maps of introgressed fragments. This will open new avenues into the exploration of ghost introgression genomic landscapes.
Graph embeddings for population genetics data visualization

Micaela Long Grosso

One of the challenges in population genetics data analysis is high dimensionality. Dimensionality reduction via principal components analysis (PCA) is commonly used for visualization of genetic variation and is usually used to make inferences on population structure in combination with other methods such as ADMIXTURE. However, PCA focuses on capturing global linear relationships in the data, and as local variation is often found on higher principal components, a two-dimensional plot does not always capture all the relevant genetic variation. Non-linear methods such as t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) have become very popular as visualization and dimensionality reduction tools. Although these algorithms emphasize local patterns and complex relationships, they usually fail at preserving global structure, because distances between clusters are not always meaningful and visualizations are usually difficult to interpret. In this work, we explore the performance of Node2Vec, an embedding algorithm that generates vector representations of nodes in a graph. Node2Vec learns low-dimensional representations for nodes, based on random walks in the graph, following the idea that nodes on a graph can be treated like sentences in a corpus. Relative to the other algorithms, Node2Vec appears to better preserve global geometry, showing results that are similar to those obtained with PCA, with an improvement in clustering groups with small sample sizes.
Applications of deep learning in population genetics typically take as input the genotype matrix or compressed summaries thereof. While these methods have proven powerful and versatile, a major hurdle is that they do not scale well to biobank-size datasets which are currently being generated. An alternative lies in using the ancestral recombination graph (ARG), which encodes the transmission of genetic material from ancestors to extant individuals across the genome. The ARG can solve many issues in population genetic inference because it is: (i) sufficient, in that it contains all the information obtainable from the genotype matrix, (ii) compact, in that shared genetic variation is stored as relationships between ancestral haplotypes, and (iii) multiscale, in the sense that the span of ancestral haplotypes and edges naturally reflects the spatial (along the genome) and temporal scales at which shared ancestry impacts variation, without arbitrary discretization into windows. We developed an architecture, tsNN, that takes ARGs (as tree sequences) as input for evolutionary inference by borrowing ideas from the field of graph neural networks (GNNs). The sequence-like nature of the ARG is captured by updating the node embedding with edge insertions and deletions. We applied tsNN to two tasks: dating mutations and inferring demographic parameters. tsNN outperforms a naive GNN despite having fewer learnable parameters, demonstrating that the biology-informed inductive bias can improve learning. Further, tsNN performs better at dating mutations than the current model-based methods. Taken together, our results demonstrate the utility of graph neural networks in ARG-based inference.
Telomeres, the protective caps at the ends of chromosomes, consist of a repetitive nucleotide sequence (TTAGGG) in humans and play a crucial role in chromosomal replication. With each cell division, telomeres shorten due to the end-replication problem. Telomere shortening is therefore associated with age-related diseases, early mortality, and has implications for cancer biology, particularly in the context of cellular senescence and tumor progression. Our preliminary analyses of Telomere Length Variation (TLV) in the 1000 Genomes Project data indicate significant variation in Telomere Length (TL) among human populations, between sexes, and age classes. A recent study (Burren et al., 2023) also performed a large-scale estimation of telomere length variation across the UK BioBank data and identified several associated rare variants that correlate with myeloid cancer precursors. Here we use data from The Cancer Genome Atlas (TCGA) in congruence with variants identified by Burren et al., 2023 to achieve three primary objectives: 1. Quantify TLV across different cancer types and tissues. 2. Identify genetic variants associated with telomere length variation, spanning both cancer and non-cancer genomes. 3. Utilize estimated telomere lengths from the whole-genome sequencing (WGS) data, rare variant data, and phenotypic data available on TCGA to train a supervised machine learning model that predicts tumor status (cancer versus non-cancer).
Interpreting deep learning methods for population genetic inference
Sara Mathieson

Presented by self
Haverford College (USA)

The broad field of generative AI has captured a worldwide audience with novel text, image, audio, and video. In evolutionary biology we have used synthetic data for decades, but created using custom simulations from evolutionary models. However, it is often challenging to create realistic simulations for populations with unique histories, which compromises the results of downstream analyses. Recently, new methods inspired by the generative AI literature have emerged as ways to automatically adapt simulations to mirror the real genomic data of any population or species. Here I will focus specifically on interpreting generative methods, including a GAN (generative adversarial network) for inferring demographic and natural selection parameters. Our method, called pg-gan, works by training a parametric generator and CNN (convolutional neural network) discriminator in concert, until there is a close match between real and simulated data. The main training algorithm requires only fast neutral simulations, which allows us to fit a demographic model. We then save the resulting discriminator and finetune it with minimal simulations of selection, which are often computationally expensive. The trained CNNs (before and after fine-tuning) produce two outputs for each region of held aside genomic data: probability of real (vs. simulated) and probability of selection (vs. neutral). We show how to interpret these networks and their output through their correlation patterns with known summary statistics. The results demonstrate implicit calculations performed by the CNN and allow us to identify interpretable features that make genomic data appear realistic.
Discovering genotype-phenotype relationships with machine learning and the Visual Physiology Opsin Database (VPOD)

Seth A Frazer

Seth A Frazer, Todd H Oakley, Mahdi Baghbanzadeh, Ali Rahnavard, Keith A Crandall
Computational Biology Institute, Department of Biostatistics and Bioinformatics, Department of Invertebrate Zoology, Milken Institute School of Public Health, National Museum of Natural History, Smithsonian Institution (USA), The George Washington University (USA), UC Santa Barbara (USA)

Predicting phenotypes from genetic variation is foundational for fields as diverse as bioengineering and global change biology, highlighting the importance of efficient methods to predict gene functions. Linking genetic and phenotypic changes has been a goal of decades of experimental work, especially for some model gene families including light-sensitive opsins proteins. Opsins can be expressed in vitro to measure light absorption parameters, including \( \lambda_{\text{max}} \) - the wavelength of maximum absorbance - which strongly affects organismal phenotypes like color vision. Despite extensive research on opsins, the data remain dispersed, uncompiled, and often challenging to access, thereby precluding systematic and comprehensive analyses of the intricate relationships between genotype and phenotype. Here, we report a newly compiled database of all heterologously expressed opsin genes with \( \lambda_{\text{max}} \) phenotypes called the Visual Physiology Opsin Database (VPOD). VPOD_1.0 contains 864 unique opsin genotypes and corresponding \( \lambda_{\text{max}} \) phenotypes collected across all animals from 73 separate publications. We use VPOD data and deepBreaks to show regression-based machine learning (ML) models reliably predict \( \lambda_{\text{max}} \), account for non-additive effects of mutations on function, and identify functionally critical amino acid sites. The ability to reliably predict functions from sequences alone using ML will allow robust exploration of molecular-evolutionary patterns governing phenotype, inform functional and evolutionary connections to an organism's ecological niche, and may be used broadly for de-novo protein design. Together, our database, phenotype predictions, and model comparisons lay the groundwork for future research applicable to families of genes with quantifiable and comparable phenotypes.
Metric learning for phylogenetic placement
Siavash Mirarab

Presented by self
San Diego (USA), University of California

NA
Interrogating the relationship between conservation and regulatory function across all human promoters using machine-learning variant effect predictions

Steven Reilly

John C Butts, Stephen Rong, Rodrigo I Castro, Sager J Gosai, Mackenzie Noon, Rohit Ghosh, Ryan Tewhey, Steven K Reilly
Dept. of Genetics (USA), Graduate School of Biomedical Sciences, Graduate School of Biomedical Sciences and Engineering, The Broad Institute (USA), The Jackson Laboratory (USA), Tufts University School of Medicine (USA), University of Maine (USA), Yale Center for Genomics Health (USA), Yale University (USA), Yale University School of Medicine

The ability to predict an allele's regulatory function would improve our identification of causal, non-coding variants underlying human traits and disease. Evolutionary conservation is commonly used as a functional proxy, but the relationship between conservation and regulatory impact are poorly understood. While massively parallel reporter assays (MPRAs) are powerful tools to identify functional regulatory variants, they can maximally assay ~100K genetic variants at a time, limiting their applicability to a small subset of currently observed human variation. Our recently developed deep learning model, Malinois, accurately predicts empirical MPRA results (R>0.89) in multiple cell types and is scalable to the whole genome. Here we present the first comprehensive analysis of regulatory function and evolutionary conservation at the single nucleotide level in all promoters. We generate saturation mutagenesis MPRA predictions for ~14M SNVs in promoters of all protein-coding genes and provide these functional scores for all promoters as prospective maps for variant interpretation. Genome-wide, we find weak purifying selection against both loss-of-function (LoF) and gain-of-function (GoF) promoter variants, even for genes that show no evidence of coding LoF-intolerance. Some promoters show strong purifying selection against LoF variants (e.g. LDLR and U2AF2), GoF (e.g. LCP2), or both (e.g. MAG), indicative of different modes of selection acting on promoter activity. Finally, we find recent purifying selection against LoF and GoF promoter variants using gnomAD. We deploy these maps to interpret rare variant allelic series in biobanks and identify causal variation underlying human selection scans.
Sequence embeddings by pretrained language model indicate adaptive convergence of high-order protein features.

Zhengting Zou

Zhenqiu Cap, Zhengting Zou
Chinese Academy of Sciences (China), Institute of Zoology, University of Science and Technology of China (China)

Various computational methods have been proposed to investigate the genetic basis for organismal functional convergence, specifically focusing on convergence of amino acid state at individual sites in proteins. However, beside site-level convergence, convergence of high-order features in protein sequence can also lead to functional similarity. In this study, we derive fixed-length embeddings of protein sequences by a pretrained protein language model (PLM), and propose that these embeddings reflect high-order features of the proteins. We investigated multiple cases of distantly related but functionally similar protein orthologs putatively having experienced adaptive convergence during evolution, e.g. proteins in hyperthermophilic bacteria and archaea species. We found that in most cases, these proteins have similar embeddings despite their site-level phylogenetic divergence, supporting that PLM embeddings can reflect high-order feature convergence of proteins. We then proposed a convergence test pipeline PEAC (protein embedding adaptive convergence) for multiple lineages of species with functional convergence, based on comparing the PLM embedding similarity of proteins with simulated neutral background. Using PEAC in a genome-wide case study of echolocating mammals, we discovered proteins with putative adaptive convergence that are not reported previously, alongside with known convergent cases such as Prestin.
S30 - New computational approaches to estimate past demographic events and natural selection.
Patterns of linkage disequilibrium contain a wealth of information about the evolutionary forces acting within a population. Much of our existing understanding has focused on pairwise correlations between alleles, and how they decay as a function of the recombination rate. However, the magnitudes of these correlations are strongly influenced by other evolutionary forces like natural selection and genetic drift, making it difficult to tease out the effects of recombination (or vice versa). Here we introduce a theoretical framework for studying an alternative class of summary statistics that explicitly quantify the homoplasy produced by recombination. We derive analytical expressions that predict how these homoplasy metrics depend on the rates of recombination and recurrent mutation, the strength of selection and genetic drift, and the present-day frequencies of the two alleles. We find that the degree of homoplasy can strongly depend on this frequency scale, which reflects the underlying timescales over which these mutations occurred. We show how these scaling properties can be used to disentangle the rates of recombination, and discuss their implications for understanding the rates of horizontal gene transfer in bacteria.
Minus the error: making selection detection robust to residual alignment errors

Avery Selberg, Nathan Clark, Maria Chikina, Sergei L. Kosakovsky Pond
Temple University (USA), University of Pittsburgh (USA)

Advancements in computational methods for genomic data analyses are crucial for unraveling the evolutionary history and adaptation of diverse species, yet errors in multiple sequence alignments (MSAs) are known to bias many comparative methods. In the context of natural selection analyses, specifically codon evolutionary models, excessive rates of false positives result. A characteristic "signature" of error-driven findings is unrealistically high estimates of dN/dS (e.g., >100), affecting only a small fraction (e.g., ~0.1%) of the alignment. Despite the widespread use of codon models to assess alignment quality, their potential for error correction remains unexplored. We present BUSTED-E ("minus the error"): a novel method designed to detect positive selection while concurrently identifying alignment errors. This method is a straightforward adaptation of the BUSTED flexible branch-site random effects model used to fit distributions of dN/dS, with an important modification: it integrates an error-sink component representing an "abiological" evolutionary regime (dN/dS > 100). With the error-sink component, the need for alignment pre-filtering is eliminated. In large-scale analyses, BUSTED-E reduces unrealistic rates of positive selection detection and provides the option for masking errors in the MSA for downstream analyses. We conducted a genome-wide scan of avian and mammalian genes utilizing BUSTED-E, resulting in a substantial decrease in the percentage of positively selected genes (from 40% to 4%), which were notably enriched in immune-related functions—highlighting BUSTED-E’s ability to detect meaningful biological patterns even in the presence of alignment errors.
NeMu: A Comprehensive Pipeline for Accurate Reconstruction of Neutral Mutation Spectra from Evolutionary Data

Bogdan Efimenko

Bogdan Efimenko, Konstantin Popadin, Konstantin Gunbin
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University (Russian Federation)

One of the most important characteristics of each contemporary model of molecular evolution is the assumption that mutations occur in a constant manner. However, in nature mutations are determined by the interaction of DNA replication and repair efficiency, which are highly diverse because of environmental changes like exposure to mutagens or naturally selected changes in DNA repair. Although the importance of mutational processes is well-recognized in human cancer research, progress in the reconstruction and utilization of mutational spectra at comparative- or intra-species levels remains limited due to data scarcity. Traditional mutation accumulation experiments used to estimate neutral mutational spectra are not only laborious and generally confined to model organisms, but they also might not be entirely neutral, thus impeding progress in filling this knowledge gap. To address this shortfall, we present the NeMu pipeline, available at https://nemu-pipeline.streamlit.app. By harnessing the vast amount of sequences from existing databases, NeMu efficiently and accurately reconstructs the neutral mutational spectrum. The pipeline encompasses multiple steps: automated sequence sampling from nucleotide databases, phylogeny and precise ancestral states reconstruction, extraction of substitutions and inference of the neutral spectrum. This efficient tool enables the identification of species-specific mutational signatures, advancing research into species-specific spectrum variation and allowing in future for their integration into more realistic evolutionary models for understanding drivers of evolution. This project was supported by the Russian Science Foundation (RSF) grant ? 21-75-20143
A genealogy-based framework to estimate population structure and demographic history

Charleston W.K. Chiang

Charleston W.K. Chiang
University of Southern California (USA)

Many methods infer population structure or demographic history from genetic data using relatively low-dimensional summaries, such as the allele frequency spectra (AFS). In principle, the complete genomic history is encoded in the sequence of gene-genealogical trees, known as the ancestral recombination graph (ARG). As a step toward utilizing all available genomic information, we introduce two methods that leverage the ARG to infer population structure and demographic histories. First, we describe a framework to infer the expected relatedness between pairs of individuals given an ARG of the sample, which we call the eGRM. We show that the eGRM captures the structure of a population better than the canonical Genetic Relationship Matrix (GRM), even when only genotyping array data is available. Second, we devised a method called gLike, which employs a graph-based structure to summarize the relationships among all lineages in a tree with all possible trajectories of population memberships through time and to efficiently compute the probability under any parameterized demographic model. Through extensive simulations of multiple admixtures, we showed that gLike accurately estimates dozens of parameters such as ancestral population sizes and admixture timing, and outperforms conventional approaches based on the AFS. We applied both methods to real-world human genomic data from Finnish, Latino American, and Native Hawaiian cohorts to gain further insights into the patterns of population structure and to estimate parameters of the admixture histories. Taken together, our studies demonstrate the power of leveraging the genealogical trees for downstream population genetic inferences.
Evolutionary Analysis of Plasmodium falciparum Across the Asian and African continents
Daniel Garcia Ruiz

Daniel Garcia Ruiz, Hsiao-Han Chang
(Taiwan), Academia Sinica (Taiwan), Bioinformatics Program, College of Life Sciences and Medicine, Institute of Bioinformatics and Structural Biology, Institute of Statistical Science, National Tsing Hua University, National Tsing Hua University (Taiwan), Taiwan International Graduate Program

Malaria remains one of the most burdening infectious diseases despite control programs, the development of vaccines, and the availability of anti-malaria drugs. While the majority of malaria cases occur in the African continent, it is noteworthy that resistance to several anti-malaria drugs initially emerged in Asia. The current study aims to understand Plasmodium falciparum’s evolutionary history, including population structure and natural selection, by performing a comparative analysis between the African and Asian continents. Variant information from the Plasmodium falciparum dataset 7 was used to estimate nucleotide diversity between populations and to perform principal component analysis (PCA) and will be used to estimate pairwise dN/dS. Antigenic genes and genes involved in drug resistance show different patterns of nucleotide diversity across different continents. Kelch 13 and MDR1, genes associated with artemisinin and multi-drug resistance as well as the antigenic genes SERA2 and TRAP show reduced diversity in the Asian continent. Contrarily, PCA shows that most of the variation occurs in the Asian continent, specifically in the Greater Meakong subregion. Our preliminary results indicate the presence of distinct evolutionary pressures in the Asian and African continents, however, dN/dS estimation is needed to better characterize the selection forces acting upon Plasmodium falciparum across different continents. We anticipate that our study will provide valuable insights to address the challenges posed by drug resistance in malaria treatment and elimination efforts.
Detecting natural selection in Holocene Europe with multi-locus genotype identity scans
Devansh Pandey

Devansh Pandey, Mariana Harris, Nandita R Garud, Vagheesh M Narasimhan
The University of Texas at Austin (USA), University of California Los Angeles (USA)

Large sample size ancient DNA studies now offer insights into human evolution by examining changes in genomes over time. Recent efforts using time-stratified aDNA for selection scans mainly focused on single locus approaches. As different types of statistics often provide complementary information about selective events, we carried out multi-locus genotype scans for natural selection on a time transect of 708 samples over the past 7,000 years of European history. We used the G12 statistic that had been designed for unphased diploid data and validated its performance on aDNA using simulations. We incorporated the missingness, ascertainment bias, DNA damage, and unphased random allele calling typical of aDNA processing utilizing a demographic model that captures broad features of the allele frequency spectrum of European genomes. With these simulations as well as using positive control loci that have been previously identified and functionally validated in modern Europeans, we found that the multi-locus statistic G12 performed significantly better than an allele frequency-based selection statistic, SweeepFinder2 that has also been recently used on ancient DNA. Applying our approach to the ancient DNA time transect, we identified 14 signals of selection across the four time periods with half of the observed signals seen in the earliest time period, not detectable in the latest time period. Taken together, our results suggest that selective events in European prehistory, including from times when humans first began to domesticate animals, have been obscured by neutral processes such as drift or by demographic events including admixture.
Hidden Diffusion: Accurate Joint Inference of Selection and Demographic History from Time-series Data for Multiple Populations
Ekaterina Noskova

Ekaterina Noskova, Madleina Caduff, Andreas Fueglistaler, Anna Parker, Christoph Leuenberger, Daniel Wegmann
Swiss Institute of Bioinformatics (Switzerland), University of Fribourg (Switzerland)

Recent advancements in sequencing technologies and ancient DNA preparation have enabled the generation of extensive time-series data—genetic snapshots at various time points—that promise to identify loci affected by selection. Time-series allele frequency data is effectively modeled with Hidden Markov Models (HMM). Inference under Wright-Fisher HMMs involves calculating transition probabilities between frequency states and data likelihood given those probabilities. However, due to the large number of possible allele frequency states, analytical inference is computationally challenging. Several methods address this challenge: ApproxWF uses a hidden Markov model on a coarse allele frequency grid, while TimeSweeper is machine learning method trained on simulated data under explicit demographic histories. We introduce a novel method based on hidden diffusion, employing a single partial differential equation solver to calculate both transition probabilities under Wright-Fisher diffusion and likelihood. This approach is faster than ApproxWF and more accurate than TimeSweeper in jointly inferring demographic parameters and locus-specific selection in a single population. Moreover, hidden diffusion identifies loci under selection in models involving multiple populations connected by migration and with potentially different selection coefficients. It can identify loci under divergent selection between populations, where single-population time-series data may be uninformative. Implemented in user-friendly software, our method also models auto-correlation in selection coefficients along the genome using a second-layer HMM to account for linkage between neighboring loci. We demonstrate its utility by identifying selection targets from a time-series of ancient human samples from southern Germany spanning from 2000 BC to modern times.
Joint modeling of demography and background selection reveals complex patterns of polymorphism

Gustavo Valadares Barroso

Gustavo Valadares Barroso, Aaron Ragsdale
University of Wisconsin-Madison (USA)

Classic background selection theory describes the deterministic effect of strong negative selection on diversity at linked neutral sites. However, the model breaks down when drift allows deleterious variants to segregate more freely. Recent models have incorporated weak selection, but predictions remain inaccurate for 2N_s values in the vicinity of ~4 (the peak of interference among deleterious variants). This is also the most interesting region of parameter space, where the reduction in diversity at neutral sites (B-values) is strongest. Therefore, reliable models of linked selection depend on accurate prediction of B-values under interference. Here we develop a new method (moments++) to achieve this goal, based on the Hill-Robertson system of two-locus statistics. We model the effect of selection acting on weakly and moderately constrained loci linked to a neutral locus at arbitrary recombination distance. We treat the joint effect of multiple selected loci independently, but we show interference among them is well captured through local rescaling of mutation, recombination and selection in an iterative procedure that converges quickly. We demonstrate by forward simulations that our formulation leads to B-value predictions that are substantially more accurate than the current state-of-the-art, bridging the gap between weak and strong selection. Moreover, moments++ is more flexible than existing methods and supports any demographic history, as well as mutation and recombination maps. We couple complex demography with variable recombination, mutation and selection intensities along the genome to show that patterns of diversity can be misinterpreted by other models.
A fundamental goal of evolutionary biology is to identify drivers of natural selection. Existing codon evolutionary models examine selection on genes by comparing the rates of synonymous (dS) and nonsynonymous substitutions (dN). However, such models overlook the nuanced variation in the relative rates for individual synonymous substitutions, driven by selection on codon usage. Crucially, the relative rates are known to correlate with key biological traits like gene expression, and can inform researchers about the underlying factors driving selection. To explore selection on codon usage, we introduce the Multiclass Synonymous Substitution (MSS) model, a novel codon evolutionary model designed to separate the monolithic synonymous rate class into its individual components. The MSS model estimates the relative rate for each pair of synonymous codons, enabling precise comparison of synonymous substitution rates within a gene or across a set of genes. Ranking synonymous pairs by their relative rates allows us to explore which pairs are under selection (i.e., those with a substitution rate <1). We applied the MSS model to several well-studied genuses with prior evidence of selection on codon usage. The MSS ranking of synonymous substitution rates in Drosophila aligns with average SNP density across codon pairs, an independent measure of selection on codon usage. Our model also recapitulates different substitution patterns in highly divergent taxonomic groups, such as Enterobacteria, Caenorhabditis, and Saccharomyces. With the MSS model, we can now study selection on synonymous substitutions in diverse taxa, independent of any a priori assumptions about the forces driving that selection.
Considerations for Inferring Demography and Detecting Selection in Understudies Species

Jazlyn Mooney

Presented by self
University of Southern California (USA)

As we progress into the genomic era, genetic and genomic datasets are being generated for species of conservation concern. These data are often utilized to detect evidence of natural selection and to infer changes in effective population size (Ne) over time and to estimate current Ne. However, the majority of species of conservation concern lack a reference genome, which greatly limits evolutionary inference. This research demonstrates the utility of genomic data for inferring natural selection and demographic histories using Canidae. First, I will discuss research conducted on Ethiopian wolves, a canid species endemic to the Ethiopian Highlands that has been steadily declining in numbers for decades. Currently, out of 37 extant species, it is one of the world’s most endangered canids. Previous conservation efforts have focused on preventing disease, monitoring movements and behavior, and assessing the geographic ranges of sub-populations. We added an essential layer by determining the Ethiopian wolf’s demographic and evolutionary history using high-coverage whole-genome sequencing data from 10 Ethiopian wolves from the Bale Mountains. Then, I will discuss recent efforts to re-infer the demographic history of North American gray foxes, Urocyon cinereoargenteus. I will compare the demographic history and other frequently inferred summary statistics when using a conspecific and non-conspecific reference genome. Our disparate results highlight the potential pitfalls of mapping to a divergent species and subsequently conducting genomic analyses.
The impact of natural selection on patterns of genetic diversity
Jesús Abad Guzmán-López

Jesús A. Guzmán-López, Parul Johri, Diego Ortega-Del Vecchyo
Chapel 11 Hill, Chapel Hill, Department of Biology, Laboratorio Internacional de Investigación sobre el Genoma Humano, México (Mexico), NC 27599 (USA), NC 27599 Department of Genetics, NC 27599 Integrative Program for Biological & Genome Sciences, Querétaro, Santiago de Querétaro, Universidad Nacional Autónoma de México, University of North Carolina

Synonymous mutations are changes in protein-coding DNA that do not modify the amino acid; these mutations often occur in the third base of a codon due to the genetic code redundancy. Hence, synonymous mutations are assumed to have a neutral or nearly neutral effect on the fitness of a population. Evidence has shown that a high proportion of synonymous sites are under strong selection in species such as Drosophila melanogaster. Here we perform extensive genomic simulations with the software SLiM and subject them to different degrees of selection, to analyze the impact of natural selection on various summaries of genetic diversity such as the site frequency spectrum (SFS). We discuss the impact of our results on estimates of genetic diversity in various species and compare them with empirical data.
Extending Flex-sweep with Random Forest to infer time and strength of selection
Jesus Murga-Moreno

Jesus Murga-Moreno, David Enard
University of Arizona (USA)

The contribution of selective sweeps to genomic adaptation remains an active area of research in population genetics. Machine learning (ML) has recently surpassed most typical approaches, such as classical summary statistics, maximum likelihood, and ABC. Although many existing methods can detect selective sweeps, most are statistically limited to specific sweep types, and subtle signals are particularly challenging to detect. One method, Flex-sweep, is a Swiss-knife army that detects sweeps of different types, strengths, and ages, allowing us to better discern past selective pressures such as ancient epidemics. Although Flex-sweep and other ML approaches can detect and classify sweeps, only a few methods can quantify the strength and timing of selection. We extend Flex-sweep with Flex-date, a Random Forest approach to infer the strength and timing of selection. We exploit Flex-sweep’s classifications, summary statistics, and theoretical differences between sweep to reduce the prior space parameter when performing the inferences. Because different summary statistics capture different sweep patterns, our method relies on a two-step approach: i) classify Flex-sweep outputs by age and ii) predict parameters through the selected sweep age model. By prioritizing each model’s most important summary statistics, we manage the high-dimensional data and focus on the most informative statistics for further parameter estimation. Flex-date performs robust inferences on the strength and timing of selective sweeps covering ancient, complete and partial sweeps. We compared Flex-date to recent methods for inferring sweep parameters including ARG-based methods.
Inference of ultra recent demography, recurrent mutation, and natural selection from biobank-scale site frequency spectra

Joshua G Schraiber

Joshua G Schraiber, Michael D Edge
University of Southern California (USA)

Modern population genetic datasets have sample sizes in the tens or even hundreds of thousands of individuals. With such large sample sizes, many mutations will have multiple independent origins and the standard infinite sites assumption of population genetics becomes untenable. In addition, the vast majority of variants are extremely rare, and therefore sensitive to recent demographic history. Moreover, discovery of ultra rare variants provides an unprecedented window into the action of natural selection. Because of this, biobank-scale data presents an unprecedented opportunity to learn about recurrent mutation, recent demography and natural selection. We developed a new approach to modeling the site frequency spectrum of ultra rare variants (frequency less than 0.1%) in biobank scale datasets (>10000 samples). Our approach leverages the branching process approximation to the Wright-Fisher process. We show that finite sample probabilities follow an inhomogeneous birth death process with immigration and present an interpretable analytical solution that allows for efficient computation. When applied to data from gnomAD, we show that recurrent mutation common, and estimate rates of population growth in the last ~100 generations. We then estimate a distribution of fitness effects across different functional categories. As examples of downstream applications, we present a method for estimating the number of independent origins of a mutation subject to natural selection, and find that many rare, deleterious mutations have arisen multiple times independently in humans.
Bayesian mediation models reveal patterns of host genomic disease adaptation driven by environment

M. Elise Lauterbur

M. Elise Lauterbur, David Enard
University of Arizona (USA)

Global patterns of pathogen prevalence are shifting in response to climate change and other human activity. Traditional epidemiological and disease ecological approaches used to examine these impacts are typically restricted to a modern timescale and narrow scope of host, pathogen, or environment diversity. By using host genomes as a record of past interactions between hosts and pathogens, we can extend disease ecological analyses to investigate these relationships over evolutionary time in the context of the host’s environment. We develop this eco-evolutionary framework applied to a dataset of genomes from over 200 host mammal species occupying a wide range of habitats. We quantify the impact of virus exposure across species and environments using genomic adaptation at virus-interacting proteins (VIPs) with a novel application of a Bayesian mediation model structure. This allows linking ecological and environmental factors with viral impact across evolutionary time, controlling for genome-wide adaptation. We show that there are broad, general effects of ecological factors including precipitation seasonality, host diet, and viral transmission mode on the impact of virus exposure as measured through host genomic disease adaptation. This work has important implications for understanding the distribution of viruses across host and virus types worldwide, and for predicting future changes in virus exposure both to humans and other organisms in response to changing environments. In addition, this framework can be generalized to other questions for which genomic adaptation is an expected response that can be correlated to ecological conditions.
In population genomics, there are a dizzying array of potential data analysis approaches to infer population history, aspects of natural selection, or other evolutionary properties from data. Although methods developers try to evaluate their approaches, those evaluations can be unconsciously biased or may not reflect the experiences of real-world users. GHIST aims to create an annual forum in which the community can test inference approaches in an unbiased fashion. Each year, the GHIST organizers will release simulated population genomic data sets and host a competition to infer various aspects of the processes that generated those data. From the competitors’ solutions, the community will learn what approaches perform well or poorly in particular circumstances. Our first competition is now live, with four demographic history inference tasks of escalating expected difficulty. We look forward to your submissions!
Amino acid physicochemical property convergence may underlie adaptive functional convergence.
Shanshan Chen

Shanshan Chen, Zhengting Zou
Chinese Academy of Sciences (China), Key Laboratory of Zoological Systematics and Evolution Institute of Zoology, University of Chinese Academy of Sciences (China)

Many studies have proposed various methods to detect the molecular basis for adaptive functional convergences between species, often attributing them to the convergence of amino acid states in functionally related proteins. Here we group amino acids into physicochemically similar property classes, to test whether convergence of amino acid properties, not states, may also underlie organismal or functional adaptive convergence. Investigating three previous case studies of different species and methodology, we found functionally related genes with putatively adaptive property convergence, providing evidence for our hypothesis, and emphasizing the importance of considering amino acid property when studying adaptive sequence convergence.
Integrating epidemiological and population genetic models to gain insights into Plasmodium vivax biology

Shyamalika Gopalan

Shyamalika Gopalan, Jillian Grassia, Amy Goldberg
Clemson University (USA), Duke University (USA)

Genomic data are increasingly being generated to understand the evolution and epidemiology of human pathogens. Despite progress in studies of non-recombining viruses and bacteria, gleaning actionable insights from eukaryotic parasite data remains difficult, largely due to their complex life cycles. Here, we develop a novel framework for jointly modeling the epidemiological dynamics and genomic evolution of the malaria parasite Plasmodium vivax. We build a two-step model across the parasite's haploid and diploid stages within the human and mosquito, respectively. First, we use a stochastic Ross-Macdonald-style model to simulate a series of human-mosquito transmission events under a given set of epidemiological parameters. Then, we use SLiM to simulate the within-human and within-mosquito stages of parasite genome evolution as they recapitulate these transmission events, including recurrent transmission bottlenecks and periods of exponential growth. By capturing the processes generating neutral genetic variation in parasites, we are able to explain empirically observed phenomena, test the relative contributions of life stages to summaries of variation, and provide neutral baselines for common summary statistics. For example, for parasite populations thought to be at constant population size, most empirical studies observe strongly negative genome-wide average Tajima’s D values. We show that, unlike for human populations, this is an expected feature under a neutral model of P. vivax evolution. Our findings have important implications for the interpretation of summary statistics and overall patterns of genetic variation used for the inference of demography and selection in eukaryotic pathogens.
Both demography and Background Selection (BGS) have profound influence on the diversity patterns along the genome. Demography has a consistent genome-wide effect while the effect of BGS on diversity depends on the local strength of selection and linkage patterns. However, the joint effects of demography and BGS are not well understood: the traditional theory of BGS leading to local diversity reduction maps (B-maps), assumes a constant population size. Also, almost all coalescent-based demographic inference methods (e.g., standard PSMC and MSMC) cannot incorporate BGS effects. Here we investigate a new theory of a distorted coalescence process that incorporates BGS and we implement an inference method for ARGs that accommodate known genomic maps of BGS. We show that this method can lead to accurate coalescent-based demography inference in the presence of BGS. We also show that ARG inference allows accurate fine-scale (kb scale) estimation of diversity levels, which otherwise is difficult to obtain from genotype data. Fine-scale diversity also facilitates solutions to the reverse problem: learning the full BGS landscape from observed diversity patterns, rather than summarizing them as Distribution of Fitness Effects (DFE). Joint modeling of demography and BGS has the potential to provide estimates of demography under BGS, obtain more accurate fine-scale B-maps, and learn region-specific BGS coefficients. We illustrate the utility of the method by exploring the effect of BGS on inference of population size changes in different populations and investigate their differential selection landscapes, using published human genomic data.
S31 - Haplotype-based methods and frameworks for inference of evolutionary history
Ancestry-based approaches for inference of evolutionary history on short timescales

Amy Goldberg

Presented by self
Duke University (USA)

One of the major insights from the genomic era is the ubiquity of migration and admixture throughout human history, producing mosaic-like ancestry along genomes. Admixed populations may be particularly informative about recent evolution because of an added layer of information: patterns of genetic ancestry. Further, the admixture process may facilitate adaptation on short timescales through introduction of already-adaptive loci. But, genomic analyses are limited by the ways we summarize ancestry data, often discarding most of the genome and requiring prior knowledge to choose statistics. Instead, we transformed genomic data into new forms that preserve the full structure and information: images. We could then train neural networks on images of chromosomes painted by ancestry to identify adaptive loci. This opens a suite of image-based machine learning tools for genomics. Our method localizes the adaptive variant to a region about ~1/3 the length of prior approaches, resulting in fewer genes to test for function and improved interpretation. Applied to humans from Cabo Verde, we detect adaptation to malaria that reduced the pool of susceptible individuals by ~half over the last 600 years. Beyond adaptation, we combined patterns of ancestry with other summaries of variation to investigate the demographic history of Cabo Verdeans. We found signatures of assortative mating by ancestry that bias inference of the timing of admixture. Finally, based on this observation of ancestry-assortative mating, we conducted a series of forward simulations to link mechanistic mate-choice models to the consequences of assortative mating on patterns of global and local ancestry.
Reconstructing parent genotypes at genotyping array accuracy using siblings and other relatives

Amy L. Williams

Ying Qiao, Ethan M Jewett, Kimberly F McManus, William A Freyman, John Blangero, The 23andMe Research Team, Amy L Williams
23andMe, Cornell University (USA), Inc. (USA), The University of Texas Rio Grande Valley (USA)

Children inherit two chromosome copies, one from each parent, with both formed via recombination. In expectation, n siblings inherit a proportion of 1-1/2n of both parents’ genomes, providing rich information about their parents’ DNA. Reconstructing ungenotyped parents can increase power in genome-wide association studies, reveal parents’ ancestral origins, and improve relatedness inference. We developed a method to reconstruct the genotypes of parents from full siblings using a combination of family-based phasing, sex-specific genetic maps to infer parent sexes, and identity by descent (IBD) sharing to close relatives. Specifically, HAPI2 jointly phases siblings, building a one-generation ancestral recombination graph (ARG), and implicitly reconstructs the parents’ DNA. Given data for 77 siblings, HAPI2 often provides chromosome-scale haplotypes for the parents. For fewer siblings, we combine phase information with IBD between the reconstructed parent haplotypes and genotyped relatives. This disambiguates parental assignment of the DNA within and between chromosomes since these IBD segments are typically inherited through only one parent. We tested this method on research participant families in the 23andMe, Inc. database and the San Antonio Mexican American Family Studies (SAMAFS). With 77 children, we reconstructed 73.7% (mean; 96.6% maximum) of the two parents’ genotypes with mean error rate of <10⁻⁴. Using 4 children and leveraging IBD sharing with close relatives, we reconstructed 62.7% (mean; 82.9% maximum), with mean error rate <10⁻⁴. As datasets grow in size, more families will be implicitly collected; our work holds promise to enable high quality reconstruction of parent genotypes.
Over the past decade, the ability to assemble high-quality genome assemblies and parse haplotype-specific differences has enabled close evaluation of genome architecture evolution between individuals, populations, and species. Advancements in long-read sequencing accuracy and genome assembly algorithms revolutionized the ability to parse parental haplotypes within trios to directly interrogate haplotype-specific evolutionary questions. These questions often target key aspects of adaptation, selection, and speciation alike. Here, we report how using trio-binned F1-interspecific hybrids for single-haplotype genome assembly enabled near Telomere-to-Telomere genomes that provide novel insights into the evolution of genomic architecture spanning the family Felidae. We find an enrichment of inversions as well as large structural differences between species on the X chromosome, within a region previously associated with hybrid dysfunction that acts as a barrier to gene flow. For the first time, we evaluate the size and architecture of multi-megabase transcriptionally active FA-SAT macrosatellites in domestic cats, which are implicated in the formation of cancer and structure of centromeres. Finally, we provide further insight into the evolution of the large, copy-number variable macrosatellite DXZ4 which has been shown to have a role in speciation and hybrid sterility. Together, we present an intriguing view into the structurally dynamic evolution of felid genomes across roughly 15 million years. With continued additions of high-quality single haplotype genome assemblies, we aim to fully characterize the repertoire of structural differences between and within all cat species to facilitate genotype to phenotype discovery.
Assessing the impact of post-mortem damage and contamination on imputation performance in ancient DNA
Antonio Garrido Marques

Low-coverage imputation is becoming ever more present in ancient DNA (aDNA) studies. Imputation pipelines commonly used for present-day genomes have been shown to yield accurate results when applied to ancient genomes. However, post-mortem damage (PMD), in the form of C-to-T substitutions at the reads termini, and contamination with DNA from closely related species can potentially affect imputation performance in aDNA. In this study, we evaluated imputation performance i) when using a genotype caller designed for aDNA, ATLAS, compared to bcftools, and ii) when contamination is present. We evaluated imputation performance with principal component analyses (PCA) and by calculating imputation error rates. With a particular focus on differently imputed sites, we found that using ATLAS prior to imputation substantially improved imputed genotypes for a very damaged ancient genome (42% PMD). Trimming the ends of the sequencing reads led to similar improvements in imputation accuracy. For the remaining genomes, ATLAS brought limited gains. Finally, to examine the effect of contamination on imputation, we added various amounts of reads from two present-day genomes to a previously downsampled high-coverage ancient genome. We observed that imputation accuracy drastically decreased for contamination rates above 5%. In conclusion, we recommend i) accounting for PMD by either trimming sequencing reads or using a genotype caller such as ATLAS before imputing highly damaged genomes and ii) only imputing genomes containing up to 5% of contamination.
Estimating relatedness in structured and admixed populations using the distributions of Identity By Descent tract lengths

Arun Sethuraman

Arun Sethuraman
San Diego State University (USA)

Genetic relatedness estimators often utilize Identity By State (IBS) as a proxy for Identity By Descent (IBD), thereby leveraging observed allele frequency spectra in their inference. However, in highly structured or admixed populations, conflating IBS and IBD results in erroneous inference with large variance in relatedness estimates. Here I derive and describe a new computational method that utilizes the distribution of IBD tract lengths under the admixture model to estimate relatedness in structured or admixed populations. I describe a maximum likelihood-based estimator of relatedness, solved using an Expectation Maximization (EM) algorithm. I utilize extensive simulations of genome-scale haplotypic data under a variety of demographic models using stdpopsim, to test the accuracy and precision of our estimator. I also compare its performance against other estimators, including those implemented in InRelate, KING and IBSrelate. Finally, I test my method to infer close first and second degree relatives in the 1000 Genomes Project data, while accounting for the presence of population structure and admixture.
Detecting introgressed fragments from Neandertals and Denisovans in modern human genomes is an important step in the study of human evolution. A common strategy, implemented e.g. in Admixfrog and HMMix, is to use a Hidden Markov model that can detect archaic ancestry in small windows along the genome (bins). These methods estimate a posterior probability of putative archaic ancestry for small bins along a genome. Nearby bins with high probability can then be merged together into introgressed segments, the specifics and parameters of these procedures can introduce biases that may influence further downstream analyses such as admixture dating. Here, we develop a new method to call archaic ancestry fragments from posterior probabilities using a wavelet decomposition. This method allows us to perform a multiresolution fragment calling using the posterior probabilities of each bin, meaning that we can identify different fragment sizes across the genome by merging bins. We tested this method on simulation of Neandertals introgression into Denisovans. We then compared the results to the penalty score method performed by Admixfrog for fragment calling. We were able to identify segments with higher sensitivity and specificity without a fragment size cutoff in comparison to the Admixfrog penalty score. This method presents a promising alternative for accurately identifying the size of fragments of archaic ancestry in modern humans.
The effect of divergent and parallel selection on the genomic landscape of divergence
Hisham Ali

Hisham AA Ali, Tim Coulson, Sonya M Clegg, Claudio S Quilodran
- (United Kingdom), Arizona State University (USA), Geneva, Switzerland. (Switzerland), University of Geneva, University of Oxford (United Kingdom)

While the role of selection in divergence along the speciation continuum is theoretically well understood, defining specific signatures of selection in the genomic landscape of divergence is empirically challenging. Modelling approaches can provide insight into the potential role of selection on the emergence of a heterogenous genomic landscape of divergence. Here, we extend and apply an individual-based approach that simulates the phenotypic and genotypic distributions of two populations under a variety of selection regimes, genotype-phenotype maps, modes of migration, and genotype-environment interactions. We show that genomic islands of high differentiation and genomic valleys of similarity may respectively form under divergent and parallel selection between populations. For both types of between-population selection, negative and positive frequency-dependent selection within populations generated genomic islands of higher magnitude and genomic valleys of similarity, respectively. Divergence rates decreased under strong dominance with divergent selection, as well as in models including genotype-environment interactions under parallel selection. For both divergent and parallel selection models, divergence rate was higher under an intermittent migration regime between populations, in contrast to a constant level of migration across generations, despite an equal number of total migrants. We highlight that interpreting a particular evolutionary history from an observed genomic pattern must be done cautiously, as similar patterns may be obtained from different combinations of evolutionary processes. Modelling approaches such as ours provide an opportunity to narrow the potential routes that generate the genomic patterns of specific evolutionary histories.
Improving Local Ancestry Inference through Neural Networks

Jazeps Medina Tretmanis

Jazeps Medina Tretmanis, Maria C Avila-Arcos, Flora Jay, Emilia Huerta-Sanchez

Brown University (USA), CNRS, Center for Computational Molecular Biology, Department of Ecology, Evolution and Organismal Biology, INRIA, International Laboratory for Human Genome Research, Laboratoire Interdisciplinaire des Sciences du Numerique, UNAM (Mexico), Universite Paris-Saclay (France)

Local Ancestry Inference (LAI) is the process of classifying the genetic ancestry of chromosomal segments within a genome. LAI enables the study of ancestry-specific disease risk, as well as the evolution of admixed populations. Historically, LAI methods have been implemented through Hidden Markov Models (HMMs) underpinned by Li & Stephen’s model. However, recent methods have started to leverage Neural Networks (NNs) for the task of LAI. We present the first comprehensive benchmarking of both traditional and NN models for LAI using a combination of real and simulated data. We also benchmark the performance of NN architectures that have not previously been used for LAI. We include model performance for scenarios that are not usually benchmarked: short ancestry tracts (<1000 SNPs), and admixture between closely related populations. We find that Convolutional NNs achieve the highest performance out of all methods for traditional LAI. While we can reliably infer inter-continental genetic ancestry within admixed individuals, current methods cannot reliably infer intra-continental ancestries. However, we find that this can be achieved by supplementing SNP data with other statistics such as minor allele frequencies. Finally, we find that short ancestry tracts still pose a challenge for LAI, but that this can be ameliorated by incorporating knowledge of recombination rates into model architectures. The work presented here is useful both for its value in helping researchers make informed decisions regarding which methods fit their needs, as well as being an important initial data point for how NNs can improve the LAI method landscape.
How to paint a dog: limits of local ancestry inference in ancient genomes
Katia Bougiouri

Recent advances in imputation of ancient genomes have created the opportunity to apply state-of-the-art statistical genetics methods to ancient genomes of different species to study various aspects of evolutionary history across large timescales. One popular class of such methods performs what is known as local ancestry inference (LAI), which allows to “paint” an individual’s genome as a mosaic of segments inherited from different ancestral source populations. Despite their power to reveal patterns of genetic ancestry across the genome, most popular LAI methods have been designed to study relatively recent admixture events in human history using large panels of present-day genomes. It is currently unknown how reliable is LAI in ancient genomes, which are characterized by scarce geographic and temporal sampling, potentially large divergence between ancient source proxies and admixing populations, and complex ancient demographic histories. In this study, we evaluated the performance of two popular LAI methods, FLARE and MOSAIC, under different demographic scenarios, various sampling schemes and sample sizes, as well as the timing of admixture events, and assess their accuracy as a function of these parameters. Furthermore, we tested to what extent the inclusion of source proxies pre- or postdating the admixture event or grouping putative sources from different time points, affect the inferences. Our study establishes the limits of LAI in ancient genomes and provides general guidelines for LAI in future studies, which we demonstrate by a proof-of-concept LAI in a set of imputed Mesolithic dog genomes from northwestern Siberia.
Inference of recent effective population size from high and low coverage DNA data

Pier Francesco Palamara

Presented by self
Oxford University (UK)

Individuals sharing recent ancestors are likely to co-inherit large identical-by-descent (IBD) genomic regions. The distribution of these IBD segments in a population may be used to reconstruct past demographic events, such as effective population size variation, but accurate IBD detection is difficult in ancient DNA data and in underrepresented populations with limited reference data. We developed an accurate method for inferring effective population size variation during the past ~2000 years in both high and low coverage DNA data, called HapNe. HapNe infers recent population size fluctuations using either IBD sharing (HapNe-IBD) or linkage disequilibrium (HapNe-LD), which does not require phasing and can be computed in low coverage data, including datasets with heterogeneous sampling times. HapNe yields higher accuracy in a range of simulated demographic scenarios compared to currently available methods for IBD-based and LD-based inference of recent effective population size, while requiring fewer computational resources. We apply HapNe to several modern populations from the 1,000 Genomes Project, the UK Biobank, the Allen Ancient DNA Resource, and recently published samples from Iron Age Britain, detecting multiple instances of recent effective population size variation across these groups.
Leveraging the Ancestral Recombination Graph to estimate the STR mutation model
Sebastián Iturbe

Sebastián Iturbe, Alan Izarraras-Gómez, Rori Rohlfís, Carolina De Lima-Adam, Luis Fernández-Luna, Diego Ortega-Del Vecchyo
Department of Data Science and Institute of Ecology and Evolution, Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México (Mexico), University of Oregon (USA)

Short tandem repeats (STRs), also known as microsatellites, are repetitive base pair sequences that constitute approximately 3% of the human genome. Over 10,000 STR variants contribute to gene expression, accounting for 10-15% of the heritability of gene expression highlighting their significance in various clinical conditions. The evolution and mutation dynamics of STRs are a crucial factor we need to understand in order to perform more comprehensive analysis of the phenotypic impact of STRs. To achieve this understanding, it is imperative to infer the mutation model determining patterns of genetic variation of STRs. The evolution of STRs relies on the overarching evolutionary history across the genome. This history is manifested in the genomic relationships among all samples represented in the Ancestral Recombination Graph (ARG). Recent advancements in methods for inferring the ARG, such as Relate and tsinfer, present opportunities to devise strategies for inferring the most plausible STR mutation model integrating the evolutionary history encoded in the ARG. Our proposal involves the development of a novel maximum-likelihood method for inferring the most probable mutation model by leveraging information encoded in the ARG. Through exhaustive simulations, we demonstrate the accuracy of our method in inferring the STR mutation model. Additionally, we show an application of our method using data from the 1000 Genomes Project.
Kinship inference between individuals from an archeological site allows pedigree reconstruction and improves our knowledge on past social structures. However, reconstructing a pedigree can be challenging: kinship inference methods often only return the degree of relatedness between two individuals, which in turn can be explained by multiple underlying genealogical structures. For instance, an uncle-niece pair or a half-sibling pair would have the same degree of relatedness, but distinct pedigrees. Recent methodological improvements in ancient DNA now allow the detection of Identical-By-Descent (IBD) segments between two ancient relatives. In this work, we developed a method that leverages the analysis of IBD segments to solve complex genealogical structures. Briefly, in cases where siblings show a 2nd-degree relatedness with a third individual, our method first infer shared IBD between the siblings and that individual and then determine whether this person is their uncle/aunt or not based on Mendelian principle. The method also takes into account the rest of the pedigree to distinguish between other types of 2nd-degree relatedness when possible. We benchmarked our approach on simulated genomes and our results show a strong precision even for extremely low coverages. Finally, we applied it to the case study of the Koszyce archeological site to correctly place a young deceased child in a reconstructed genealogy.
S32 - Not just Ne Ne-more: New applications for SMC from ecology to phylogenies.
Gaining insights into the historical demographic dynamics helps to identify potential threats to species survival and to implement targeted conservation measures. Typically, the main metrics to assess the conservation status from genetic data is the variation in effective population size through time. This is related to the inverse of the coalescence rate (CR) which ultimately depends on the true demographic history of a species. SMC-based algorithms predict the variation of CR through time employing population genetics summary statistics obtained from genomic data. However, their performance in structured meta-populations remains poorly investigated. Moreover, under panmixia, variations in CR correspond to changes in population size over time, which has a simple biological interpretation. Conversely, the interpretation of CR variation in structured population is challenging, because it depends on the joint effect of all evolutionary forces acting on the meta-population. Fully understanding of the CR variation is of paramount importance, claiming for novel approaches to interpret the CR and integrate it with other population genetics statistics. Here we evaluated the accuracy of SMC-based algorithms by testing their performance in complex structured demographic scenarios for which the true CR was generated by coalescent simulations. We observed that SMC-based algorithms performance varies depending on the demographic parameters. We then used a Deep Learning (DL) model to correct SMC-predictions to be used in an AI model integrating information with other summary statistics, such as site frequency spectrum (SFS) to infer the demographic history of a population from its observed genetic variation.
Detecting structural variation in reconstructed genealogies

Anastasia Ignatieva
Anastasia Ignatieva, Martina Favero, Jere Koskela, Jaromir Sant, Simon R Myers
University of Glasgow (United Kingdom), University of Oxford (United Kingdom), University of Warwick (United Kingdom)

In the presence of recombination, the genealogy of a sample of sequences can be fully captured in the form of an ancestral recombination graph (ARG). Following recent breakthroughs, it is now possible to reconstruct genome-wide ARGs from large-scale sequencing datasets, and new ARG-based statistical inference methods promise powerful insights into evolutionary events and parameters. Genomic structural variants (SVs) are a ubiquitous and evolutionarily important type of mutation, however their presence is ignored by all currently available ARG reconstruction tools. Through developing new theoretical results on the distribution of the genomic span of a haplotype block, using the powerful framework of the SMC’ model, we present an accurate ARG-based tool for detecting SVs which result in the suppression of recombination in heterozygotes (such as chromosomal inversions). By applying this tool to ARGs reconstructed using Relate for the 1000 Genomes Project data, we detect several known and putative new inversions with high confidence, allowing for the genealogy-based analysis of their age and evolutionary history. We also detect other SVs, in particular copy number variants, by identifying and scanning for the characteristic artefacts they induce in reconstructed ARGs. Our results definitively demonstrate that reconstructed ARGs capture the signal of SVs, even though their presence is not explicitly modelled by the ARG reconstruction method. Identifying and analysing the evolution of SVs at all scales is an incredibly important goal, and our results provide the first evidence that reconstructed ARGs can be leveraged for this purpose.
Accurate inference of population history in the presence of background selection
Arun Durvasula

Trevor Cousins, Daniel Tabin, Nick Patterson, David Reich, Arun Durvasula
Cambridge University (United Kingdom), Harvard University (USA), USC (USA)

Most published methods for learning about demographic history make the simplifying assumption that the genome evolves neutrally, and do not seek to account for the effects of natural selection on patterns of variation. This is a major concern, as ample work has demonstrated the pervasive effects of natural selection and in particular background selection (BGS) on patterns of genetic variation in diverse species. Simulations and theoretical work have shown that methods to infer changes in effective population size over time (Ne(t)) become increasingly inaccurate as the strength of linked selection increases. Here, we introduce an extension to the Pairwise Sequentially Markovian Coalescent (PSMC) algorithm, PSMC+, which explicitly co-models demographic history and natural selection. We benchmark our method using forward-in-time simulations with BGS and find that our approach improves the accuracy of effective population size inference. Leveraging a high resolution map of BGS in humans, we infer considerable changes in the magnitude of inferred effective population size relative to previous reports. Finally, we separately infer Ne(t) on the X chromosome and on the autosomes in diverse great apes without making a correction for selection, and find that the inferred ratio fluctuates substantially through time in a way that differs across species, showing that uncorrected selection may be an important driver of signals of genetic difference on the X chromosome and autosomes.
Brown bears (Ursus arctos) colonized North America from Eurasia in two distinct and temporally separated waves. Once in North America they encountered endemic American black bears (U. americanus) during range expansions from eastern Beringia southwards into the interior of the continent. The establishment of sympatry between these species provided the opportunity for hybridization and introgression, which was previously identified using D-statistics. Both species have broad spatial ranges that should limit the extent of introgression, such that it is found primarily between sympatric populations. I used whole genome sequencing from samples collected across the ranges of both bear species to test for spatial variability in introgression. I identified two pulses of introgression between brown and American black bears, and demonstrate the introgressed segments occur across spatially structured lineages in both species. The first pulse occurred 240 – 130kya, near the estimated intraspecific divergence times of both species. This pulse was also estimated to occur prior to the establishment of sympatry in western North America, and thus implicates another lineage mediating gene flow between Eurasian brown bears and North American black bears. The second pulse occurred only between western black bears and North American brown bears. Introgressed genomic segments occur in both species, yet North American brown bears contain 2-4% of black bear segments, compared to 0.2-0.8% of brown bear segments in American black bears. This asymmetry may be linked to range expansion of brown bears across North America, but may also implicate adaptive introgression.
Mosquito evolution and the emergence of mosquito-borne disease in the human era
Noah H. Rose

Presented by self
San Diego (USA), University of California

Of thousands of species of mosquitoes, just a handful present the greatest threats to public health: those species that have recently evolved a strong preference for human hosts and habitats. I will discuss the results of a set of continent-scale studies documenting massive diversity and rapid evolution in Aedes aegypti, the globally invasive primary vector of dengue, Zika, chikungunya, and yellow fever, across its native range in sub-Saharan Africa. Cross-coalescent analyses suggest that these mosquitoes likely first became specialized on humans in response to a climatic shift towards long, hot dry seasons in the Sahel region of West Africa, where they came to depend on human water storage as habitat for their aquatic larvae. More recently, mosquitoes in dense urban habitats appear to be undergoing a rapid shift towards greater human preference. A large number of genes are involved in specialization, but they are concentrated in just a few key genomic loci, some of which are also associated with increased Zika competence. The expansion of cities into nearby forest areas, combined with changes in underlying vector genetics, present a major and growing risk for the zoonotic emergence of arboviruses.
Joint inference of evolutionary transitions to self-fertilization and demographic history using whole-genome sequences
Stefan Strütt

Presented by self
Max Planck Institute for Plant Breeding Research (Germany)

The evolution from outcrossing to selfing occurred recently across the eukaryote tree of life in plants, animals, fungi, and algae. Despite short-term advantages, selfing is hypothetically an evolutionary dead-end reproductive strategy. The tippy distribution on phylogenies suggests that most selfing species are of recent origin. However, dating such transitions is challenging yet central for testing this hypothesis. We build on previous theories to disentangle the differential effect of past changes in selfing rate or from that of population size on recombination probability along the genome. This allowed us to develop two methods using full-genome polymorphisms to (1) test if a transition from outcrossing to selfing occurred and (2) infer its age. The teSMC and tsABC methods use a transition matrix summarizing the distribution of times to the most recent common ancestor along the genome to estimate changes in the ratio of population recombination and mutation rates overtime. First, we demonstrate that our methods distinguish between past changes in selfing rate and demographic history. Second, we assess the accuracy of our methods to infer transitions to selfing approximately up to 2.5Ne generations ago. Third, we demonstrate that our estimates are robust to the presence of purifying selection. Finally, as a proof of principle, we apply both methods to three Arabidopsis thaliana populations, revealing a transition to selfing approximately 600,000 years ago. Our methods pave the way for studying recent transitions to self-fertilization and better accounting for variation in mating systems in demographic inferences.
Drivers of historical population declines in Australian Marsupials: are humans behind the wheel?

Toby G. L. Kovacs

Toby G. L. Kovacs, Simon Y. W. Ho, Carolyn J. Hogg

The University of Sydney (Australia)

Characterising the historical context of extant biodiversity can help to explain the vulnerability of certain groups to extinction and provide insight into how populations will respond to future environmental change. Owing to the historic geographical isolation of Australia, the island harbors a unique and outstanding mammal biota. Nevertheless, Australia has the worst mammal extinction rate in the world, with many remaining species now restricted to small, isolated populations. Previous demographic estimates have suggested that populations of Australian marsupials declined in the late Pleistocene, coinciding with the arrival of humans on the continent. However, the reliability of these demographic inferences is uncertain, given that they need to be scaled using estimates of species-specific mutation rates. To address this, we sequenced parent-offspring trios for two marsupial species of high conservation concern, and modeled mutation rates across marsupials. Using these new mutation rate estimates, we inferred population sizes across marsupials over the last million years. Our results suggest that many marsupial species experienced large population declines before the arrival of humans, while highlighting the impact of historical climate change on past population health. These findings could explain the vulnerability of Australian marsupials to extinction and raises concern for already struggling marsupial populations under a currently changing climate.
S33 - Science in the Spotlight: Empowering Education and Public Engagement with Cutting-Edge Science in Molecular Evolution.
Disease Detectives: Using Minecraft to explore virus evolution and epidemiology
Kate Duggan

Kate Duggan, Maria Fanourgiaki, Richard Fitzpatrick, Perrin Grant
University of Edinburgh (United Kingdom)

The COVID-19 pandemic illustrated the importance of clear, accessible public engagement, amidst mistrust and conspiracy theories surrounding SARS-CoV-2. However, the pandemic also served as a high-profile example of evolution in action. Our public engagement project uses now-familiar concepts such as variants of concern and infection waves to engage audiences with evolution and its applications to public health. Based around the popular game Minecraft, we created a workshop that introduces young people to data collection, sequence alignment and public health decisions. Our main aim was to increase familiarity with the use of genomics to respond to viral outbreaks and monitor pathogen evolution. Participants investigate a series of simulated outbreaks, which spread in Minecraft recreations of Scottish cities. The workshop is highly accessible, requiring no coding experience. The popularity of Minecraft and the exploration-based nature of the workshop ensured that even students with minimal interest or experience in evolutionary biology were engaged in gathering and analysing data. Our pilot phase saw the workshop reach over 1000 people across various settings, from drop-in events at music festivals to workshops with schools and community groups. In addition to helping us develop our outreach materials, the pilot allowed us to build new connections with young people and partner organisations across the Scottish Highlands. By using participants’ lived experience of COVID-19 and familiarity with Minecraft, our workshop serves as an accessible, adaptable method of communicating key concepts in evolution and public health.
Active Learning Experience using an analogy with the lac Operon model (PAPIME PE216224)

Marco Antonio Carballo-Ontiveros

Marco Antonio Carballo-Ontiveros, América Nixtin Castañeda-Sortibrán, Marcos Nahmad-Bensusan
Biofísica y Neurociencias, CINVESTAV (Mexico), Departamento de Fisiología, Facultad de Ciencias, UNAM (Mexico)

The analogy of the lac operon with a vending machine is intriguing. Analogies can be powerful tools for teaching complex concepts by relating them to something more familiar. The analogy to a vending machine might help students understand the concept of gene regulation and expression more concretely. To test this analogy, we designed an active learning class, which included a Learn-Before-Lecture (LBL) activity, an in-depth study of some aspects taught by the subject teacher, and, as a central strategy, the analogy of the lac operon with a coffee vending machine, which was accompanied by activities in which students submitted their answers digitally. In this regard, our students showed astonishment at the change in class dynamics, presenting resistance to answering a questionnaire associated with the LBL activity. On the other hand, a favorable response was observed to the analogy of the lac operon and a coffee vending machine. The activities of this analogy were solved quickly and favorably by the students. In our experience, implementing an analogy to an abstract topic, such as the lac operon, within an active learning class allowed students to assimilate knowledge quickly and enthusiastically.
The Network of Researchers on the Chemical Emergence of Life (NoRCEL) is an interdisciplinary collaborative network focused on understanding the origins and evolution of life from a chemical perspective. Founded by Sohan Jheeta, NoRCEL aims to bring together scientists from various disciplines, including chemistry, biology, astronomy, and Earth sciences, to investigate one of the most fundamental questions in science: how life began. In recent times, NoRCEL has successfully organised annual debate forums on the other side of human evolution, its own extinction and that of other species due to overexploitation of natural resources. Finally, looking for preservation of human health, a few NoRCEL scientists have examined the microbiome and its evolution. In short, like the community’s motto, NoRCEL is about: Emergence, Evolution, Extinction. NoRCEL fosters collaboration irrespective of geographical boundaries. A cornerstone of NoRCEL’s mission is supporting early-career scientists and students, particularly from underrepresented regions, in astro-sciences and life emergence studies. By connecting emerging researchers with established scientists, the network aims to nurture the next generation of interdisciplinary scholars. Overall, NoRCEL endeavours to advance understanding of life’s origins while promoting collaboration, inclusivity, and public outreach in the scientific community. Through colloquiums, workshops, and conferences, it facilitates the exchange of ideas and findings, providing platforms for researchers and students to present work and forge new partnerships. Moreover, NoRCEL prioritizes public engagement, making the study of life’s chemical evolution accessible through lectures, publications, and online resources. Gradually, NoRCEL will be transformed from only an academic community into a scientific institute for education.
Engaging the Public on Biology’s Moonshot: The Earth BioGenome Project

Nicolette Caperello

Presented by self

Davis (USA), University of California

The Earth BioGenome Project (EBP) is a moonshot for biology that aims to sequence, catalog, and characterize the genomes of all of Earth’s eukaryotic biodiversity over a period of ten years. This grand vision will create a new foundation for biology to drive solutions for preserving biodiversity and sustaining human societies. Public engagement and education are critical to the success of this project. The EBP Secretariat, housed at Arizona State University in the US, along with the EBP Communication and Public Affairs Committee, uses a multimedia approach to spread science messaging around the globe to engage scientists and educate broad audiences to fuel scientific curiosity. Our approach includes 1) typical platforms like websites and social media, 2) innovative in person workshops that harness Justice, Equity, Diversity, and Inclusion outcomes, 3) an annual Biodiversity Genomics conference that is online, free, and open to all, and 4) in person and online trainings, publications, and presentations from our Affiliated Projects that share EBP stories and best practices throughout their networks and beyond. With 54 affiliated projects headquartered in 14 countries and scientists working in over 55 countries on 6 continents, the EBP has considerable global coverage and cooperation with dynamic science stories to tell. The key to a successful conclusion of this project is synergy; this session will weave the various methods of public engagement used by the EBP, the largest project in the history of biology, to construct a comprehensive and impactful strategy for the communication of cutting-edge research.
Evaluating community perceptions and ethical considerations in genetics research in small scale northern Kenyan populations
Rebecca L Siford

Rebecca L Siford, Carla Handley, Sarah Mathew, Melissa Wilson, Jason Robert, Mercy Y Akinyi, Joseph Kamau, Anne C Stone
Arizona State University (USA), Institute of Human Origins, Kenya Institute of Primate Research (Kenya), One Health Centre, School of Human Evolution and Social Change, School of Life Sciences

As genomic research expands to be inclusive of global diversity, indigenous perspectives about genetic research and data management should be considered, and the success of dissemination efforts to participants should be evaluated. Kenya, where participation in genetic research is amongst the highest in Africa, is also one of the most understudied for participant perspectives from small scale societies. Here, we address how Western ethical practices used in genetic studies accurately align (or fail to align) with the understandings, attitudes, and perceptions of the Turkana, Borana, Rendille, and Samburu pastoral populations of northern Kenya. Genetic research by our group has been conducted in collaboration with these populations which facilitates the opportunity to explore ethical practices and concerns. We collected semi-structured interviews from previous genetic study participants and from those who had not previously participated, to explore the effectiveness of dissemination efforts and to characterize views on privacy, consent, data use, and governance of data. From over 550 interviews collected in 2023, we generally find that dissemination of genetics results is expected and should be presented with cultural and religious considerations. Also, broad consent is an inappropriate model and permission for new studies should be sought. Finally, communities favor a participatory model of research. We provide a general set of guidelines and population specific governance frameworks that may aid Kenyan research ethics committees and other researchers. Our study highlights that participant and community perspectives offer culturally relevant motivations and are necessary to ensure ethical governance of genetic data.
Melanogaster Catch The Fly! Rural high schools committed to frontier research on genomics and evolution through citizen science

Roberto Torres

Presented by self
Barcelona (Spain), La Ciència Al Teu Món

The citizen science network in adaptation genomics "Melanogaster Catch the Fly!", through its hands-on citizen science practices have already demonstrated its efficiency to establish a long-term relationship between evolutionary biologists and high school students, teachers, farmers and other interested parties from rural areas across Europe (and beyond), in order to expedite, and make more participative, cutting-edge scientific research in adaptation genomics. Citizen science practices, also contribute to generate the knowledge to build on citizens the capacities that are necessary to face societal challenges such as monitoring and analyzing the impact of climate change in global and local biodiversity. Melanogaster Catch the Fly! evidences the relevance and convenience of citizen science practices to involve society in all the different stages of the scientific process: from co-creation of research processes, development of fieldwork technological solutions, to data collection and bioinformatic analysis, validation of experimental data through molecular biology technics, policy-making, and the dissemination of results. As such, Melanogaster Catch the Fly! allows citizens to become the enablers of a global change in the perception of the implications of basic science, and in the correct understanding of evolutionary processes with implications for biodiversity conservation, personalized medicine, and food security, all relevant for citizens daily lives. Moreover, this presentation will share a compendium of citizen science recommendations on promoting scientific literacy in Evolution through citizen science.
The analysis of repetitive regions (RRs) throughout the genome is a formidable challenge in bioinformatics and computational biology. Not only do RRs (including transposable elements (TEs), endogenous viruses, and repeats of as-of-yet-unknown origin) constitute a significant portion or the majority of some genomes, they can also be the most dynamic components of genome architecture due to their ability to replicate, mobilize, horizontally transfer, or facilitate non-homologous recombination. Many biologists hope to analyze RRs of the genomes they study, either to understand their impact, history, or function. Because of their ubiquity and importance, a dizzying array of tools have been developed to find, characterize, and quantify repetitive DNA. Unfortunately, most programs depend on species-specific libraries or focus on a particular type of repeat. TE-Hub is a community of scientists working on repeats formed to address the ongoing challenges associated with analyzing and understanding genomes, and with the goal of sharing resources and knowledge with those new to this sector of DNA. Our aims are to provide connections among TE researchers, make new tools and resources, facilitate the development of community norms and benchmarks for software development, and offer online and in-person workshops to help introduce and train the next generation of biologists on extant tools. We hope these efforts will advance our understanding of DNA repeats, and further support the growing community of scientists dedicated to the challenge of understanding this dynamic part of the genome.

Sishuo Wang

Sishuo Wang, Jianhao Lv
Nanfang College of Sun Yat-Sen University (China), The Chinese University of Hong Kong (Hong Kong)

Molecular evolution and phylogenetics have rapidly become crucial domains in biology in the past decades, yet pose a high entry barrier due to their interdisciplinary nature, encompassing molecular biology, evolutionary biology, statistics, computer science, etc. Ziheng Yang's books, "Computational Molecular Evolution" (http://abacus.gene.ucl.ac.uk/CME/) and "Molecular Evolution: A Statistical Approach" (http://abacus.gene.ucl.ac.uk/MESA/), stand as cornerstone references in this interdisciplinary domain, offering not only comprehensive content but exceptional exercises covering proofs/computations, programming, and software operation. However, no solutions to the problems are available, posing significant challenges for biologists to further understand the problems and the books. We present an over 140-page comprehensive solutions manual for all the 75 exercises included in Ziheng Yang's books. This manual provides detailed, step-by-step derivations, along with extensive code examples and datasets for hands-on practice. Additionally, it provides alternative solutions/derivations where applicable, incorporating state-of-the-art methods to tackle some of the "classic" problems, enabling readers to experience the evolution of phylogenetic methods. To facilitate engagement and collaboration, we have established an online platform hosted on GitHub (https://github.com/sishuowang/Solutions_Manual_CME2006_MESA2014), inviting feedback, alternative solutions, and error reporting from the community. Embracing e-learning principles and a flipped classroom approach, this platform empowers learners to drive their learning journey actively. This work has never received any financial support, but we believe that our initiative will catalyze a community-driven effort to help researchers and students from various disciplines interested not only in Yang's seminal works but also in computational molecular evolution and phylogenetics.
PhyloGenome: a gamified and participatory approach to genomic education

Sònia Casillas, Noah Jornet-Calomarde
Department of Genetics and Microbiology, Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona (Spain)

PhyloGenome is an innovative educational platform that integrates genomic science with the interactive elements of a collectible card game, fostering active learning and collaboration among students. It allows learners to become protagonists in their educational journey, applying theoretical concepts to real-world genomic sequencing, assembly, and annotation. As part of the course, which complements traditional lectures, students choose a genome to investigate, culminating in the creation of an informative species card. This card is then added to the PhyloGenome deck, contributing to a growing educational resource. The rich game mechanics, inspired by the open-access Phylo game, combine scientific concepts with strategic gameplay, immersing players in their very own narrative of the genome sequencing era through a diverse array of species, event, and generation progress cards. The PhyloGenome project is supported by a dedicated website (phylogenome.omicsuab.org), which acts as a central hub for educators and learners. It provides free access to the complete card set, game rules, glossaries, and teaching materials. The website also encourages a collaborative community by inviting educators to involve their students in contributing new cards to the project. By distributing all materials under a Creative Commons license, PhyloGenome stands at the forefront of inclusive, participatory science education. It innovates genomic education, merging academic learning with gameplay to spark scientific curiosity and collaboration. Beyond educational outcomes, PhyloGenome aims to impact broader societal issues, such as environmental awareness, open education access, and community development, paving the way for a more sustainable future.
A vast amount of population genomic data in a growing number of species is currently being generated by the newest sequencing technologies. This data contains precious information about the evolutionary history of life on our planet. To help in the evolutionary interpretation of this information, we have launched the PopLife project, a free online resource for population genomics to describe and explain genetic variation within and between populations across any species in the tree of life. PopLife provides a highly performing pipeline to compute population genetics statistics for large-scale genomic datasets, such as variation and divergence metrics, linkage disequilibrium parameters, and neutrality tests. Its novel user-friendly genome browser interface allows interactive visualization and data retrieval. The available information applies not only to evolutionary genomics, but also to fields such as functional genomics, conservation genetics, genetic breeding, pandemic risk, and countless applications in other research fields. PopLife aims to address broader impacts on Environmental, Social, and Governance (ESG) issues. We plan to address environmental concerns by utilizing artificial intelligence algorithms to detect early warning signs of genetic bottlenecks or inbreeding in vulnerable populations, providing proactive strategies for conservation efforts. Our social initiatives include educational programs, outreach workshops, and accessible online resources for undergraduates, high school students, and the general public to promote understanding of population genomics and its importance. Finally, governance initiatives focus on applying open data standards and establishing a collaborative network of researchers with complementary knowledge and skills to establish PopLife as an international reference resource in population genomics.
S34 - Going local: Using engaged research practices to understand regional-scale interactions.
Conversations in community-based genomic research: Showcasing a platform from Chile to integrate Indigenous engagement and representation

Constanza de la Fuente Castro

Constanza de la Fuente Castro, Tomás González Zarzar, Felipe Martínez, Nicolás Montalva, Maanasa Raghavan, Constanza Silva-Gallardo, Ayelén Tonko-Huenucoy

Department of Human Genetics, Escuela de Antropología, Facultad de Ciencias Sociales. Pontificia Universidad Católica de Chile (Chile), Facultad de Medicina. Universidad de Chile (Chile), ICBM, Independent researcher (Chile), Millennium Nucleus on Sociomedicine (Chile), Museo Nacional de Historia Natural (Chile), Programa de Genética Humana, Society and Health Research Center, Universidad Mayor (Chile), University of Chicago (USA)

Human genomic research has greatly increased our understanding of biological diversity. Yet, the underrepresentation of certain regions limits its global impact. In Latin America, initiatives to diversify sampling, particularly focused on minoritized and marginalized groups aim to democratize genetic research. However, sampling strategies do not always consider a key problem in research: the level of engagement by communities and their involvement in research design, implementation, and result interpretation. Overcoming this issue requires reconsidering how research is formulated and performed within a region-specific context as well as boosting strategies that level the field and establish long-term relationships between academia and Indigenous researchers, representatives, and communities. With this goal in mind, we designed and implemented two 4-day workshops in 2022 and 2023 in Chile titled ‘Genomics and Identities in Chile’ focusing primarily on the interplay between genetic histories and self-identities. Both workshops were attended by 30 participants from various Indigenous communities and 12 genetic and anthropological researchers from Chile and abroad. These workshops, first of their kind in Chile, were not designed around an ongoing or new research project, but as a way to facilitate and promote spaces for dialogue on genetic research in the country, especially when involving Indigenous participation. In this presentation, we will discuss these workshops’ rationale, content, and format, our experience implementing them, and the limitations and difficulties we faced as organizers. We also discuss participants’ feedback to motivate similar initiatives for working towards more equitable and collaborative relationships between academia and Indigenous peoples.
Paleogenomic analyses of archaeological remains reveal sex roles and mobility of ancient families in central Mexico
Daniela Orozco-Perez

Daniela Orozco-Perez, Sofía Ivonne Vieyra Sanchez, Viridiana Villas-Islas, Maria A. Nieves-Colón, Maria C. Ávila-Arcos, Rafael Montiel Duarte, Gabriela Zepeda-García Moreno, Alberto Aveleyra Talamantes, Karla Sandoval-Mendoza, Andres Moreno-Estrada
Amigos del Museo de San Miguel A.C./Proyecto cultural Artesanos del Tiempo (Mexico), CINVESTAV (Mexico), Department of Anthropology, Equality and Gender Office of the Centre for Research and Advanced Studies (CODIGO-C), International Laboratory for Human Genome Research (LIIGH), National Institute of Anthropology and History (INAH) (Mexico), National Laboratory of Genomics for Biodiversity (UGA-LANGEBIO), UNAM (Mexico), UNAM City: Querétaro (Mexico), University of Minnesota Twin Cities (USA)

Ancient DNA recovery remains a challenge in temperate environments across Latin America. A recently excavated archaeological site in Central Mexico, Cañada de la Virgen, has been hypothesized to be a multicultural complex inhabited from 540 to 1050 CE., but its origins remain elusive. Using aDNA we inferred the genomic origins, kinship dynamics and biological sex of the individuals buried at the site. We sequenced 19 human skeletal remains, with six samples having >1% endogenous DNA and undergoing enrichment and deep sequencing. Chromosomal sex showed that the most notable burial -“The Hierarch”-, previously assigned as male by osteology, was genetically inferred as female and, according to C14, is surprisingly ~1000 years older than the complex. We also uncovered a mother-son relationship involving two skeletons buried next to each other and assessed their genetic origins by combining our data with reference panels of modern Indigenous Mexican populations. Using PCA and F-statistics we found that the mother exhibits genetic proximity to southern Oaxaca and Guerrero Indigenous groups, while her son showed genetic contribution from northern groups traced back to Sonora (near the US border), which is not observed in his mother. Overall, our results suggest that this complex had multiple genetic contributions from places as distant as Mexico’s northern and southern tips, converging around 800 CE and bringing burials from before the settlement. Our work illustrates a unique perspective on sex roles in prehispanic societies, challenging the gender narrative of leading characters and mobility dynamics in ancient Mexico.
An ancient genetic insight into pre-colonial Trinidad.

Eleni Seferidou, Zara Ali, Ashleigh Morris, Kathrin Naegele
Leiden University (Netherlands), Max Planck Institute for Evolutionary Anthropology (Germany), The National Trust of Trinidad and Tobago (Trinidad and Tobago)

Recent ancient genomic research into the Caribbean region has shed light on the complexity of past Caribbean populations, possible settlement routes, and human movement into and between the islands. Successive dispersals from the mainland to the islands have been recognised, with one of these dispersals originating from northern South America, following a northward route into the islands, possibly via Trinidad, due to its geographic location. Genetic research thus far has taken a much-needed pan-Caribbean approach to get a grasp of population structure and admixture in the region. Given the fact that diversity in pre-colonial Caribbean populations has become apparent from multiple lines of evidence, this study is aimed at a localised approach, by focusing on a specific island, with the aim of reconstructing individual genomic histories and local interactions. This more localized research framework has led to a collaborative project, together with local archaeologists, researchers, and stakeholders to provide a valuable contribution to the local history through genetic data. The two sites under investigation - Manzanilla and Red House - are found on opposite sides of the island, with the former located on the eastern shore and the latter near the western shore in the capital, Port of Spain. Thus, they provide a valuable opportunity to study genetic connections between their inhabitants, and among them and neighboring, contemporaneous populations.
Distinct positions of genetic and oral histories: Perspectives from India

Esha Bandyopadhyay

Esha Bandyopadhyay, Arjun Biddanda, Constanza de la Fuente Castro, David Witonsky, José Antonio Urban Aragon, Nagarjuna Pasupuleti, Hannah M. Moots, Renée Fonseca, Suzanne Freilich, Jovan Stanisavic, Tabitha Willis, Anoushka Menon, Mohammed S. Mustak, Chinnappa Dilip Kodira, Anjaparavanda P. Naren, Mithun Sikdar, Niraj Rai, Maanasa Raghavan

Anthropological Survey of India (India), Birbal Sahni Institute of Palaeosciences, Uttar Pradesh (India), Cincinnati Children’s Hospital Medical Center (USA), Cystic Fibrosis Research Center, Department of Anthropology, Department of Applied Zoology, Department of Archaeology, Department of Human Genetics, Department of Pediatrics, Division of Pulmonary Medicine, Facultad de Medicina, Instituto de Ciencias Biomédicas, Mangalore University (India), Programa de Genética Humana, PureTech Health (USA), Universidad de Chile (Chile), University of Cambridge (United Kingdom), University of Chicago (USA), University of Vienna (Austria)

Several studies over the past decade have demonstrated the power of human genomic datasets to provide insights into population histories. Concurrently, there is potential for tension arising from genetic histories being prioritized over oral histories or used to confirm community-based knowledge and ethnography, especially if they differ. To investigate the interplay between oral and genetic histories, we focused on the southwestern region of India. We generated whole-genome sequence data from 158 individuals identifying as Bunt, Kodava, Nair, and Kapla. We additionally considered historical records and oral histories, focusing on references to non-local origins such as ancient Scythians for Kodava, Nair, and Bunt, members of Alexander the Great’s army for the Kodava, and an African-related source of ancestry for Kapla. Our results showed that the Bunt, Kodava, and Nair exhibit strong genetic similarity to other Indian populations, with Kapla being more similar to tribal groups from South India that harbor a greater proportion of genetic ancestry related to Ancient Ancestral South Indians. We did not find evidence of additional genetic sources in the study populations other than those identified in other present-day South Asians. Our results demonstrate that oral and genetic histories may not always provide a consistent account of population origins. Moreover, our study highlights the importance of community-engaged, multi-disciplinary investigations to motivate avenues for ethnographic and anthropological follow-up of the origin stories in these communities’ oral histories, while also cautioning against the dangers of conflating self-identities and genetic histories.
Uncovering the global origins of an Industrialized microbiome using regionally focused studies
Laura Weyrich

Presented by self
Pennsylvania State University (USA)

Industrialization is considered as the single largest global factor globally driving healthy human oral and gut microbiome variation. This was identified using comparative population studies with people living in Industrialized countries contrasted to rural areas or Indigenous communities. However, these populations have distinct evolutionary histories, genetics, behaviors, diets, and cultural practices, which can misrepresent the factors and practices that shaped Industrialized human microbiomes. Local, historical approaches may help identify specific processes that shape Industrialized microbiomes. Here, we reconstruct ancient oral microbiomes within calcified dental plaque to trace the impacts of Industrialization in Great Britain, Germany, Netherlands, and Australia in a localized context. Dietary shifts (i.e., sugar, milk, and meat consumption) and historical events (i.e., Second Plague Pandemic) directly shaped pre-Industrial populations in Great Britain in unique ways, resulting in an unexpected loss of oral microbial diversity prior to the Industrial Revolution. The onset of the Industrial Revolution (1770) further shifted oral microbiome composition in Great Britain, but not other locations in Europe until later. Post-industrial impacts on oral microbiomes were imported into British colonies, such as Australia, and maintained, even without active Industrial practices, likely due to an increase in Fusobacterium nucleatum – a keystone species in dental plaque microbiome. Lastly, a single population study over the past 150 years in Australia indicates that modern Industrial signatures occurred sometime after the 1930s and are tied with the Great Acceleration – and not Industrialization. Great Acceleration shifts are also linked to rises in the HACEK microbes today linked with cardiovascular disease, suggesting that environmental, dietary, and lifestyle shifts over the last 70 years may contribute to ‘Industrial’ health signatures today. This has important implications for tracing and identifying the specific factors and policies that led to ‘Industrial’ microbiomes today, especially in those experiencing differential rates of microbiome-associated diseases.
Community engagement experiences of the Afromexico Genomics Project
María Ávila-Arcos

Presented by self
International Laboratory for Human Genome Research-UNAM (México)

Approximately 250,000 enslaved Africans were forcibly brought to New Spain between the 16th and 17th centuries as part of the Trans-Atlantic Slave Trade. Today, more than 3 million Mexicans self-identify as “Afro-Mexicans”. They were only officially recognized in the Mexican constitution in 2019. The Afromexico Genomics Project collaborated with Afro-Mexican communities to characterize their genetic ancestry using dense genome-wide genotyping. Participants in the study were 380 people from three Mexican states: Guerrero, Oaxaca, and Veracruz, who self-identified as either Afro-descendants, Indigenous or as having mixed heritage. To complement the genome-wide genotype data, we collected genealogical data and self-identification information. We engaged with the communities through informative talks and brochures and returned the genetic ancestry results to participants. We carried out ethnographic interviews to learn about their expectations from the project. Also, we interviewed participants after returning the genetic ancestry results to investigate the impact of these on their self-identification. The study was carried out with utmost consideration of the vulnerable status of these populations, which include marginalization, discrimination, and limited access to health services. Therefore, an additional goal of the project is to contribute to the appreciation of Afro-Mexicans as part of Mexico’s mosaic of diversity and help set the stage for health interventions.
Small-scale genetic histories from Roman Britain
Marina Silva

Marina Silva, Thomas Booth, Joanna Moore, Kyriaki Anastasiadou, Christopher Barrington, Sharon Clough, Alexandre Gilardet, Paolo Guarino, Michael Henderson, Sarah Johnston, Monica Kelly, Jesse McCabe, Alex Smith, Leo Speidel, Pooja Swali, Frankie Tait, Don Walker, Mia Williams, Alistair Barclay, David Bowsher, Janet Montgomery, Pontus Skoglund
Ancient Genomics Laboratory, Bioinformatics and Biostatistics, Cotswold Archaeology (United Kingdom), Department of Archaeology, Genetics Institute, Headland Archaeology (United Kingdom), Museum of London Archaeology (MOLA) (United Kingdom), The Francis Crick Institute (United Kingdom), University College London (United Kingdom), University of Durham (United Kingdom)

The Imperial Roman period witnessed significant political and societal changes, fueled by road and maritime network development associated with military campaigns, governance and trade, resulting in an intensification of human movement across Western Eurasia, including long-range mobility. We aim to study the Roman period in Britain, at the northwestern fringe of the Roman Empire, from a genomics perspective by building a bottom-up sampling approach, focusing on intra-site questions arising primarily from the archaeological context and stemming from close partnerships with collaborators from academic, commercial and cultural heritage sectors. Here we showcase two examples of how our multidisciplinary approach, combining multiple strands of evidence (e.g. aDNA, osteological assessment, radiocarbon dating, isotope analysis), has contributed to illuminate social practices within small and rural communities and uncover life histories of single individuals and family groups in Roman Britain: 1) the life history of long-distance mobility of a male individual in a isolated inhumation in Cambridgeshire, who did not carry traceable ancestry related to local populations in Britain and was instead genetically closer to Caucasus and Sarmatian groups; and 2) the reconstruction of small family trees spanning up to four generations in a small rural cemetery in Childrey Warren, Oxfordshire. These two contrasting archaeological contexts – isolated inhumation versus a formal cemetery – highlight the diversity of burial traditions and provide insights on different aspects of life and death during this period, ultimately contributing for a more nuanced view of Roman Britain.
S35 - Genetics, Molecular Biology, and the Future of Forensic Science.
The recovery of DNA from burned forensic contexts.

Anne C Stone

Cody Parker, Amber Coffman, Erin Rawls, Adam Ben Rohrlach, Katelyn Bolhofner, Giovanna Vidoli, Joanne Devlin, Anne Stone

Arizona State University (USA), Max Planck Institute for Evolutionary Anthropology (Germany), University of Adelaide (Australia), University of Tennessee (USA)

Innovation in molecular methods provides opportunities to increase the likelihood of successful downstream DNA identification in challenging forensic contexts such as those with thermal alteration. Since ancient DNA research is heavily invested in optimizing recovery of DNA from challenging samples under similar contexts (e.g., low yields of highly fragmented/degraded DNA), we used and evaluated aDNA as well as forensic extraction methods for 36 paired burned bone and overlying tissue samples that had been classified into different levels of burning based on bone characteristics (i.e. color, cracking). For 109 bone samples, we also constructed both double stranded and single stranded DNA libraries and then used these for targeted enrichment of the mitochondrial DNA genome (H. sapiens Representative Global Diversity Panel) and genome-wide single nucleotide polymorphisms (FORCE-v2, Arbor BioSciences). We found that charred tissue samples when available consistently returned higher concentrations of both total DNA (Qubit HS DNA assay and Agilent TapeStation D5000 HS) and endogenous DNA (Quantifiler Trio). In addition, these assays were not a reliable predictor of successful DNA recovery, likely due to high levels of co-extracted inhibitors. These inhibitors also affected NGS results, reducing recovery in burn category 2.5 but lessening in categories 3 and 3.5. DNA recovery and mapped reads dropped significantly at temperatures greater than 350°C (burn categories 4 and 5). Our results show that innovative molecular methods can improve DNA recovery from burned samples for human identification; however, we will also discuss the barriers that exist to implementing these in routine forensic analyses.
How should we report genetic matches following an Investigative Genetic Genealogy search?

Feriel Ouerghi, Dan E Krane, Nathan Adams, Michael D Edge
Forensic Bioinformatics (USA), University of Southern California (USA), Wright State University (USA)

Traditional forensic genetics practice in the United States has historically relied on short tandem repeat (STR) analysis 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. Over the past seven years, a new approach called investigative genetic genealogy (IGG) has been instrumental in solving hundreds of criminal cases. IGG 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 IGG are genotyped at CODIS (STR) markers, and a summary of the evidence from the match, such as an RMP or a likelihood ratio (LR) is subsequently reported in court. However, the calculated statistics do not account for the fact that the suspect was initially identified through an IGG match. We demonstrate that, since an IGG search potentially includes identical-by-descent (IBD) segments over the CODIS markers, the probability of a CODIS match increases with relatedness. We also study how population structure, isolation by distance, and assortative mating increase the probability of CODIS matches, even in the absence of overlap between the IBD segments used in IGG and the CODIS markers. Based on this, we argue for a conservative framework for evaluating RMPs and LRs following identification via IGG.
High-throughput sequencing (HTS) offers new opportunities in forensic genetics, providing several advantages over traditional methodologies. We have assessed the applicability and performance of HTS for the identification of victims from mass graves of the Spanish Civil War (1936–1939). Here, we present the genetic identification of victims buried in 2 mass graves located in Valencia which were analyzed previously by traditional technologies. A total of 102 post-mortem remains and 76 relatives (41 families) were sequenced using HTS with Verogen-Qiagen reagents and sequencer, which included 59 STRs (28 autosomal, 24 Y-chromosome, 7 X-chromosome) and 94 SNPs. We used the DVI module of Familias software for kinship analysis. Informative profiles, with >50% genotyped markers, were obtained in 70 remains. To the 30 identifications obtained with capillary electrophoresis sequencing of 32 STR markers, we have added the identification of 14 victims. These identifications were obtained after comparison with reference samples from first-, second- and third-degree relatives. As expected, parent-child relationships were more robustly identified than other relationships. The main limitations for the identification of victims from these mass graves were the low number of first-degree living relatives and the low DNA concentration and high levels of DNA degradation in the remains. Despite this, NGS-based genotyping provides numerous advantages that improve the interpretation of challenging or degraded DNA and have the potential to increase the power of discrimination for identification and kinship analysis, especially for low-quality samples and distant relatives.
Familial searching in forensic genetics involves testing a query DNA profile against a database to identify potential relatives of database entrants. Traditionally, this method requires both profiles to be typed at overlapping markers to determine compatibility with a biological relationship. We extend this approach to scenarios where query and target profiles are typed with nonoverlapping marker sets in linkage disequilibrium. Using a likelihood-based algorithm, we explore the feasibility of leveraging linkage disequilibrium to identify relatives when profiles rely on distinct genetic markers. Considering data on individuals genotyped with both CODIS microsatellites used in forensic applications and genome-wide SNPs, we demonstrate that familial relationships can be identified using the SNPs of one member of the pair and the microsatellites of the other. We further show that genetic record-matching using SNP data, even with low genomic coverage, achieves similar accuracy to whole-genome sequencing data. Additionally, we assess the impact of genetic ancestry on the accuracy of record matching. Our work introduces a novel approach for identifying degraded DNA samples in criminal justice, mass disasters, missing persons, and ancient DNA contexts, even when genotyping standard forensic STRs is not feasible. Furthermore, we highlight that privacy concerns arise from computations across multiple databases without shared genetic markers, posing risks to both database entrants and their close relatives, which are further influenced by genetic ancestry.
Comparing accuracy of forensic DNA mixture analysis across groups with varying genetic diversity.

Kamillah T Felix

Maria Flores, Cara Ly, Evan Ho, Niquo Ceberio, Kamillah Felix, Hannah Mariko, Miguel Guardado, Matt Paunovich, Chris Godek, Carina Kalaydjian, Rori Rohlfs

George Washington University (USA), Los Angeles (USA), San Francisco (USA), San Francisco State University (USA), Santa Cruz (USA), University of California, University of Oregon (USA)

Police are increasingly using trace amounts of DNA in investigations, often when samples contain DNA from multiple contributors. The reliability of interpreting these DNA mixtures is complex. A likelihood ratio (LR) is a value used to assess whether genetic evidence at the crime scene supports the defense hypothesis (the suspect did not contribute to the DNA mixture), or the prosecutor’s hypothesis (the suspect did contribute to the mixture). The LR is calculated based on the allele frequency distribution of the assumed group that the person of interest (POI) belongs to, which will be referred to as the reference group. We hypothesized higher error rates when the reference group is incorrectly assumed (different from the POI’s true genetic background), and even higher when the true group has low genetic diversity. We also predict that as the number of contributors in a mixture increases, the amount of false positives increases. To test our hypothesis, we used Forensim – a free open source R package – to simulate individual genotypes, forensic DNA mixtures, and calculate the LR. We observed that when there are more individuals in the DNA mixture, there is less reliability in assessing the evidence where the POI did not contribute. With the correct reference group, there was an increase in the false positive rate for groups with low genetic diversity compared to groups with high genetic diversity. Our results indicate that forensic DNA mixture analysis tools used today may falsely identify individuals with certain genetic backgrounds.
Construction of Epigenetic Clock using Cell-Free DNA

Maria Flores

Maria Flores, Matteo Pellegrini
University of California Los Angeles (USA)

In recent years, DNA methylation for the use of investigative leads in forensic cases, specifically age estimation, has become promising and is increasingly being studied. Several studies have shown that about one-third of methylated sites in the genome are affected by age, leading to the attempted usage of age-related differentially methylated regions for the prediction of chronological age or otherwise known as “epigenetic age clocks”. While many of these studies have developed their own mathematical models for epigenetic age prediction, most methods have only considered optimal coverage samples particularly in whole blood and saliva. Here, we present an epigenetic clock using a binomial estimator approach from whole genome bisulphite sequencing data of plasma (cell-free DNA) retrieved from pregnant women that will be used as a proxy for samples that can be found in forensic casework due to its varying degrees of coverage and sparseness. Preliminary results show that of the eight samples that were randomly selected, six of the samples were accurately assessed for a fair accuracy rate of 75%. The model presented here serves to provide a proof of principle so that a similar model may be considered for sparse forensic samples as degraded DNA is becoming more frequently analyzed among law enforcement.
Highly accurate predictions indicating sexual activity using microbiome-based analyses in forensic settings
Meghna Swayambhu

Meghna Swayambhu, Sarah Ahannach, Sarah Lebeer, Eirik Natås Hanssen, Lars Gustav Snipen, Fardin Javanmard, Larissa Walser, Rolf Kümmerli, Natasha Arora
Department of Bioscience Engineering, Department of Quantitative Biomedicine, Norwegian University of Life Sciences (Norway), Oslo University Hospital (Norway), University of Antwerp (Belgium), University of Zurich (Switzerland), University of Zürich (Switzerland), Zurich Institute of Forensic Medicine

Advances in sequencing technologies and bioinformatic tools have shown the potential of microbiome-based analyses for several applied settings like healthcare and forensics. In forensics, particularly the body-site specificity of bacterial communities has been exploited in various studies investigating body fluid identification. Typically, all these studies focus on sequencing specific variable regions of the 16S rRNA gene and use machine learning tools to report predictions. However, different studies focus on different stretches of the 16S rRNA gene, limiting the integration of datasets. In addition, the classifiers presented in recent studies lack systematic testing on forensically relevant settings like samples deposited on substrates, mixtures, and aged samples. In this study, we combined data from studies using different regions of the 16S rRNA gene with a closed-reference OTU clustering approach to train a random forest classifier. Subsequently, we systematically tested the classifier on mixtures and substrate samples generated in a controlled manner in the laboratory and on uncontrolled forensic settings. Uncontrolled forensic settings included samples from couples with extensive metadata on the body fluids to be expected in the sample, last intercourse, and showering habits among other information. We obtained high prediction probabilities particularly for high biomass samples like saliva, skin from hand and vaginal swabs. In addition, the classifier could predict whether sexual intercourse in couples was penetrative, oral, or both. Our results demonstrate the exciting potential of using a classifier trained on a heterogeneous training dataset in forensic settings especially in sexual assault cases.
Investigative Genetic Genealogy (IGG) is the practice of identifying people of interest using distant relatives' genetic data from genetic genealogy datasets. This tool has grown in popularity for forensic identification after it was used to identify the Golden State Killer in 2018. Still, the accuracy and reliability of this technique are unknown, with existing evaluations of the limitations focused on individuals with European ancestry and simple two-children-per-family pedigree models. We aim to quantify the accuracy of IGG with imperfect SNP panel data using population genetic simulations of realistic pedigrees with varying genetic ancestry. We present py_ped_sim, an open-source software with a dynamic pedigree structure and genome simulation framework. py_ped_sim is designed to support investigation of the accuracy of IGG, but can also be applied to genetic pedigree analyses in evolutionary and medical genetics. We validate our genomic pedigree simulation, showing a strong correlation between observed kinship and expected genetic kinship across six simulated families (average R²=0.90). We then utilize py_ped_sim to compare how kinship estimates were impacted by the genetic ancestry of founders across five super-continental population groups from the 1000 genomes consortium. We quantified differences in kinship estimates across the genetic ancestry of founders, considering the five super-continental groups. In addition, we use this simulation framework to test the accuracy of other popular kinship methods by generating pedigrees, including up to 6th-degree relationships. Overall, we provide a template for criminal justice practitioners to develop guidelines on the accuracy/limitations of IGG methodologies.
Recently developed (and developing) forensic genetic identification technologies have dramatically increased the reach of DNA evidence. However, the accuracy of these technologies varies with population genetic diversity in ways that have not be fully explored. This variation in accuracy must be quantified over realistic human population genetic diversity in order to reasonably interpret the results of these analyses in their impactful social context. One such technology is DNA mixture analysis, where a forensic sample contains DNA from multiple contributors, and investigators attempt to determine if a particular person of interest (POI) is one of those contributors. We show that misidentifications of non-contributors as contributors are more common for groups with relatively low genetic diversity. Further, using an inappropriate allele frequency can lead to inflated evidence for POI contribution, particularly when the reference allele frequency has relatively high genetic diversity. Another emerging technology is Investigative Genetic Genealogy (IGG) where a sample from a POI is compared to a genetic genealogy database to identify distant relatives. Through a genealogical investigation, a set of individuals who could be the POI are identified (POI pool), and then confirmed with direct genetic comparisons. We question if similar trends of differential accuracy are observed for IGG. To that end, we developed tools to quantify accuracy of the genetic kinship estimates across diverse genetic backgrounds, as well as to measure the size of the POI pool across family structures.
Investigative genetic genealogy (IGG) relies heavily on the accuracy of pedigrees used for analysis. Following its pivotal role in the identification of the Golden State Killer in 2018, this method has gained significant traction in forensic identification. However, little work has been done to understand how the structure of a pedigree influences the accuracy of results. Py_ped_sim has been created to simulate pedigrees structures and the genomes of individuals in a pedigree. One limitation of py_ped_sim and other pedigree simulators is that only descendants from one starting pair of root founders are simulated. This leads to an underestimation of cousins for individuals because we would have family data on one parent, but none on the other parent who is a non-root founder, that is, not a descendant of the root founder. I present two software programs that will aid in generating dynamic pedigrees via adding in parameters for misattributed paternity (MAP) and extending the family’s breadth. Family broadening simulates families onto these non-root founders and gives us a more accurate number of cousins for each individual. This software is made with a python interface and will be incorporated into the py_ped_sim software framework. We present validations of how extending the breadth of the family creates more distant cousin relationships providing a more accurate representation of genetic relatives. Overall, these suites of software will offer a valuable resource for advancing the field of IGG by providing dynamic pedigrees for comprehensive genetic investigation.
S36 - Greener and Sustainable Computing in Molecular Evolution: Methods, Algorithms, Tools, and Protocols
Green Computing in Molecular Evolution and Phylogenetics for the Global South

Beatriz Mello

Presented by self
Federal University of Rio de Janeiro (Brazil)

The computational demands imposed by modern science come with a significant environmental cost. The development and adoption of energy-efficient computing infrastructure, algorithms, and data storage are crucial for sustainable computing in molecular biology and evolution. Whenever computational tools developed for the same purpose provide similar accuracy, choosing greener options should be a commitment. For the Global South, where access to computational resources is frequently a limiting factor, green algorithms have a more profound impact on research development. This aspect of green computing promotes equity and diversity in our field. However, benchmarking efforts to determine the most environmentally efficient approaches are still needed. Studies establishing green methods will lead to more environmentally friendly bioinformatic analyses, besides prompting the Global South to play a bigger role in science. To illustrate this, I will provide a comparative example of the impact of green methods on molecular dating, which has been ranked among the top carbon footprint analyses in phylogenetics. Computationally efficient methods provide divergence time estimates equivalent to those obtained by time-demanding approaches. Among them, the relative rate framework implemented in RelTime was significantly faster and, consequently, more environmentally friendly than the penalized likelihood implemented in treePL. RelTime was orders of magnitude faster than statistically sophisticated Bayesian software. Thus, molecular dating, a mandatory step in many evolutionary studies, demonstrates how the adoption of green solutions not only addresses environmental concerns but also contributes to the advancement of research within scientific communities in low-resource settings.
Green Computing in Molecular Evolution and Phylogenetics for the Global South

Beatriz Mello

Beatriz Mello
Federal University of Rio de Janeiro (Brazil)

The computational demands imposed by modern science come with a significant environmental cost. The development and adoption of energy-efficient computing infrastructure, algorithms, and data storage are crucial for sustainable computing in molecular biology and evolution. Whenever computational tools developed for the same purpose provide similar accuracy, choosing greener options should be a commitment. For the Global South, where access to computational resources is frequently a limiting factor, green algorithms have a more profound impact on research development. This aspect of green computing promotes equity and diversity in our field. However, benchmarking efforts to determine the most environmentally efficient approaches are still needed. Studies establishing green methods will lead to more environmentally friendly bioinformatic analyses, besides prompting the Global South to play a bigger role in science. To illustrate this, I will provide a comparative example of the impact of green methods on molecular dating, which has been ranked among the top carbon footprint analyses in phylogenetics. Computationally efficient methods provide divergence time estimates equivalent to those obtained by time-demanding approaches. Among them, the relative rate framework implemented in RelTime was significantly faster and, consequently, more environmentally friendly than the penalized likelihood implemented in treePL. RelTime was orders of magnitude faster than statistically sophisticated Bayesian software. Thus, molecular dating, a mandatory step in many evolutionary studies, demonstrates how the adoption of green solutions not only addresses environmental concerns but also contributes to the advancement of research within scientific communities in low-resource settings.
CMAPEL: efficient phylogenetic inference in the pandemic era
Bui Minh

Presented by self
Australian National University (Australia)

We have recently introduced MAPLE (MAXimum Parsimonious Likelihood Estimation), a new pandemic-scale phylogenetic inference method exclusively designed for genomic epidemiology. In response to the need for enhancing MAPLE's performance and scalability, here we present two key components: (1) CMAPLE software, a highly optimized C++ reimplementation of MAPLE with many new features and advancements; and (2) CMAPLE library, a suite of Application Programming Interfaces for easy integration of the CMAPLE algorithm into existing phylogenetic inference packages. Notably, we have successfully integrated CMAPLE into the widely used IQ-TREE2 software, enabling its rapid adoption in the scientific community. These advancements serve as a vital step towards better preparedness for future pandemics, offering researchers powerful tools for large-scale pathogen genomic analysis.
CMAPLE: efficient phylogenetic inference in the pandemic era
Bui Quang Minh

Bui Quang Minh, Nhan Ly-Trong, Chris Bielow, Nicola De Maio
Australian National University (Australia), European Bioinformatics Institute (EMBL-EBI) (United Kingdom), Freie Universität Berlin (Germany)

We have recently introduced MAPLE (MAximum Parsimonious Likelihood Estimation), a new pandemic-scale phylogenetic inference method exclusively designed for genomic epidemiology. In response to the need for enhancing MAPLE’s performance and scalability, here we present two key components: (1) CMAPLE software, a highly optimized C++ reimplementation of MAPLE with many new features and advancements; and (2) CMAPLE library, a suite of Application Programming Interfaces for easy integration of the CMAPLE algorithm into existing phylogenetic inference packages. Notably, we have successfully integrated CMAPLE into the widely used IQ-TREE2 software, enabling its rapid adoption in the scientific community. These advancements serve as a vital step towards better preparedness for future pandemics, offering researchers powerful tools for large-scale pathogen genomic analysis.
A spectral step on the divide and conquer path to energy efficient phylogenetic reconstruction

Gavin Huttley

Robert McArthur, Michale Charleston, Ahad Zehmakan, Gavin Huttley
Australian National University (Australia), University of Tasmania (Australia)

The algorithms for phylogenetic reconstruction are central to computational molecular evolution. The relentless pace of data acquisition has exposed their poor scalability and the conclusion that the conventional application of these methods is impractical and not justifiable from an energy usage perspective. This is true even for state-of-the-art tools such as IQ-TREE 2. Furthermore, the drive to improve the statistical performance of phylogenetic methods produces increasingly parameter-rich models of sequence evolution, which worsens the computational performance. Is the statistical robustness versus computational performance tradeoff inescapable? Decades-old results, both theoretical and algorithmic, suggest we may be able to divide and conquer an escape. The combination of the Disk Coverage Method (DCM, Huson et al. 1999) and Min-Cut supertree (MCS, Semple and Steel 2000) provides a formal basis for expecting that large phylogenetic problems can be split up (DCM) and the results merged (MCS) to obtain a final result. MCS explicitly accommodates rooted topologies, which can arise from the more biologically plausible non-stationary models of sequence evolution. Here we present Spectral Cluster Supertree (SCS), a novel supertree method for merging overlapping rooted phylogenetic trees. It offers significant improvements over previous methods in terms of time complexity and accuracy, particularly for large problems. While Bad Clade Deletion (BCD, Fleischauer and Böcker 2017) generates more correct clades, SCS’s generated tree is generally topologically closer to the true tree. For datasets containing 10000 taxa and ~500 source trees, BCD takes ~2 hours, while SCS takes ~15 seconds.
Causes and impacts of the widespread lack of topological convergence in Bayesian phylodynamic inference on large viral datasets

Jiansi Gao

Jiansi Gao, Guy Baele, Frederick A Matsen

Fred Hutchinson Cancer Center (USA), Howard Hughes Medical Institute, KU Leuven (Belgium), Seattle (USA), University of Washington (USA)

Bayesian phylodynamic analysis of genomic datasets has been key for elucidating evolutionary and transmission dynamics of pathogens. However, topological convergence of these analyses on large viral datasets has not been comprehensively assessed. By carefully rerunning and analyzing 15 classic large phylodynamic analyses we show that: 1) convergence and mixing issues are widespread in tree topology sampling, which makes phylodynamic inference more computationally challenging; 2) despite apparent failure in topological convergence, for most datasets only a small fraction of clades and nucleotide sites exhibit poor mixing, suggesting that the inefficient exploration of tree space may frequently stem from a small number of viral sequences that possibly have undergone recombination or convergent evolution; 3) conflicting information in the substitution and branching processes may further exacerbate the difficulty in sampling viral time trees, reflected by the negative correlation between phylogenetic and coalescent likelihoods observed in most datasets, and; 4) the inferred treewise molecular and demographic processes appear to be minimally affected by poor exploration of tree space, whereas impacts on the estimated origin time and introduction history of particular clades are more pronounced. We identify biological properties of the viral datasets that may impede tree exploration and suggest new sampling mechanisms targeting these properties to improve the computational performance of Bayesian phylodynamic inference. Outputs from our long analyses (over one trillion generations across datasets) may serve as a comprehensive training dataset for machine-learning-based phylogenetic methods and provide valuable benchmarks for evaluating phylogenetic inference mechanisms, facilitating phylodynamic inference of large viral datasets.
Shifting Focus from Gene-centric to Evolution-centric Analyses in Cancer Genomics

Li Liu

Hai Chen, Li Liu
Arizona State University (USA)

Introduction: The absence of known driver genes in many tumors hampers precision treatment. We propose that a tumor’s somatic evolutionary characteristics, derived from aggregated mutations, can reveal hidden risk factors. Methods: We developed a novel method, I-SEE-IT to Identify Significant Evolutionary Events In a Tumor. Given whole-exome or whole-genome sequencing data, this method estimates the tumor-specific mutation rate and a set of subclone-specific parameters, including time at emergence, selection coefficient, duration of expansion, cell population fraction, and dN/dS. Using this method, we analyzed 517 multiple myeloma samples. Results: We discovered that early emergence and prolonged expansion of an advantageous subclone were associated with poor patient survival. We further identified a germline genetic variant in the TBKBP1 promoter region that predisposed a tumor to these evolutionary risk factors. RNA-seq analysis confirmed that this genetic variant was associated with differential expression of the TBKBP1 gene, and high expression of TBKBP1 was linked to poor patient survival. Given that TBKBP1 facilitates tumor-mediated immunosuppression, and treating multiple myeloma cells with Bortezomib induced anti-tumor immune responses, our results collectively imply that the TBKBP1 gene and its related signaling pathways are promising prognostic biomarkers and potential drug targets. Discussion: I-SEE-IT offers a comprehensive view of tumor evolutionary dynamics, uncovering novel genetic risk factors and potential therapeutic targets overlooked by gene-centric methods.
MEGA 12: Advancing Green Computing and Phylogenomics

Sudhir Kumar, Glen Stecher, Sudip Sharma, Koichiro Tamura
Temple University (USA), Tokyo Metropolitan University (Japan)

Molecular evolutionary and phylogenetic analyses play a pivotal role in evolutionary biology and functional genomics studies. However, the computational demands of maximum likelihood methods, frequently used in these studies, become high with increasing lengths of contemporary sequence alignments. The computational burden often exceeds the processing capabilities and memory of personal computers. This limitation is particularly acute in resource-constrained environments, impeding broad participation in evolutionary research. The 12th version of the Molecular Evolutionary Genetics Analysis, MEGA, software suite introduces novel statistical and computational advancements that dramatically enhance computational efficiency—achieving large reductions in time and memory required for selecting the best-fit substitution model, conducting bootstrap tests of phylogeny, and estimating evolutionary parameters such as branch lengths and divergence times. These advances will make phylogenomic analyses feasible on personal computers. They will also promote wider research participation, accelerate discoveries, and minimize energy consumption, resulting in accessible, sustainable, rigorous, and reproducible phylogenomics. MEGA 12 software suite will be made available for download at www.megasoftware.net.
S37 - IDEA
Contextual Analysis of Genetic Studies of Gender, Sex, and Sexuality
Miriam Miyagi

Presented by self
Brown University (USA)

Through two case studies, we will examine the importance of considering context in the study of sex, gender, and sexuality, not only due to the potential societal impacts of such work, via its political and social relevance, but also to ensure methodological and scientific rigor via careful analysis of sex-related variables. First, we will focus on modern, GWAS-driven studies of sexuality, tracing the lineage of this work to demonstrate that these studies experience the same pitfalls and ethical issues as historical research on the etiology of human sexuality, while also failing to capture variability in sexuality as we now know it. Second, we will explore what genetic studies of sex differences can learn from common methodological errors in general sex difference research, and consider how a contextualist approach to sex—that is, one that interprets sex as a context-dependent categorization system, offers solutions and paths forward for the rigorous study of sex-related variation.
Science Wise: From ‘hidden figures’ to scientific foremothers getting their flowers

Rori Rohlf

Presented by self

University of Oregon (USA)

Well before JDEI was assigned any value in academia, scientists from excluded groups (based on gender, race, ethnicity, ability, etc) found ways to navigate adversarial systems to make scientific contributions. One example is how in the 1970s women programmers and statisticians made substantive contributions to theoretical population genetics, even though their work was disproportionately credited only in acknowledgements, rather than by authorship. The humans who were so motivated that they went against the odds and found ways to share their scientific curiosity and insights changed our field, both with their technical contributions, and by reshaping scientific culture. Motivated by the urgent need to build upon the work and insights of these tenacious impactful scientists, we created Science Wise, an interview podcast with women from excluded groups who navigated academic science to earn PhDs in the 1970s and 80s. In this session, we will connect with their stories, learn from their experiences, and identify ways to integrate take home lessons to create a scientific culture where we thrive individually and collectively.
S38 - Open Symposium
Origin and evolution of specialised ribosomes across eukaryotes

Alan J. S. Beavan

Alan J. S. Beavan, Mary J O’Connell
University of Nottingham (United Kingdom)

Ribosomes are essential to all cellular life. Evidence is emerging that ribosomes are not the unprejudiced translational machines we once thought, thus challenging the idea that gene regulation is driven principally by transcriptional regulation. Instead, heterogeneous pools of ribosomes exist within and between cells and these structurally different ribosomes appear to have differential capacities for translation of different sets of mRNAs. It has been previously shown that ribosomes can be “specialised” in this way through the inclusion of closely related ribosomal protein paralogs with distinct sequences. To identify all such duplication events, we use gene tree-species tree reconciliation methods to elucidate the relationships between all ribosomal protein coding homologs, mapping gene duplication events that facilitate ribosome specialisation onto a eukaryote phylogeny. Further, we identify sites that have undergone rapid evolution in cases of specialisation, and in conjunction with 3D models of ribosomal complexes, propose specific regions of the ribosome that are most likely to facilitate specialisation. Our high-throughput method features the integration of standard phylogenetic methods with 3D protein structural methods - putting the evolution of specialised ribosomal complexes into a 3D context. In the future, these methods will be used not only to elucidate how specialisation has emerged in known cases but also to identify potential new cases of specialisation.
Gene expression variation is a major contributor to phenotypic variation across species. Numerous studies have sought to understand both the evolutionary processes and molecular mechanisms shaping across-species variation in gene expression. At the molecular level, mRNA abundances are a key regulator of protein abundances, which are the “functional units” of gene expression. Although these two levels of gene expression are undoubtedly correlated, comparative analyses revealed distinct evolutionary patterns of mRNA and protein abundances. Of particular note is the greater divergence across macroevolutionary timescales of mRNA abundances compared to protein abundances, leading to speculation of compensatory evolution across regulatory levels to maintain stable protein abundances. We developed a novel, mechanistic phylogenetic model of trait evolution to analyze the coevolutionary dynamics between mRNA and protein abundances across 11 mammalian species. Our model was fit to the data in a hierarchical manner, allowing us to share information across genes and compensating for the relatively small number of species analyzed. By doing so, we tease apart the relative contributions of mutation and natural selection in shaping mRNA-protein correlations across genes and species.
Effects of pesticides are driven by differences between species, not their chemical structure
Alicja Witwicka

Alicja Witwicka, Yannick Wurm, Federico Lopez-Osorio, Courtney May, Yeahji Jeong
Alan Turing Institute (United Kingdom), Queen Mary University of London (United Kingdom), UCL (United Kingdom)

The extensive use of pesticides unintentionally harms insect pollinators, but how their effects differ among species is not well understood. Pesticide toxicity is typically assessed using model species. Results are then extrapolated to thousands of other insects, underscoring the need for immediate investigation into species-specific effects. Comparing behaviour across species is challenging due to significant morphological and life history variations. However, high-resolution genetic tools provide more reliable and consistent results. By examining brain gene expression data, we explored the impact of pesticides on four insect species from three different orders: a butterfly, a blowfly, a solitary bee, and a bumble bee. Contrary to expectations, sulfoxaflor, considered safe for bees, elicited stronger responses in non-bee species than clothianidin, a neonicotinoid known for its high toxicity. Intriguingly, while each species showed overlapping sets of genes and metabolic pathways affected by different pesticides, the overlap between species was minimal. We show that pesticide response is driven by evolutionary differences between species rather than the chemical makeup of pesticides. Our results reveal that it is incorrect to extrapolate the effects of pesticide exposure from one species to others and call into question the current pesticide classification system based on the chemical structure rather than biological targets or effects. Highlighting species-specific reactions to human-made stressors, our research underscores the importance of incorporating molecular data to enhance pesticide safety evaluations and inform broader conservation strategies.
The strength and effects of stochastic and deterministic evolutionary forces are mediated by the size of the populations in which they act. To date, the most detailed efforts at describing population genomic variation have focused on species with small to moderately sized historical populations (e.g., human, Drosophila), with fewer characterizations at extremely large population sizes. Here, we use the Pacific acorn barnacle (Balanus glandula) as a model for the study of evolution across both large population sizes and geographic scales. B. glandula is widespread across intertidal ecosystems in the Pacific coast of North America, from Baja California to Alaska. This species lives at densities between thousands and tens of thousands of individuals per square meter, and have pelagic larvae capable of long range dispersal. We provide the first chromosome-level genome assembly for B. glandula, which we combine with large-scale spatial sampling, to describe the landscape of genome-wide genetic diversity across the species’ range. Additionally, we leverage spatial population genetic simulations to model the specific life history of barnacles to more accurately assess the combined effects of evolutionary forces in the observed patterns of genetic variation. By pairing empirical population genomics and spatial genetic simulations we hope to disentangle the dynamics of evolutionary processes in large barnacle populations and set the stage for similar studies in other large-population systems.
The first reference level genome of the Caribbean long-spined black sea urchin (Diadema antillarum)

Audrey Majeske

Audrey Majeske, Carlos Farkas Pool, Juliet Wong, Alejandro Mercado Capote, Alondra Diaz-Lameiro, Jose Eirin-Lopez, Nikolaos Schizas, Tarás Oleksyk
Duke University (USA), Florida International University (USA), Oakland University (USA), Universidad Católica de la Santísima Concepción (Chile), University of Puerto Rico Mayagüez (Puerto Rico)

We generated the first reference level nuclear genome assembly of the keystone Caribbean black long-spined sea urchin species (Diadema antillarum). We employed different assembly strategies using whole genome sequences acquired from multiple sequencing platforms (Illumina, PacBio and Oxford Nanopore) to generate a high-quality near-complete assembled genome. We compared the output metrics of several different assembly programs (hifiasm, flye and SPAdes), and then improved these through scaffolding using the RagTag tool. The efficacy of our assembly strategy was underscored by a BUSCO completeness score of 98.4% (based on metazoa_odb10), with the genome comprising 2789 contigs. The length of this resulting genome assembly is 1.75 Gbps with an N50 of 1.57 Mbps.
Single-nucleus multi-omics analyses reveal cellular and molecular innovations in the anterior cingulate cortex during human evolution

Bing Su

Jiamiao Yuan, Kangning Dong, Haixu Wu, Xuerui Zeng, Xingyan Liu, Yan Liu, Jiapei Dai, Jichao Yin, Yongjie Chen, Yongbo Guo, Wenhao Luo, Na Liu, Yan Sun, Shihua Zhang, Bing Su

Academy of Mathematics and Systems Science, Chinese Academy of Sciences (China), Kunming Institute of Zoology, Kunming Institute of Zoology (China), Renmin University of China (China), School of Mathematics, South-Central Minzu University (China), Wuhan Institute for Neuroscience and Neuroengineering

The anterior cingulate cortex (ACC) of the brain involves higher-level cognitive functions such as emotion and self-awareness, and is closely related to human mental disorders. The ACC of Old World anthropoids contains von Economo neurons (VENs), one of the noticeable cellular innovations during primate brain evolution. Here, we present the profiles of human and macaque ACC gene expression and chromatin accessibility at single-nucleus resolution. We characterized the conserved gene expression, chromatin accessibility, and transcription factor binding patterns in different cell types and neuronal subtypes. Importantly, by combining the published mouse data, we discovered the molecular identities and the possible cell lineage origin of the primate VENs. In particular, with both in vitro and in vivo experimental validations, we reported a group of human-macaque-shared and human-specific VEN marker genes, such as PCSK6, ADAMTSL3 and CDHR3, potentially contributing to the morphogenesis and function of VENs. Furthermore, we established the connections between the genome-wide human-specific mutations and the specialized gene regulation in the human ACC, delineating the genetic basis of cellular and functional innovations in the ACC during primate evolution and human origin. These findings provide new insights into understanding the cellular composition and molecular regulation of ACC and its evolutionary role in shaping human-owned higher cognitive skills.
Selective sweeps and deleterious mutations affect nucleotide diversity at linked sites, termed linked selection. Linked selection shapes nucleotide diversity through background selection and sweeps; however, the quantitative effect is not well understood. Through genetic linkage, natural selection reduces surrounding nucleotide diversity by modulating nearby allele frequencies. The span of this reduction depends on recombination rate, creating an association between recombination rate and diversity. Using the resulting relationship, we investigate the quantitative effects of linked selection within genomes and across species and how the impact changes with effective population size. We studied diversity patterns from recently published genome-wide diversity data from 233 primate species as a function of recombination rate. We lifted the data to hg38 coordinates and superimposed the human pedigree-based recombination map at a 100 kb scale where it is well conserved across primates. We then inferred the relationship between recombination rate and diversity in each species and modeled the effective population size for each species, to study how this relationship depends on an estimate of the overall Ne. We find a strong relationship between diversity and recombination rate in all species, highlighting the importance of linked selection shaping diversity in all primates. Additionally, the relative reduction in diversity in low/high recombination regions and the strength of the diversity recombination relationship increases with estimated population size, supporting the theory of stronger/more selection in larger populations.
Intrinsic expression and subcellular localization specificity in de novo emerged proteins.

Carly Houghton, Nelson Castilho Coelho, Annette Chiang, Aaron Wacholder, Allyson F. O’Donnell, Anne-Ruxandra Carvunis
Center for Evolutionary Biology and Medicine (USA), Department of Biological Sciences (USA), Department of Computational and Systems Biology (USA), School of Medicine, University of Pittsburgh

De novo gene birth is the process by which new protein-coding genes arise from previously non-coding DNA. Recent findings have unveiled thousands of de novo open-reading frames (ORFs) that are translated in cells. Upon emergence of a de novo ORF, the new protein encoded has had no prior natural selection on its protein sequence. Although certain de novo proteins adopt secondary structures and impact cellular fitness, little is known about their fate of following translation. Are they rapidly degraded or sequestered in specific subcellular niches? Do these unevolved proteins display a similar localization diversity to that observed for the conserved proteome? To address these questions, we used an integrated approach in which we genetically engineered yeast to express approximately 250 fluorescently-tagged de novo proteins under an inducible expression system. We then systematically determined their in vivo protein abundance and subcellular localization. Our findings reveal that, despite being expressed from a common promoter, de novo proteins exhibit sweeping differences in protein abundance, suggesting that they vary greatly in either translation efficiency or protein stability. Moreover, de novo proteins display distinct localization patterns compared to the conserved proteome, including being over and underrepresented in the mitochondria and the nucleus, respectively. Finally, we found that particular localizations are associated with unique sequence features within de novo proteins. This study provides the first systematic examination of de novo proteins in vivo and paves the way to understanding the potential phenotypic consequences and the evolutionary implications of their existence.
Environmentally driven changes in hunter-gatherer population interconnectivity explain African genetic diversity

Cecilia Padilla-Iglesias, Michela Leonardi, Zhe Xue, Johanna Paijmans, Margherita Colucci, Anahit Hovhannisyan, Lucio Vinicius, Andrea B Migliano, Andrea Manica

Department of Zoology, Evolutionary Ecology Group, Human Evolutionary Ecology Group, Human Paleosystems Group, Max Planck Institute for Geoanthropology (Germany), Trinity College Dublin School of Genetics and Microbiology (Ireland), University of Cambridge (United Kingdom), University of Zürich (Switzerland)

Emerging archaeological, anthropological and genetic evidence have challenged the traditional view that Homo sapiens originated from a single African region. Instead, it has been proposed that our species emerged and diversified within numerous, geographically distinct populations in Africa that became in contact at different points in time, exchanging genes and culture. Whilst analyses of contemporary and ancient genomes from African hunter-gatherer groups support deep divergence times between them, coupled with intermittent episodes of gene flow, these events are lacking a geographic and, perhaps more importantly, the climatic context in which they took place. Our study employs a Climate Informed Spatial Genetic Modelling (CISGeM) framework, integrating Species Distribution Models (SDMs) built from African hunter-gatherer archaeological assemblages and paleoclimatic reconstructions with data from all available contemporary and ancient African hunter-gatherer genomes. This allows formally assessing whether inferred geographic range changes from SDMs, inter-regional migrations and demographic fluctuations can be reconciled with genetics. We utilise Approximate Bayesian Computation to determine which demographic parameters (migration rates, admixture dates, growth rates...) are most supported. Our findings suggest that temporal patterns of African genetic diversity are attributable to climatic influences on population dynamics, and do not require the necessity for hypothesized introgression from "ghost" populations. We pinpoint key environmental factors promoting and hindering gene-flow between African regions and outline the historical geographical range of ancient hunter-gatherer populations. This research highlights the value of incorporating genomic data into spatially explicit models to elucidate the complex processes shaping human evolutionary trajectories.
Population genomics and spatial modelling provide insights into African human history

Cesar A Fortes-Lima, Carina M Schlebusch
Johns Hopkins University (USA), SciLifeLab (Sweden), University of Johannesburg (South Africa), Uppsala University (Sweden)

For the last decade, computational methods have been developed to investigate mass migrations in human history using genetic data. Here, we applied allele-frequency, haplotype-based, and spatial modelling approaches to reconstruct and visualise putative migration routes throughout the African continent. We first collected genetic and geographical data for 5,600 individuals from representative African populations from the whole continent as well as comparative Eurasian populations. Genetic data was generated in previous studies using Illumina arrays of over two million variants and whole-genome sequencing methods. Our spatial visualisation of ancestral components estimated using ADMIXTURE and effective migration rates estimated using FEEMS show i) putative migration routes of Bantu-speaking populations from west-central to east and south sub-equatorial Africa, ii) complex migration patterns of populations with different linguistic affinities across the Sahel belt, and iii) low migration patterns in North African populations due to low-migration rates across the Sahara Desert. We investigated haplotype data using MAPS, to infer both migration rates and population sizes of the studied populations and to disentangle their migration patterns over time. To better understand large expansions of African populations, we tested events of spread-over-spread, genetic continuity and genetic replacement. Our spatial modelling approach refines our understanding of corridors of higher and lower human migration or expansion across the African landscape that were confirmed with other sources. Therefore, this research highlights how spatially explicit analyses in population genomics can complement and investigate hypotheses from archaeology, biological anthropology, and historical linguistics.
The evolutionary history of the cycloamanide biosynthesis genes within Agaricomycetes species

Christian Quintero

CHRISTIAN QUINTERO-CORRALES, MELANIA VEGA, GRISELDA NALLELY HERNÁNDEZ-RICO, PABLO OCTAVIO-AGUILAR, ROBERTO GARIBAY-ORIJEL
Centro de Investigaciones Biológicas, Instituto de Biología, Jardín Botánico, UNAM (Mexico), Universidad Autónoma del Estado de Hidalgo (Mexico)

Horizontal gene transfer (HGT) is one of Eukaryotic life’s most interesting and complex evolutionary events. There is evidence that microorganisms share genes with other organisms, conferring potential evolutionary advantages. The cycloamanide, like amatoxins, are one of the most relevant fungal metabolites responsible for most worldwide deaths by mushroom poisoning. It has been suggested that involved genes passed to Galerina, Lepiota and Amanita sect. Phalloideae by HGT. Since several poisonings’ cases by the consumption of A. rubescens complex species in Mexico have been reported, we used de novo genome assembling and gene mining for cycloamanide homologous genes. The first record for Amanita bruennneolocularis and Amanita flavorubes in Mexico, together with the “A. cruentilemurum” samples has structurally similar genes to the known cycloamanide synthesis pathway. Among all studied samples, species from one particular site showed most of the cycloamanide synthesis pathway genes. Also, high cryptic diversity was found for the prolyl-oligopeptidase gene family showing a more complex evolutionary history than previously reported. Our results suggested that several old horizontal transfer and duplication events among Agaricomycetes fungal lineages, such as Piloderma, Russula, Cortinarius, Hebeloma, and more Amanita happened.
Compensatory mutations potentiate constructive neutral evolution by gene duplication

Christian R Landry

Christian R Landry, Philippe C Després, Alexandre Dubé, Rong Shi, Marie-Ève Picard, Jordan Grenier
Université Laval (Canada)

Protein functions generally depend on their assembly into complexes. During evolution, some complexes have transitioned from homomers encoded by a single gene to heteromers encoded by duplicate genes. This transition could occur without adaptive evolution through intermolecular compensatory mutations. Here, we experimentally duplicate and evolve an homodimeric enzyme to examine if and how this could happen. We identify hundreds of deleterious mutations that inactivate individual homodimers but produce functional enzymes when co-expressed as duplicated proteins that heterodimerize. The structure of one such heteromer reveals how both losses of function are buffered through the introduction of asymmetry in the complex that allows them to subfunctionalize. Constructive neutral evolution can thus occur by gene duplication followed by only one deleterious mutation per duplicate.
From diffusion to network: A Neurexin view of the Origin of Neural Synapse

Christine Guzman

Christine Guzman, Kurato Mohri, Hiroshi Watanabe
Okinawa Institute of Science and Technology (Japan), University of Fribourg (Switzerland)

In the nervous system, direct information exchange occurs through specialized intercellular junctions called synapses, whose protein organization and functional properties have been studied in Bilateria. However, the knowledge of when or how the functional synapse first emerged is still lacking. In this study, we used the non-bilaterian model Nematostella vectensis to investigate the role of Neurexins (Nrxns), a family of core pre-Synaptic Cell Adhesion Molecules (SAMs), in an early nervous system. We observed two classes of Nrxns: a broadly expressed ‘classical’ Nrxn (cNrxn), and a neuronal-specific Nrxn (nNrxn). Functional analysis of the cNrxn revealed its major involvement in cell-cell adhesion between ectodermal and endodermal epithelia. Meanwhile, knockdown of nNrxns resulted in abnormal behaviors associated to polyps’ muscle contractions. Further experiments suggested that nNrxns are involved in chemical transmission rather than peptidergic signaling. This study provides molecular, functional, and cellular insights into the ancestral non-neural function of Nrxns and may explain how and why this cell adhesion molecule family was employed in the synaptic machinery of the early nervous system.
On the evolution of gene product diversity
Daohan Jiang

Daohan Jiang, Nevraj Kejiou, Tina Qiu, Alexander Palazzo, Matthew Pennell
University of Southern California (USA), University of Toronto (Canada)

RNA and protein expressed from the same gene can have diverse isoforms due to various post-transcriptional and post-translational modifications, yet functional significance of the substantial gene product diversity generated by these processes remains largely unknown, raising the question whether it is generally adaptive or errors in biochemical processes. Additionally, testing these hypotheses has been challenging because it is often unclear how variation isoform abundance relates to the underlying evolutionary processes. Here, we introduce a generalized theoretical model where cis- loci (sequence motifs recognized by a modification enzyme), trans- loci (loci underlying the modification enzyme's expression and/or activity), gene expression level, and isoform decay rates collectively shape isoform abundances to predict evolutionary dynamics under different regimes of selection. We derive distributions of isoform abundances across species and genes at mutation-drift-selection balance, focusing on two types of gene product diversity: the first, exemplified by RNA editing, generates modified isoforms that may or may not be functional; the second, exemplified by RNA splicing, produces functional isoforms alongside potentially non-functional ones. We demonstrate that factors beyond the selection regime significantly influence evolutionary dynamics; importantly, a population is more likely have a sub-optimal mean phenotype when the effective population size is small, the number of cis-acting loci is large, and/or different genes are subject to different regimes of selection. Our modeling framework provides a quantitative approach for investigating gene product diversity in future empirical studies to advance our understanding of the evolutionary forces shaping molecular diversity.
The Codon Statistics Database: A Database of Codon Usage Bias

David Alvarez-Ponce

Krishnamurthy Subramanian, Bryan Payne, Felix Feyertag, David Alvarez-Ponce

Reno (USA), University of Nevada

We present the Codon Statistics Database, an online database that contains codon usage statistics for all the species with reference or representative genomes in RefSeq (over 15,000). The user can search for any species and access two sets of tables. One set lists, for each codon, the frequency, the Relative Synonymous Codon Usage, and whether the codon is preferred. Another set of tables lists, for each gene, its GC content, Effective Number of Codons, Codon Adaptation Index, and frequency of optimal codons. Equivalent tables can be accessed for (1) all nuclear genes, (2) nuclear genes encoding ribosomal proteins, (3) mitochondrial genes, and (4) chloroplast genes (if available in the relevant assembly). The user can also search for any taxonomic group (e.g., “primates”) and obtain a table comparing all the species in the group. The database is free to access without registration at http://codonstatsdb.unr.edu.
Genetic architecture of adaptation varies with the strength of selection
Deepa Agashe

Shivansh Singhal, Deepa Agashe
National Centre for Biological Sciences (NCBS) (India)

In new environments, populations face novel selection pressures, and adaptation to such conditions is of interest to evolutionary biologists, conservation biologists and ecologists alike. The genetic architecture of adaptation – i.e., the number and effect sizes of loci involved in adaptation, varies between species and sometimes even between populations of the same species. Theoretical models predict that the genetic architecture of adaptation changes with the strength of selection, but empirical evidence for this phenomenon is scarce. Under weak selection, we expect to see many loci with small effects on adaptive traits; whereas with strong selection, a few highly beneficial loci are likely to drive adaptation. We tested this using laboratory populations of the red flour beetle Tribolium castaneum adapting to novel resources (ancestral resource: wheat, novel resources: corn, finger Millet and sorghum, in decreasing order of selection pressure). After 35-60 generations of adaptation, we sequenced pooled genomic DNA from ancestral and evolved lines and quantified the change in allele frequencies over time. Compared to the ancestor, evolved lines had reduced nucleotide diversity and many more alleles at extreme frequencies. Supporting theoretical predictions, the magnitude of allele frequency change (a measure of effect size) as well as the number of alleles with large effect was correlated with the strength of selection imposed in each resource. Allele frequency changes were also highly parallel across populations, suggesting widespread beneficial pleiotropy. Thus, the genetic architecture of adaptation is associated with the strength of selection, suggesting broad predictability of adaptation in different environments.
Exploring independent evolution of light-dependent redox regulation in the plastids of Paulinella

Duckhyun Lhee

Duckhyun Lhee, Hwan Su Yoon
Republic of), SungKyunKwan University (Korea

Redox regulation plays an important role in the control of metabolic pathways in response to environmental changes. Within chloroplasts, thiol oxidation is linked to the activation of photosynthetic electron transport, allowing the modulation of chloroplast functions in response to light stimuli. In addition, evolutionary trends suggest an expansion in the number of redox-sensitive cysteines within the proteome, particularly coinciding with the emergence of plastids. In this study, we investigate the redox regulatory system in the plastid of Paulinella micropora, an organism possessing photosynthetic organelles derived from an independent primary endosymbiosis of Archaeplastida plastids. Using a quantitative proteomic approach employing resin-assisted enrichment coupled with isobaric tandem mass tag labeling, we investigate the redox dynamics of cysteine sites under light and dark conditions. Our results suggest that in Paulinella, the light-dependent redox regulation of key metabolic pathways within plastids has evolved independently from that observed in plants. However, the mechanism underlying the transfer of light-induced electrons to target proteins remains unclear. Furthermore, our analysis reveals a high percentage of oxidation among cysteines located in the plastid transit peptide regions of Paulinella, suggesting a possible role for in these cysteines.
Exploring the role of tRNA regulation in the transition to multicellularity of the amoeba Dictyostelium discoideum.

Dulce I. Valdivia Martínez

Dulce I. Valdivia Martínez, Peter F. Stadler
National University of Colombia (Colombia), The Santa Fe Institute (USA), University of Leipzig (Germany), University of Vienna (Austria)

In eukaryotes, transcription and translation occur in distinct cellular compartments. This division presents an intricate adaptation of regulation in which transfer RNAs (tRNAs) have to maintain the link between these two processes, as they must be transcribed within the nucleus and then exported to the cytoplasm to carry the amino acids needed to satisfy the ribosomal demand for proper messenger RNA-to-protein translation. Interestingly, this suggests that tRNA abundance could be an alternative way to regulate protein synthesis and their downstream associated phenotypes that it is independent of the regulation of their coding genes. Here, we explore the mechanisms that control changes in tRNA pools and their translational consequences during the development of unicellular Dictyostelium discoideum amoebae into multicellular fruiting bodies, a response that allows them to survive periods of starvation. Studying tRNAs at a transcript level is challenging because they are encoded in large gene families (e.g. ~400 genes in the 34Mb nuclear genome of D. discoideum) with many identical copies that produce short transcripts (~75 bp) and, additionally, premature tRNA molecules are cleaved into smaller fragments with potential alternative functions. To overcome this, we employed a combination of epigenomic datasets (ATAC-seq and ChIP-seq) to identify transcriptionally active tRNA gene loci and, in addition, we used LOTTE-seq data, a newly developed method to capture mature tRNAs. Our findings shed light on the fine-tuned contribution of tRNA abundance to adaptive phenotypes, either solely to meet amino acid demands or, potentially, as a crucial regulatory factor.
Nucleotide bias between leading and lagging strand in bacteria is caused by asymmetric mutagenesis
Eldar Badamshin

Eldar Badamshin, Alexey Skudnov, Stepan Denisov, Konstantin Gunbin, Konstantin Popadin
Higher School of Living Systems, Immanuel Kant Baltic Federal University (Russian Federation), The University of Manchester (United Kingdom)

The Mutation Accumulation Experiment (MAE) is a commonly employed technique to obtain the most pure mutational spectrum (Anjali Mahicar et al, 2022). These studies usually provide a 6-component mutational spectrum (2 base pair types*3 possible mutations) for different bacterial species. According to the parity rule 1 (hypothesis) frequencies of complementary mutations have to be equal (Sueoka, 1995). For this reason the 6-component mutation spectrum is commonly used in research. The approach doesn’t differentiate between complementary mutations (e.g. C>A and G>T), thus reducing 12 possible mutations to 6. Using MAE data we created 12-component mutation spectrum of four species: Vibrio cholerae, Aliivibrio fischeri, Bacillus subtilis and Dienococcus radiodurans (Dillon et al., 2016, Sung et al., 2015, Long et al., 2015), taking into account replication asymmetry and that many mutations may arise due replication errors (Horton et al., 2023). We showed that frequencies of complementary mutations are not equal on leading and lagging strands. We discovered that the GT-rich leading strand is characterized by C>T and A>G mutations that outnumber the corresponding complementary mutations. This observation is true for all chromosomes (main and secondary) for C>T and A>G mutations in MutS/L-deficient lines, and for C>T in wild-type of all studied species. We assume that C>T, A>G mutations occur more frequently in the template ssDNA for lagging strand (Frederico et al., 1990), which spends more time being single-stranded compared to the leading strand template.
Conservation Genomics of North American Jaguars
Eldridge Wisely

Eldridge Wisely, Natalie Payne, Ivonne Cassaigne, Karla Vargas, Victor H Luja, Melanie Culver
Arizona State University (USA), Primero Conservation (USA), U.S. Geological Survey (USA), Universidad Autónoma de Nayarit (Mexico), University of Arizona (USA)

The jaguar (Panthera onca) was once found throughout the Americas. Now, jaguars are found in only a fraction of their former range. They suffer from habitat fragmentation, human-wildlife conflict, and illegal wildlife trafficking driving them toward extinction. Along with habitat fragmentation comes the potential for increased inbreeding and deterioration of genetic diversity - necessary for continued adaptation to a changing environment and climate. Jaguars living at the edge of their range are even more at risk because of reduced migration opportunities, but may also harbor unique adaptive and neutral genetic variation due to selection or drift. Whole genome sequencing can aid management decisions for endangered species through increased resolution compared to traditional population genetics, which improves the ability to scan for local adaptations or alleles involved in reduced fitness. We used an individual-based sampling scheme to investigate the genomic characteristics of 14 free-ranging North American jaguars representing 7 Mexican states and Arizona over the last 30 years. We sequenced whole genomes of these jaguars from blood or tissue samples and mapped reads to the annotated tiger (Panthera tigris) genome. We inferred levels of connectivity and gene flow between the North American sub-populations represented by these jaguars, which reflects the level of landscape connectivity in North America, with respect to jaguar dispersal and migration. Additionally, we leveraged 30 years of temporal sampling to trace the levels of heterozygosity and inbreeding of these North American jaguars over time. Here we present our goals, methods, and preliminary findings.
Unraveling natural Variants in the Gene Regulatory Network underlying Root Hair Formation: A Comprehensive Analysis of 855 Arabidopsis thaliana Accessions

Elsa Herminia Quezada Rodríguez

Elsa H Quezada Rodríguez, Alma Piñeyro Nelson, Adriana Garay Arroyo, Ana M Almeida, Elena Alvarez Buylla
Centro de Ciencias de la Complejidad, Ciudad Universitaria, Coyoacán, Departamento de Ecología Funcional, Departamento de Producción Agrícola y Animal, East Bay. 25800 Carlos Bee Blvd. Hayward/CA. US. 94542 (Mexico), Green Biome Institute. California State University, Instituto de Ecología, México D.F. 04510. (Mexico), Universidad Autónoma Metropolitana Xochimilco (Mexico), Universidad Nacional Autónoma de México. 3er Circuito Exterior

A plant’s capacity to adapt and respond to diverse challenges is enhanced by the plasticity of the root. Root hairs, thin tubular structures elongating in the root’s differential zone, play a pivotal role in water absorption, nutrient uptake, and microorganism interaction. In this research, insights into the genetic diversity of genes that govern root hair development in diverse Arabidopsis accessions was analyzed. We employed computational methods to assess genetic variations in genes that comprise the Gene Regulatory Network (GRN) underlying the development of the hair/no-hair pattern in primary root, analyzing 855 accessions distributed around the world. At the protein level, we conducted an exhaustive analysis that included constructing phylogenetic trees, identification of domains, motifs, families, Posttranslational modifications (PTMs), protein features, secondary and tertiary protein structure. Through a Single Nucleotide Polymorphism (SNP) search in the GRN, SABRE, EGL3, GL3 and GL2 were found to have a higher number of SNPs. In these four genes, we identified 22 accessions with greater variability in SABRE, 44 for EGL3, 48 for GL3, and 28 for GL2. Moreover, our study highlights five accessions with at least two significant variabilities in one of the four analyzed genes. Finally, Chat-1, Lov-1, Sq-1, and Old-1 accessions were utilized for phenotyping, revealing variations in the trichoblast and atrichoblast patterns, as well as in the quantity of root hairs. These findings provide valuable insights into the intricate molecular processes contributing to root hair developmental plasticity in naturally occurring Arabidopsis accessions.
Chloroplast Phylogenomics and the Taxonomy of Mesoamerican Agaves

Elsa Peters Ruiz de Chávez

Elsa Peters Ruiz de Chávez, Abisaí García-Mendoza, Alejandro Casas, Daniel Piñero, Logan Kistler, Marcela Sandoval-Velasco

Anthropology Department, Centro de Ciencias Genómicas, Instituto de Biología, Instituto de Ecología, Instituto de Investigaciones en Ecosistemas, Jardín Botánico, National Museum of Natural History, Posgrado en Ciencias Biológicas, Smithsonian Institution (USA), Universidad Nacional Autónoma de México (Mexico)

Agaves are one of the most emblematic plant groups in Mexico, where they thrive and dominate various ecosystems. Their importance resides not only in the great diversity and broad ecological interactions these plants have but also in the multiple uses humans have given to them for thousands of years. Consequently, Agaves have been in the spotlight of intensive research for many years which has led to discussions about their taxonomy and attempts to reconstruct the molecular phylogeny of the genus. However, most studies have been based on few and small chloroplast and nuclear DNA sequences and important phylogenetic relationships have been left undetermined due to a lack of data resolution, which limits testing some relevant evolutionary hypotheses. Since taxonomy and molecular systematics provide frameworks for understanding biological interactions within a genus, both should be considered when trying to understand evolutionary relationships. However, for Agave these approaches have been insufficiently integrated. In this study, we generated chloroplast data for 160 agave samples from both herbarium and living specimens representing 65 species. We de novo-assembled and circularized chloroplast genomes which we use to reconstruct a robust phylogeny. We use this phylogeny as a biological framework for testing evolutionary hypotheses about niche conservatism and niche diversification for species historically managed by humans in Mesoamerica. Such integration looks to understanding how human management has impacted the geographic distribution and molecular phylogenetic relationships among Agave species, which in turn has modified the ecological niche and evolutionary history of the genus.
Seascape genomics and genetic basis of adaptation to low salinity of Harbour porpoises (Phocoena phocoena) across environmental gradients in the North Atlantic and adjacent waters

Enrique Celemin Amaro

Enrique Celemin Amaro, Ralph Tiedemann
Potsdam University (Germany)

The Harbour porpoise (Phocoena phocoena) is a highly mobile cetacean species primarily occurring in coastal and shelf waters across the Northern hemisphere. It inhabits heterogeneous seascapes broadly varying in salinity and temperature. Here we produced 82 whole genomes at intermediate coverage to study Harbour porpoise's evolutionary history and investigate the role of local adaptation in the diversification into subspecies and populations. We identified ~6 million high quality SNPs sampled at 8 localities across the North Atlantic and adjacent waters, which we used for population structure, demographic, and genotype-environment association analyses. Our results suggest a genetic differentiation between three subspecies (P.p. relicta, P.p. phocoena and the recently proposed P.p. meridionalis), and three distinct P.p. phocoena populations occurring in the Baltic: Atlantic, Belt Sea and Proper Baltic Sea. Effective population size and Tajima's D levels suggest a population contraction in Black Sea and Iberian porpoises, but a population expansion in the P.p. phocoena populations. Genotype-environment association analysis identified salinity as the major environmental driver in genomic variation and we identified candidate genes putatively underlying adaptation to different salinity levels in Baltic porpoises. Our study highlights the value of whole genome resequencing to unravel the genetic basis underlying adaptation to brackish waters and shows how strong environmental gradients may lead to population differentiation. The results have great conservation implications as we found inbreeding and low genetic diversity in the endangered Black Sea subspecies and identified the critically endangered Proper Baltic Sea porpoises as a separate population.
Dinotoms possess two evolutionary distinct autophagy-related ubiquitin-like conjugation systems

Euki Yazaki

Euki Yazaki, Tadaaki Uehara, Hirokazu Sakamoto, Yuji Inagaki
Center for Computational Sciences, Chiba University (Japan), Department of Infection and Host Defense, Department of Pathology, Graduate School of Life and Environmental Sciences, Graduate School of Medicine, National Agriculture and Food Research Organization (Japan), RIKEN iTHEMS (Japan), Research Center for Advanced Analysis, Stanford School of Medicine (USA), University of Tsukuba (Japan)

Autophagy is an intracellular degradation mechanism by which cytoplasmic materials are delivered to and degraded in the lysosome-fused autophagosome (autolysosome) and proposed to have been established at an early stage of eukaryotic evolution. Dinoflagellates harboring obligate endosymbiotic diatoms (so-called “dinotoms”), which retain their own nuclei and mitochondria in addition to plastids, have been investigated as an intermediate toward the full integration of a eukaryotic alga into the host-controlled organelle (i.e., plastid). Pioneering studies systematically evaluated the degree of host governance on several metabolic pathways in the endosymbiotic diatoms (ESDs). However, little attention has been paid to the impact of the endosymbiotic lifestyle on the autophagy operated in the ESDs. In this study, we searched for ATG3, ATG4, ATG5, ATG7, ATG8, ATG10, and ATG12, which are required for autophagosome formation, in the RNA-seq data from dinotoms Durinska baltica and Kryptoperidinium foliaceum. We detected two evolutionally distinct sets of the ATG proteins in the dinotom species, one affiliated with the dinoflagellate homologs and the other with the diatom homologs in phylogenetic analyses. The results suggest that the ATG proteins descended from the diatom taken up by the dinoflagellate host persist for autophagosome formation and, most likely, autophagy.
The manifestation of hybrid sterility requires a greater number of sterility genes in heterozygous introgressing segments than in homozygous introgressing segments.

Ezel Jacome Galindo-Pérez

Ezel J. Galindo-Pérez, Blanca E. Chávez-Sandoval, Lilian M. Barandica-Cañon, Francisco García-Franco
Departamento El Hombre y su Ambiente. Universidad Autónoma Metropolitana. Unidad Xochimilco. (Mexico), Departamento de Ciencias Básicas, Departamento de Zoología, Universidad de Cartagena. (Colombia), de la Escuela Nacional de Ciencias Biológicas. Instituto Politécnico Nacional. (Mexico)

In this bioinformatic study we identified all the orthologous genes of D. buzzatti, D. mojavensis and D. melanogaster present in six homozygous sterility-producing segments, of different sizes and distributed along chromosome 4 of D. buzzatii. By their association with the respective mutant sterility alleles, each of these genes was classified into one of three categories: female sterility genes, male sterility genes, and both sex sterility genes. Surprisingly, we identified a greater number of female sterility genes and a smaller number of male sterility genes and both sexes sterility genes. All of them distributed heterogeneously along chromosome 4. We conclude that: there are twice as many major oogenesis genes as there are major spermatogenesis genes, heterogeneously distributed across all autosomes; in each introgressive segment, between 3 to 11 spermatogenesis genes and at least 20 oogenesis genes are required to be placed in a homozygous recessive state to produce the respective sterility of each sex; each female sterility segment always includes the minimum number of male sterility genes sufficient to produce sterility in both sexes; To produce male sterility, at least 45 male sterility genes must be put into recessive heterozygosity (30% heterozygous segments, 15 times larger than the 2.21% homozygous segments); The large number of major oogenesis genes allows for a very robust oogenesis process that practically prevents the manifestation of female sterility and/or lethality by introgresant heterozygous segments.
Low complexity regions across the Tree of Life: sources of diversity or just noise?

Fabia Ursula Battistuzzi

Fabia Ursula Battistuzzi, Vineet Saravanan, Krysta Krawczyk
Cranbrook Upper school (USA), Oakland University (USA)

Low complexity regions (LCRs) are loosely defined as regions of a genome with lower than expected nucleotide or amino acid diversity. Traditionally considered non-functional parts of a genome, they have also been suggested as potential sources of genomic variability. Most of our knowledge regarding LCRs comes from eukaryotic genomes that, with their large amount of non-coding DNA host relatively large numbers of LCRs. However, despite their more compact nature, prokaryotic genomes also have a consistent presence of LCRs, suggesting that similar evolutionary forces driving their origin and evolution may be shared by the three domains of life. The availability of thousands of prokaryotic genomes provides a unique opportunity to evaluate the evolutionary history of LCRs at the population and species levels, thus shedding light on their potential role in genomic variability. In this study, we will present a computational analysis of thousands of prokaryotic species and some example populations to reconstruct patterns of presence/absence and composition of LCRs within these genomes.
Germline mutations arise from a combination of DNA damage and replication errors. One hypothesis posits that variation in the efficiency of DNA repair in the germline explains the observed variation in mutation rates across species, although this has mainly been explored in mammals. Another possibility is that exogenous mutagens also contribute to mutation rate variation. Different mutational processes cause different types of mutation “signatures”, and we can exploit this to measure their relative importance. To explore the distribution of mutational processes across a broad phylogenetic context, we applied Baymer, a Bayesian hierarchical tree model, to infer windows of nucleotide sequence contexts (up to 4 flanking nucleotides, i.e., ‘9-mers’) that capture variation in mutation rates across 70 eukaryote species using polymorphism datasets. We applied phylogenetic linear mixed models to explore the role of biological and environmental modifiers of C > T mutations at CpG sites. We confirmed that variation in CpG > T mutation rates is highly correlated with germline methylation levels, with patterns consistent with the phylogeny. Additionally, we analyzed mutational signatures from the Catalogue Of Somatic Mutations In Cancer (COSMIC), which are believed to represent different mutational mechanisms that could contribute to germline mutagenesis. We explore the distribution of the 86 COSMIC signatures across species, reflecting both genetic mechanisms and potential environmental mutagens. These results allow us to explore the factors affecting mutation rates with data that encompasses a wide range of biological and environmental diversity.
Gene and allele specific expression underlying the electric signal divergence in African weakly electric fish

Feng Cheng

Feng Cheng, Alice B Dennis, Otto Baumann, Frank Kirschbaum, Salim Abdelilah-Sefried, Ralph Tiedemann
Humboldt University (Germany), University of Namur (Belgium), University of Potsdam (Germany)

In the African weakly electric fish genus Campylomormyurus, electric organ discharge (EOD) signals are strikingly different in shape and duration among closely related species, contribute to pre-zygotic isolation and may have triggered an adaptive radiation. We performed mRNA sequencing on electric organs (EOs) and skeletal muscles (SMs; from which the EOs derive) from three species with short (0.4 ms), medium (5 ms), and long (40 ms) EODs and two different cross-species hybrids. We identified 1,444 up-regulated genes in EO shared by all five species/hybrids cohorts, rendering them candidate genes for EO-specific properties in Campylomormyurus. We further identified several candidate genes, including KCNJ2 and KLF5, their up-regulation may contribute to increased EOD duration. Hybrids between a short (C. compressirostris) and a long (C. rhynchophorus) discharging species exhibit EODs of intermediate duration and showed imbalanced expression of KCNJ2 alleles, pointing towards a cis-regulatory difference at this locus, relative to EOD duration. KLF5 is a transcription factor potentially balancing potassium channel gene expression, a crucial process for the formation of an EOD. Unraveling the genetic basis of the species-specific modulation of the EOD in Campylomormyurus is crucial for understanding the adaptive radiation of this emerging model taxon of ecological (perhaps even sympatric) speciation.
Host-microbe systems are unique evolutionary niches that produce complex biological interactions that neither participant could evolve on their own. Over the last two decades the influence of host-microbe interactions on global health has begun to emerge and with it the greater biological research community's interest in host-microbe interactions and the evolution of these systems. However, these systems have historically been a complex and difficult field of biological research. Impactful advances in global health will be obtained by progressing our understanding of these systems and laying the groundwork for future bioengineering projects. Protein mimicry, wherein a microbe encodes a protein that closely resembles and mimics the structure and function of a host's, is an important aspect of host manipulation in these systems. We have utilized novel protein structure prediction and alignment tools in order to explore the structural proteome for examples of protein structure mimicry previously undetectable by sequence based approaches. By leveraging the Legionella pneumophila proteome and its many known structural mimics, we have developed and validated a framework that can be applied to virtually any host-microbe system to uncover signals of protein mimicry, identifying candidates of host control in microbial proteomes. We successfully identified undiscovered effector candidates as well as determined host targets of previously known effectors in three microbe proteomes important to human health and disease; Legionella pneumophila, Helicobacter pylori, and Wolbachia. We also further supported the evidence of one of the Wolbachia mimic candidates through functional assays in the Drosophila-Wolbachia system.
Transcriptome reference-based SNP calling as an alternative for SNP annotation in the absence of a reference genome for invasion genomics studies.

Gabriela Castellanos-Morales

José Francisco Miranda-Vidal, Yocelyn T. Gutiérrez-Guerrero, Anahí Canedo-Téxon, Maricela García-Bautista, Everardo Barba, Gabriela Castellanos-Morales

El Colegio de la Frontera Sur, UC Berkeley (USA), Unidad San Cristóbal de las Casas (Mexico), Unidad Villahermosa (Mexico)

For many invasive species, as in other non-model species, reference genomes for SNP annotation are lacking. This limits our ability to identify candidate loci to advance our understanding about how these species adapt to new environmental conditions. In such cases, available transcriptomic data can be used to annotate candidate SNPs. We used this approach to study the invasive genomics of the suckermouth armored catfish (Pterygoplichthys sp.) in the Grijalva-Usumacinta River Basins. These South American fish have been introduced and become invasive in many areas of Asia and North America. In invaded areas, they threaten biodiversity and cause economic loss for freshwater fisheries. Previous studies based on morphological and genetic data suggested that hybridization could be playing a role in invasion success for these taxa. In the present study, we conducted genotyping-by-sequencing for 103 samples from 5 localities representing an environmental gradient in the Grijalva and Usumacinta River Basins in Mexico. We assembled and annotated the transcriptome for Pterygoplichthys pardalis from available data at the NCBI database. We conducted de novo and reference SNP calling, using P. pardalis reference transcriptome to map reads. We obtained 22,595 and 3,083 SNPs, respectively. Patterns of genetic variation and genetic structure were similar between datasets, and our results support the hypothesis of a hybrid origin for this invasive population, and rapid population expansion. Moreover, the transcriptome referenced database allowed identifying candidate loci associated to stress response. Transcriptome reference-based SNP calling is an alternative in the absence of a reference genome.
Exploring genomic correlates of multi-trait phenotypic convergence in trap-jaw ants

Gaurav Agavekar

Gaurav Agavekar, Yoshihiko Suzuki, Michael Hiller, Gene Myers, Evan P Economo
LOEWE Centre for Translational Biodiversity Genomics (Germany), Okinawa Institute of Science and Technology (Japan), Senckenberg Research Institute (Germany)

Convergent evolution has produced remarkable adaptations throughout the tree of life. A central question is whether convergence at the genomic level underlies phenotypic convergence, and if so, what are the relative contributions of protein-coding vs. regulatory convergent changes? We ask this question in Strumigenys, a hyperdiverse ant genus in which trap-jaw mandibles (TJ), a phenotype characterized by morphological, behavioral, and neural adaptations, has evolved independently in different biogeographic regions. These ants employ TJ as a high-velocity weapon to hunt fast prey items such as springtails, and produce the fastest acceleration observed in resettable animal movements with their mandibles. We took a comparative genomic approach to analyze three independent instances of trap-jaw evolution in Strumigenys. To this end, we newly sequenced 18 chromosome-scale, highly accurate genomes (~99.9999% base-level accuracy) and generated comprehensive annotations. We then separately analyzed coding genes and conserved non-exonic elements (CNEs) to explore the genomic signatures of phenotypic convergence. In our analyses, we do not find strong evidence for convergence in protein-coding genes, but we do see multiple independent accelerations in CNEs in trap-jaw clades. These CNEs exhibit substantial overlap with ATAC-seq peaks, corroborating their functional significance in regulatory processes. Although our exploration of convergence within these CNEs is ongoing, preliminary findings suggest that regulatory regions, rather than protein-coding sequences, play an important role in multi-trait phenotypic convergence among Strumigenys ants.
Genomes often exhibit heterogeneity across chromosomes, and distributions of protein-coding genes, repetitive elements, and polymorphisms are not uniform along chromosomes in multiple species. One explanation for these patterns is recombination rate variation. As recombination interacts with selection to shape the evolutionary fates of alleles, variation in recombination rate within and between chromosomes could promote differences in the distribution of genomic features across chromosomes. In the nematode Caenorhabditis elegans, recombination rate correlates with multiple genomic features that are non-uniformly distributed along chromosomes. Specifically, in C. elegans, recombination rates are higher on chromosome ends compared to chromosome centers. Its closest known relative, C. inopinata, harbors a radically altered genome with nearly uniform chromosomal distributions of repetitive elements and protein-coding genes; it likewise has a less heterogeneous chromosomal distribution of polymorphisms. Is this dramatic change in genomic organization in any way connected to the evolution of recombination rates? Here, we describe ongoing efforts to infer genetic maps in C. inopinata. We performed whole-genome sequencing on 180 individual F2 recombinants. Tentative, preliminary results reveal some chromosomes may have conserved recombination rate domain structure whereas other chromosomes may harbor divergent, more uniform recombination rate distributions. We intend to compare these chromosome-level recombination rates with genomic landscapes of genes, repeats, and polymorphisms across multiple species to understand the extent to which recombination rates predict genomic organization.
A geological timescale for bacterial evolution and oxygen adaptation
Gergely J Szöllösi

Gergely J Szöllösi, Adrian A Davin, Philip Hugenholtz, Tom A Williams
(Hungary), Centre for Ecological Research, Institute of Evolution, Okinawa Institute of Science and Technology Graduate University
(Japan), The University of Queensland (Australia), University of Bristol (United Kingdom), University of Tokyo (Japan)

Microbial life has dominated the biosphere throughout Earth’s history but has left a meagre fossil record, greatly hindering our understanding of evolution in deep time. However, the co-evolution of life and the Earth system has left signatures of microbial metabolism in the geochemical record, most conspicuously the Great Oxidation Event (GOE) ~2.33 billion years ago (Ga, (Poulton et al. 2021)). The GOE transformed Earth’s biosphere from dominantly anoxic to oxic via oxygenic photosynthesis and carbon burial via tectonism (Eguchi, Seales, and Dasgupta 2019). Here, we combine machine learning and phylogenetic reconciliation to infer ancestral transitions to aerobic lifestyles during bacterial evolution. Linking these transitions to the GOE provides new constraints to infer the timetree of Bacteria. We find that extant bacterial phyla are truly ancient, tracing their diversity back to the Archaean and the Proterozoic: the oldest include Bacillota (Firmicutes) that radiated 3.1-3.7 Ga, Cyanobacteriota (3.3-3.5 Ga) and Patescibacteria (3-3.5 Ga). We show that most bacterial phyla were ancestrally anaerobic and transitioned to aerobic lifestyles after the GOE. However, we infer that the earliest oxygen-adapted Bacteria pre-dated the GOE and facilitated the evolution of oxygenic photosynthesis in the cyanobacterial ancestor.
Physiographic features on Earth’s surface are long thought to limit and structure populations. That structuring occurs through the non-random reduction of gene flow associated with those features, resulting in partitioning of genomic variation across space. This can be contextualized as a form of information transfer between Earth's landscape and the lineages evolving with that landscape. As a proof of this concept, we simulated genomic data of diverging lineages, varying migration, mutation and recombination rates, and effective population size. We calculated standard pairwise genetic metrics genome-wide and then calculated the partial information decomposition of those genetic metrics to determine if any of the decomposed nodes held information that predicted the evolutionary history of those lineages, namely whether they diverged with or without gene flow. Considering the 7,648,800 nodes, the information content of nodes involving Tajima’s D and nucleotide diversity performed predicted gene flow history better than current SFS-based tools. This suggests that the context of lineage divergence (with or without gene flow) can be recovered with information theoretic measures calculated on the evolving sequences. If physiographic features do limit some inter-population gene flow in nature, then this suggests there is information transfer from Earth’s landscape to populations of species which is encoded in the partitioning of genomic variation across space. Each species is a separate receiver of this communication, where species evolving with high fidelity to the landscape receive this information through a ‘noiseless’ channel, while others (e.g., panmictic species) receive this message through a noisy channel.
Incipient genetic differentiation of A. ludens (Diptera: Tephritidae) as a result of its recent geographic and host expansion leading to the attack of commercially grown apples in Mexico.

Helena Socorro Hernández-Rosas

Helena S. Hernandez-Rosas, Daniel Helena S. Piñero, Nancy Helena S. Nancy Gálvez-Reyes, Alma Altúzar-Molina, Martín Aluja
AC–INECOL, Clu?ster Cienti?fico y Tecnolo?gico BioMimic (Mexico), Instituto de Ecologi?a, Instituto de Ecología, Universidad Nacional Autónoma de México (Mexico)

The fruit fly Anastrepha ludens is a significant citrus pest in Mexico, primarily feeding on citrus pulp and other cultivated and wild species. Typically found in humid tropical regions below 500 meters altitude, its range has expanded to higher elevations due to warming, even reaching apple crops. Through population genetics, focusing on single nucleotide polymorphisms (SNPs), we discovered A. ludens exhibits greater genetic diversity (\(\theta = 0.176\)) than its sister species, A. obliqua (\(\theta \sim 0.0105\)). Flies infesting apples displayed extensive polymorphism (99%), contrasting with grapefruit-infesting flies, which were more conserved (50%). Principal Component Analysis identified genetically distinct apple parasites. Admixture analysis revealed genetic differentiation in apple-collected flies. Despite low differentiation (FST = 0.0125) among samples, a distinct group of apple-infesting flies closely related to pear parasites emerged. Evidence suggests recent diversification of apple-infesting fly lineages. Our study supports A. ludens possibly undergoing early population differentiation driven by host plants. Focusing on apple varieties, our findings indicate ongoing ecological divergence.
A new scenario for the macroevolution of the Caulimoviridae based on the analysis of endogenous caulimovirids using CAULIFINDER

Héléna Vassilieff

Héléna Vassilieff, Nathalie Choisne, Florian Maumus, Pierre-Yves Teycheney, Saad Serfraz, Andrew D.W Geering
CABB, CIRAD, INRAE, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland (Australia), UMR PVBMT (France), URGI (France), University Paris-Saclay, University of Agriculture Faisalabad (Pakistan)

Endogenous viral elements (EVEs) result from the integration of viral sequences into the genome of their hosts. EVEs can be considered molecular fossils and provide access to ancient or unknown viral sequences that can be used in paleovirological approaches to reconstruct the evolution of related viruses over extended periods of up to tens of millions of years. In plants, most characterized EVEs belong to the family Caulimoviridae, the only family of retrotranscribed plant viruses, and are termed endogenous caulimovirids (ECVs). Previous studies of ECVs have proposed two alternative scenarios to explain the evolutionary history of the Caulimoviridae: a coevolution scenario and a host-switching scenario. Although different, these two scenarios assume a very early emergence of the Caulimoviridae, which may well have preceded that of the euphyllophytes. However, the bioinformatic analyses on which these studies were based were limited by the lack of automated tools for annotating ECVs in plant genomes and the underrepresentation of basal plant genomes in their datasets. To address these issues we developed CAULIFINDER, a tool for the automated detection and annotation of ECVs in plant genomes, and used it to analyze a dataset of 56 genome assemblies covering all subdivisions of the tracheophytes, including basal ones such as ferns and lycophytes. We discovered 1449 representative ECVs that were organized in 35 new taxonomic units among the Caulimoviridae. By intersecting their phylogenetic relationships with the taxonomy of host plants, we propose a new macroevolutionary scenario for the family Caulimoviridae, reconciling host-virus coevolution and host swaps.
The effect of genetic inheritance impact on de novo mutation patterns in Macaca fascicularis.

Hye Ri Park

Hye-ri Park, Hyeon-Mu Cho, Se-Hee Choe, Min-Gyeong Ko, Yun-Jung Lee, Sang-Je Park, Young-Hyun Kim, Jae-Won Huh

Korea Research Institute of Bioscience and Biotechnology(KRIBB) (Korea), Korea University of Science and Technology (UST) (Korea, Republic of)

Mutations arise from errors in the DNA replication process or exposure to environmental elements like radiation or chemicals. And occurring within germ cells possess the potential to be passed down to successive generations, perpetuating their genetic legacy. Whether the process of transferring unrepaired mutations to the next generation is repeated, it can proceed to microevolution that changes population gene pool at the species level. In the case of sexual reproduction, compared to female, male germ cells have more opportunities to mutate due to spermatogenesis during whole lifetime. In the present study, we perform whole genome sequencing (WGS) to analysis of somatic/germline mutations accumulated within individuals and patterns of mutations transmitted to offspring by parents’ age in crab-eating monkeys (Macaca fascicularis). We selected 4 families (total 16 monkeys; 10 males, 6 females) around various age to get data at all age point and considered that they can produce 3rd generation for detecting transmission to next generation. Also, we considered families with males as 2nd generation offspring in order to directly confirm the effect of transmitted mutations on the reproductive cells of the child. As a result of 3 years process, WGS data of monkey blood was built up, and based on that, diverse pipelines were devised to analyze mutations accumulated at each sampling time. From these analyses, we demonstrate the parental effect of de novo mutations accumulating with age in crab-eating monkeys throughout whole life on next generation.
Adaptive Genomic Signatures and Evolutionary Mechanisms in Anguillid Eels

Hyeongwoo Choi, Jiwon Nam, Seong-il Eyun
Chonnam National University (Korea, Chung-Ang University (Korea, Republic of))

Anguillid eels undergo a catadromous life cycle, transitioning from deep-sea larvae to adults in estuaries and freshwater environments. Their unique life cycle makes them an ideal model for exploring evolutionary shifts between saltwater and freshwater habitats, as well as between deep-sea and shallow-water environments. Despite inhabiting diverse ecological niches for over 21 million years, the evolutionary mechanisms of anguillid eels remain ambiguous. Hence, we generated the chromosome-level genome sequence of the short-finned eel, Anguilla bicolor pacifica, aiming to identify positively selected genes (PSGs) and elucidating their functions. Using the branch model, we detected 82 genes exhibiting positive selection on the anguillid branch (p < 0.05). It is widely known that adult eels cease feeding and experience gut degeneration during their migrating back to the deep-sea. Significantly, we observed PSGs associated with pathways involved in autophagy, longevity regulating, and peroxisome, suggesting their role in energy production during nutrient scarcity. Also, eels that migrate to the oxygen-minimum layer for spawning experience relaxed evolutionary constraint on the AMPK- and HIF-1 signaling pathways among the 82 identified PSGs, potentially contributing their adaptation to hypoxic environments. Comparative genomic analysis revealed that four expanded gene families in the four eel species, predominantly enriched in visual pathways such as NTN1, SIXs, and MAP1B (p < 0.05). In conclusion, this study provides valuable insights into the adaptability mechanisms in anguillid eels, shedding light on how they manage to thrive in diverse environment.
Development of a bioinformatics Pipeline for the Analysis of Complete Genomic Sequencing Data of Mycobacterium tuberculosis

Ikuri Alvarez-Mayá

Ikuri Alvarez-Mayá, Manuel García-Ulloa, Jaime Martinez-Urtaza
Autonomous University of Barcelona (Spain), Center for Research and Assistance in Technology and Design of the State of Jalisco (CIATEJ) (Mexico), Department of Genetics and Microbiology, Faculty of Biosciences, Medical and Pharmaceutical Biotechnology

The rapidly growing area of sequencing technologies, and more specifically bacterial whole genome sequencing, could offer new applications in clinical microbiology, such as the identification of bacterial species, prediction of genetic susceptibility to antibiotics, and virulence simultaneously. We developed a bioinformatics tool for the analysis of data obtained by Complete Genomic Sequencing that allows us to know the phylogenomic behavior. Mycobacterium tuberculosis (MTB) raw Illumina short reads were collected from projects PRJEB30933, PRJNA824124, and PRJEB44165 on the website www.ebi.ac.uk/ena repository. Additionally, the scaffolds from project PRJNA751891 at www.ncbi.nlm.nih.gov. were also collected. Raw reads were trimmed and quality filtered with fastp v0.12.4 using default parameters and a minimum read length of 50 bp. Genome assemblies were performed with SPAdes v3.15.5 using the --careful --cov-cutoff automatic options. Validation of the assemblies was carried out with QUAST v5.0.2 and BUSCO v5.4.2 using the actinobacteria_class_odb10 single-copy ortholog database as a reference. Sequencing depth was calculated by mapping the filtered reads to their respective assemblies using Bowtie2 v2.2.9 and samtools v1.7. Assemblies were scanned for contaminants by blasting their scaffolds against the NCBI nr database. Variable sites from all assemblies were extracted and aligned using the M. tuberculosis H37Rv strain (GCF_000195955.2) Multidrug resistance profiles, lineages, and sublineages were assigned to strains in silico with tb-profiler v4.3.0. The bioinformatic pipeline developed in the study was able to adequately analyze and identify not only the phylogenetic behavior of M. tuberculosis strains circulating in Mexico but also mutations related to resistance to drug treatment.
Sex determination (SD) and differentiation processes in sea lampreys present a complex interplay of genetic and environmental factors. Recent studies have shown that genes on highly repetitive small gonad-restricted chromosomes (GRCs) exhibit sex-biased expression in animals undergoing sexual differentiation, suggesting specialized role in SD similar to sex chromosomes in other species. The objective of this research project is to analyse structural and genetic variation of GRCs and their evolutionary patterns in sea lampreys. We analyzed population-level data to compare genetic differentiation between anadromous and land-locked lamprey populations, focusing on both somatic and GRC genome parts. We investigated the mode of inheritance by examining sex-specific heterozygosity in somatic chromosomes versus GRCs and used long-read data to compare structural variations in the GRC between male and female genomes. Additionally, we compared gene evolution rates on GRCs and somatic chromosomes by calculating dn/ds ratios. Our findings showed minimal differentiation between anadromous and land-locked populations in the GRC region and ten-fold higher Fst in the somatic region across sexes. Interestingly, at the macroevolutionary level, GRC genes evolve faster than those on somatic chromosomes, as shown by higher dn/ds values. Males exhibited two-fold higher GRC heterozygosity compared to females, and we identified 2 Mb of sequences exclusive to one sex, suggesting significant sex-specific structural variation. Overall, our findings indicate distinct molecular and evolutionary properties of GRCs in sea lamprey, and imply the role of this genomic region in the male-specific sex determination and differentiation pathways.
Elucidating the Evolution of the Recombinational Landscape of Placental Mammals Using Comparative Genomics

Isabella Rose Childers

Isabella R Childers, Nicole M Foley, Kevin R Bredemeyer, Terje Raudsepp, William J Murphy
Texas A&M University (USA)

Genomes are composed of varying evolutionary histories, influenced by processes like gene flow and ILS. Emerging studies have shown that the regions of the genome with lower meiotic recombination rates possess less phylogenetic conflict and have a tendency to preserve species histories. However, we have a poor understanding of how recombination rates have evolved over tens of millions of years and how we can use this information to better select informative genomic regions for phylogenomics. Here, we performed a comparative genomic analysis of recombination maps from divergent placental mammals to identify genomic regions that have retained historically low and high recombination rates. Low-rate regions should be useful for resolving species relationships, while high-rate regions may be useful for adaptive gene studies. To track the evolution of recombination rates, we reconstructed the ancestral placental mammal karyotype from mammalian genomes previously identified as having slowly evolving karyotypes and representing the four superordinal clades. We combined the ancestral chromosomes and recombination maps from extant species to identify syntenic regions where high and low recombination rates in the ancestor have persisted for 90-100 million years of independent evolution and characterized phylogenetic signal and functional attributes from these regions. These findings will enhance our understanding of the chromosomal architecture of speciation and natural selection within mammals.
Insights into the genetic bases of differential organ size in cacti flowers

Isaura Rosas Reinhold

Isaura Rosas-Reinhold, Gustavo Rodríguez-Alonso, Cristian Cervantes, Genaro Ramírez-Castro, Ulises Rosas, Alma Piñeyro-Nelson, Salvador Arias

Center for Genomics and Systems Biology, Centro de Investigación en Dinámica Celular. Instituto de Investigación en Ciencias Básicas y Aplicadas. Universidad Autónoma del Estado de Morelos (Mexico), Centro de Investigación en Dinámica Celular. Instituto de Investigación en Ciencias Básicas y Aplicadas. Universidad Autónoma del Estado de Morelos. (Mexico), Departamento de Producción Agrícola y Animal, Instituto de Biología, Jardín botánico, Universidad Autónoma Metropolitana-Xochimilco (Mexico), Universidad Nacional Autónoma de Mexico (Mexico), new york university (USA)

Floral morphological variation is a main driver of angiosperm evolution, changes in perianth pigmentation, floral organ number and size have been related to diversification and speciation. Developmental genetic studies in model species unveiled genes that directly control organ size. The roles of homologs of these genes in angiosperms with complex ontogenies such as the flower-shoot of Cactaceae have not been analyzed. Here we present a comparative transcriptomic analysis across flower buds of different growth stages from two cacti species with contrasting flower size: Disocactus speciosus and D. eichlamii. Differential gene expression analyses show that genes associated with flower size have different expression patterns in these species. Furthermore, a Weighted Gene Co-expression Network Analysis showed differential enrichment in processes related to organ growth between the initial size and the final size. Cell measurements and scanning electron microscope photographs of tepals of growing flower buds suggest that different cellular dynamics underlie final perianth size and support the transcriptome results. Our data suggest that differential regulation of specific genes could play a major role in defining flower size. In particular, BIG BROTHER and BIG PETAL, which are negative regulators of cell proliferation and cell expansion, respectively, were found to be differentially expressed in D. eichlamii and D. speciosus. These results suggest that these two genes could play an important role in flower size diversification in these species.
A complex landscape of introgressions in a diverse lineage of neotropical yeasts

J Abraham Avelar-Rivas


3Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Departamento de Ingeniería Genética, Irapuato, Juriquilla, Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), Mexico. (Mexico), Unidad Irapuato, Unidad de Genómica Avanzada (Langebio), Universidad Nacional Autónoma de México, na (Mexico)

Saccharomyces cerevisiae from the tropical Americas has scarce representation in genome-wide population studies of the species, which includes wild, domesticated, and feral yeasts from diverse habitats. To gain insights into the evolution of these yeasts living at the intersection of fermentative and wild contexts, we sequenced the genomes of over 200 strains isolated from open traditional agave fermentation across Mexico. Analysis of genetic variation in the context of isolates from all around the globe revealed a specific neotropical clade with strains isolated from Brazil to Northern Mexico. These genomic sequences formed a highly divergent branch with widespread introgressions from the sister species S. paradoxus. Phylogenomic analyses identified structured subpopulations within the clade, which correlated with geography and population parameters. Intriguingly, each subpopulation showed a specific genomic distribution of introgressions and the origins of such components were from different lineages of S. paradoxus. This suggests a still changing introgression landscape within the lineage, arguably reflecting ongoing hybridization, loss of heterozygosity, and backcrossing. Our findings underscore the contribution of population structure and introgressions to yeast diversification in a region characterized by its high biodiversity.
Life at extreme elevations: genomic and physiological insights into mechanisms of hypoxia adaptation in Andean mice

Jay F Storz, Schuyler Liphardt, Naim M Bautista, Nathanael D Herrera, Marcial Quiroga-Carmona, Juan C Opazo, Timothy B Wheeler, Zachary A Cheviron, Guillermo D'Elía, Jeffrey M Good

Integrative Biology Group (Chile), Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD) (Chile), Universidad Austral de Chile (Chile), Universidad de San Sebastián (Chile), University of Montana (USA), University of Nebraska (USA)

We report results of integrated genomic and physiological analyses to identify mechanisms of hypoxia adaptation in Andean leaf-eared mice (Phyllotis vaccarum) that live at extreme elevations (>6700 m [>22,000']) that were previously considered to be completely uninhabitable by mammals. Our high-altitude surveys in the Central Andes have yielded numerous specimens of P. vaccarum from extraordinarily high elevations, including the summits of multiple >6000 m (>20,000') volcanoes, far surpassing all previous specimen-based records for mammals. Population genomic analysis of specimens collected from sites spanning >6700 m of elevation – from the Pacific Coast of northern Chile to the crest of the Andean Cordillera – revealed relatively low levels of genome-wide differentiation. The analysis also revealed evidence for altitude-related selection on specific genes and pathways that we then targeted for functional experiments on captive mice reared in common-garden conditions. Our results provide insights into the genetic and physiological mechanisms that enable P. vaccarum to survive at the environmental limits of vertebrate life.
Long-read sequencing simplifies genome assembly and enhances the discovery of structural variants (SVs). Employing Drosophila melanogaster as our model, we investigated SV evolution at a population level using over sixty whole-genome long-read sequencing datasets. Comparative analysis of contemporary SV-calling methods yielded population statistics, unfolded site frequency spectra and selection signals. Traditionally, single nucleotide polymorphisms (SNPs) have been pivotal in inferring phylogeny, population evolution, and disease causation, owing to their detectability with cost-effective methods. Unlike SNPs, SVs, which often span and impact multiple genes, are more likely to influence an organism's phenotype. Despite their significance, SVs have eluded scrutiny due to their complex nature and the limitations of short reads from next-generation sequencing (NGS). This study bridges the gap by leveraging state-of-the-art long-read sequencing to unveil SVs in Drosophila melanogaster populations worldwide. Utilizing Pacific Biosciences and Oxford Nanopore Technology sequences from NCBI SRA, we assessed two major SV detection approaches. Mapping-based programs, though computationally efficient, exhibited lower sensitivity, while assembly-based methods, dependent on well-constructed genomes, captured larger-scale changes. Our validated pipelines are open source on GitHub. Diversity and selection signals varied among SV types, and we sought SVs contributing to population differentiation and chromosomal evolution. SVs demonstrated signals of negative selection while a small fraction was probably positively selected. Additionally, we constructed the largest Drosophila pangenome for comprehensive representation of polymorphisms.
Human ZEB2 orchestrates a network of genes with stronger contribution to neuronal development than non-human primate ZEB2

Jeong-Eun Lee, Vladimir M Jovanovic, Amanda Jager Fonseca, Sebastian Streblow, Stefano Berto, Katja Nowick
Bioinformatics Solution Center; Freie Universität Berlin (Germany), Chemistry and Pharmacy; Freie Universität Berlin (Germany), Human Biology and Primate Evolution Group; Institute for Zoology; Department of Biology, Medical University of South Carolina (USA)

Humans differ from other primates in many phenotypes, despite their nearly identical protein sequences. Therefore, investigating human-specific changes at the level of transcriptional regulation becomes imperative for understanding species-specific traits and adaptations. In this study, we explore the functional evolution of ZEB2, a key transcription factor involved in neural development and other functions, across primates. Using ChIP-seq in lymphoblastoid cell lines (LCLs) from three great ape species, we demonstrate ZEB2’s conserved binding near transcriptional start sites but also species-specific variation in target site choice and binding affinity. Human-specifically altered genes followed by ZEB2 knockdown in LCL are involved in neural development, as well as synapse and neuron projection organization, pointing to a human-specific functional change related to brain developmental processes. Consequently, we analyzed single cell RNA-seq and bulk RNA-seq data of developing great ape brain organoids. We revealed species differences in ZEB2 co-expressed genes, which might underlie evolutionary alterations in ZEB2 coordinated developmental pathways related to brain size and cognitive abilities. Our study offers valuable candidates for future studies and furthers our understanding of the role of ZEB2 in the evolution of the human brain.
Patterns of miRNA presence and absence in mammals have implications for placental phenotypes

Jonathan Fenn

Jonathan Fenn, Nkemdilim Nwaobi, Vladimir Ovchinnikov, Haidee Tinning, Zenab Butt, Olga Amelkina, Niamh Forde, Mary O'Connell
Queen Mary University (United Kingdom), University of Leeds (United Kingdom), University of Nottingham (United Kingdom)

microRNAs (miRNAs) are small regulatory RNAs which inhibit translation of mRNA - but how does the presence or absence of different miRNAs in the genome affect phenotype? The evolution of placental implantation in Eutheria was accompanied by massive diversification in miRNAs, and there are miRNA families that are entirely unique to – and conserved across- mammals. We revise previous estimates of Eutheria-synapomorphic miRNA families, using a dataset of 300 mammal genomes, homology searches and machine learning-based miRNA identification methods. We identify 10 core toolkit miRNA families whose functions are likely essential for successful pregnancy in Eutheria, and 314 miRNA families unique to mammals but with phylogenetically heterogeneous distributions. We hypothesised that the diversity of placental phenotypes could be underpinned by miRNA family presence/absence across the mammal phylogeny. We used a combination of phylogeny-aware phenotype diversity metrics (Evolink) and machine-learning approaches (Random Forest) using our purpose built placental phenotype dataset and our miRNA family presence/absence matrix. We identify 34 miRNA families strongly associated with placental phenotype. Many of these miRNAs are associated with aspects of mammalian reproduction and development. Amongst the strongest associations includes Mir-11968, present in all species examined within Ruminantia (distinctive in its cotyledonary placental shape), but completely absent outside this clade. We present our findings for the functional analysis of Mir-11968 here. In summary, our findings provide insight into the evolutionary relevance of miRNAs associated with placental phenotype, and provide targets for future work that are relevant to developmental pathologies.
When branching and evolution are tightly coupled

Jordan Douglas

Department of Physics, University of Auckland (New Zealand)

Standard temporal phylogenetic methods make the assumption that lineages evolve independently, and their evolution is unaffected by speciation events. However, this assumption is often unmet. There are many known cases when branching drives the rate of evolution, or when evolution drives the rate of branching. At the molecular level, gene duplication events enable the emergence of novel function(s) (neofunctionalisation or subfunctionalisation). At the organism level, speciation is often accompanied by rapid evolutionary changes (punctuated equilibrium). And, at the cultural level, languages often undergo rapid changes when one population breaks off from another (schismogenesis), perhaps as a means to establish cultural identity. However, these granularities of biological and cultural evolution are generally studied in distinct fields of academia. This presentation will describe a unified Bayesian phylogenetic method that accounts for these processes. This model combines phylogenetic (strict and relaxed) clock models with the birth-death process. Unobserved extinction events, or “stubs”, are accounted for.
Evolutionary diversification of dimeric transcription factors by evolving DNA binding specificity
José Fabricio López Hernández

Jose Fabricio López Hernández, Brian P Metzger, Joseph W Thornton
Deparment of Biological Sciences (USA), Purdue University, University of Chicago (USA), University of Chicago. Deparment of Ecology and Evolution (USA), University of Chicago. Deparment of Human Genetics (USA)

Variation in gene regulation contributes to the phenotypic diversity present in extant species. To regulate distinct gene targets, transcription factors (TFs) bind to specific DNA sequences. Estrogen receptors (ERs) and ketosteroid receptors (kSRs), critical for sexual development and physiology, regulate transcription by forming homodimers. Each subunit binds to one of two response elements (RE) using their DNA binding domain (DBDs). These receptors shared a common ancestor (AncSR1) before a gene duplication event. While ERs retained the ancestral profile binding to DNA, the ancestor of kSRs (AncSR2) recognized a different RE by acquiring substitution in the DNA recognition helix, the homodimer interface, and the rest of the DBD. Using an evolutionary approach that measures the effect of each substitution and their interactions will help us identify what changes were important during their functional transition. To answer this question, we are conducting a comprehensive study to quantify the effect of each possible substitution and their combinations, between AncSR1 and AncSR2. To understand the genetic architecture of the functional transition, we are using deep mutational scanning (DMS), to infer the effect of all combinations of substitutions between AncSR1 and AncSR2. We assay the function of mutants using cell sorting, by GFP expression, coupled with high-throughput sequencing of each sorted bin, to measure the ability of DBD mutants to induce transcription GFP activation using distinct REs. We will infer the effect of each substitution and their interactions to model the possible evolutionary paths during the functional transition of AncSR1 to AncSR2.
Challenging the Gram-Positive/Gram-Negative Dichotomy: Discovery of Gram-Negative Monoderm Bacteria

José Norberto García Miranda

José Norberto García Miranda, Luis José Delaye Arredondo, Gabriela Olmedo Álvarez
Centro de Investigación y de Estudios Avanzados de Instituto Politécnico Nacional Unidad Irapuato (Mexico)

The classification of bacteria into Gram-positive and Gram-negative groups has long been considered a fundamental divide in microbiology, reflecting differences in cell envelope structure. This dichotomy is closely tied to the presence of one or two cell membranes, leading to the classification of bacteria as monoderms or diderms, respectively. Historically, all diderm bacteria have been characterized as Gram-negative, while monoderm bacteria were presumed to be Gram-positive. Here, we present groundbreaking findings that challenge this paradigm. We have identified and characterized a previously unrecognized group of Gram-negative monoderm bacteria through comprehensive genomic and phenotypic analyses. What is more interesting is that these gram-negative bacteria occur as three intercalated clades within the Bacillacea clade, where the well-known Bacillus subtilis and Bacillus cereus clades occur. Our findings also have revealed that, Teichoic Acid (TA), once deemed a hallmark feature of Gram-positive cell walls, is absent in many Gram-positive bacteria and present in the "gram negative" bacillacea. TA is thought to be indispensable for maintaining cell structure, as well as playing roles in cell division, cell wall synthesis, and interactions with the environment. This work not only contributes to the ongoing discourse on bacterial evolution but also lays the foundation for future investigations into the ecological significance, evolutionary origins, and physiological adaptations of these enigmatic Gram-negative monoderm bacteria. By elucidating the existence of Gram-negative monoderm bacteria, we provide a novel perspective on bacterial cell envelope diversity.
Evolutionary and functional characterization of a new TRPV1 splicing variant originated in the ancestor of catarrhine primates

Juan C Opazo

Sofia Cacino, Deny Cabezas-Bratesco, Julia York, Kattina Zavala, Michael Vandewege, Sebastian Brauchi, Juan C. Opazo
North Caroline State University (USA), Universidad Austral de Chile (Chile), Universidad San Sebastian (Chile), University of Illinois (USA)

The evolutionary origin of new exons provides an opportunity for a more diverse repertoire of protein isoforms from a single gene. This process is a fundamental mechanism in evolution, increasing the functional complexity of genomes. The TRPV1 (Transient Receptor Potential Vanilloid 1) channel is a non-selective cation channel, a member of the TRPV gene family. It receives its name from being the receptor for capsaicin. It is also activated by temperature (>43°C), protons, and endogenous ligands. Physiologically, TRPV1 plays a critical role as a nociceptor, part of the mammalian pain pathways. Most studies have been conducted for a canonical splice variant of 839 amino acids; much less is known about the other variants. In this work, we studied the evolution of TRPV1 gene structure in primates. Our results show the existence of an exon (encoding for 11 amino acids) in the sixth position, which is only found in catarrhine primates. The splice variant incorporating the sixth exon is preferentially located in the endoplasmic reticulum; however, it is localized in the plasma membrane when co-expressed with the canonical variant. From an electrophysiological perspective, it presents lower currents than the canonical isoform, and it is not activated in the presence of capsaicin. Cells that co-expressed both variants showed currents similar to the canonical variant, although the activation in the presence of capsaicin was lower. Thus, our results suggest that the splice variant incorporating the sixth exon would be a negative dominant of the canonical isoform.
Genome evolution in parasitic plants; Insights from the first chromosome assemblies in Cuscuta

Juan Daniel Cerda

Juan Daniel Cerda, Claude dePamphilis, Eric Wafula, Michael Axtell, Jim Westwood, Paula Ralph
The Children's Hospital of Philadelphia (USA), The Pennsylvania State University (USA), Virginia Tech (USA)

The parasitic plant Cuscuta play both positive (diversity drivers) and negative (pests) roles in the ecosystem and in agriculture and is a model for mobile RNA (miRNA & microRNA) between host and parasite. Current studies on Cuscuta genome discovered large scale gene loss as well as numerous genes associated to the haustoria (feeding structure), but no study has done an extensive exploration of gene evolution in Cuscuta, much less genome evolution. In this study, we will attempt to answer the questions: How did parasitism evolved in Cuscuta? And how does a parasitic lifestyle affect the genome evolution of Cuscuta gronovii? To answer these questions, we created a chromosome level genome assembly for Cuscuta gronovii and Cuscuta campestris. We incorporated the genomes of Cuscuta australis, five Ipomoea species and Solanum lycopersicum with a comparative genomics approach to discern gene and genome evolution. Chromosome level assembly showed chromosome number close to expected for Cuscuta campestris (n = 30 vs n = 31), but half the previously established numbers for Cuscuta gronovii (n = 30 vs n = 15). We found that 46% and 81% of the genomes of Cuscuta campestris and Cuscuta gronovii respectively are comprised of repetitive elements. We deduced that the extreme genome size in Cuscuta gronovii is caused by a recent expansion of LTR repeats. This study has shown that Cuscuta genome plasticity is a complex issue worth revisiting in depth.
Enhanced DNA Damage Response in a Clade of Long-Lived Bats Resolved Using Chromosome-Length Genome Assemblies

Juan Manuel Vazquez

Juan M Vazquez, Mary E Lauterbur, Peter J Sudmant, David Enard
Berkeley (USA), University of Arizona (USA), University of California

Lifespan is one of the most variable traits across the entire tree of life, and especially in mammals. Differences in lifespans between closely-related species provides a promising avenue for discovering novel pro-longevity pathways using evolutionary techniques. Previous studies focused on the evolution of longevity-associated traits, such as DNA damage response, have been hampered by a combination of low-quality genomes, low-phylogenetic coverage, or long evolutionary times, all of which can negatively affect their power to detect genes associated with longevity. In order to comprehensively study the evolution of aging and aging-associated traits in bats, we generated chromosome-level reference genomes and primary cell line libraries from a 10-million-year-old clade of 9 California Myotis species spanning a 3-fold range of lifespans. Increases and decreases in longevity independent of body size have evolved multiple times in this clade, providing a dynamic range which can be studied through functional genomics. Leveraging both genomes and cell lines, we identify genetic variants in several pathways specifically associated with longevity - including a polymorphic TP53 whole-locus duplication. We additionally find adaptations in other longevity-associated traits such as DNA repair and immunity; and show that these changes are associated with cellular resistance to various forms of chemically-induced DNA damage. These pathways represent new targets for exploration using primary cell cultures, and contribute to our understanding of how both agonistic and antagonistic pleiotropy play a role in the evolution of longevity.
Immune modulation is crucial for invasive species' success and house sparrows (Passer domesticus) exemplify this. Colonization likely relies on inflammation control, with Toll-like receptors (TLRs) playing a key role in pathogen recognition. We found elevated TLR expression (TLR2 and TLR4) in invading sparrows; however, excessive expression of TLRs may lead to an overly robust inflammatory response. We hypothesize that non-native house sparrows will exhibit an elevated baseline TLR expression, while concurrently relying on the co-expression of anti-inflammatory cytokines (IL-10) to attenuate downstream inflammatory pathways and non-native house sparrows will hasten the resolution phase of their inflammatory response. To examine the differences in immune modulation, we captured house sparrows and measured TLR2, TLR4, IFN
Illuminating the Abyss: Cultivation Strategies for Uncharted Cave Dwarf Bacteria
Karen Sevilla Landaverde

Karen Sevilla Landaverde, Corina Diana Ceapa, Arturo Jiménez Sanchez, Colin Bonduelle, Daniel Ríos Barrera
Biomedical Research Institute (Mexico), Institute of Chemistry (Mexico), University of Bordeaux (France)

Microbial Dark Matter is a term used to describe the vast majority of microorganisms on Earth that are yet-to-be-cultured, representing over 99% of the estimated microbial diversity. While these microorganisms are abundant and diverse, particularly in soils, there is a significant proportion of them that remain unexplored, especially in unique environments such as caves. Among this unexplored microbial diversity, dwarf bacteria, which have cell sizes smaller than 0.6µm, are particularly challenging to grow due to their nutrient-hoarding ability and limited microenvironment access. To address this challenge, cultivation methods have been proposed to simulate the conditions found in caves, which have yielded two populations of cultivated bacteria: larger and facultative small-cell size bacteria, and dwarf bacteria. While the former have been successfully isolated, the cultivation rate and morphotype diversity for the latter appear to be reduced. To accurately estimate the cell size of the bacteria, four new fluorescent molecules have been tested and compared against DAPI in this study, resulting in improved size estimation. Using this methodology, several cell populations of dwarf bacteria that grow in caves were isolated and characterized. This research highlights the significance of exploring unexplored environments to expand our understanding of microbial diversity, which could lead to the discovery of novel microbial species. The proposed cultivation method could also be useful for studying other environments, such as extreme habitats and even well-known ones. The findings of this study emphasize the importance of developing novel cultivation methods to study the vast majority of unexplored microbes.
Dissecting an ancient stress resistance trait syndrome in the compost yeast Kluyveromyces marxianus

Kaylee E Christensen

Kaylee E Christensen, Abel Duarte, Zhenzhen Ma, Judith L Edwards, Rachel B Brem
Berkeley (USA), Stanford University (USA), UC Berkeley (USA), University of California

In the search to understand how evolution builds new traits, ancient events are often the hardest to dissect. Species-unique traits pose a particular challenge for geneticists—cases in which a character arose long ago and, in the modern day, is conserved within a species, distinguishing it from reproductively isolated relatives. In this work, we have developed the budding yeast genus Kluyveromyces as a model for mechanistic dissection of trait variation across species boundaries. Phenotypic profiling revealed robust heat and chemical-stress tolerance phenotypes that distinguished the compost yeast K. marxianus from the rest of the clade. We used culture-based, transcriptomic, and genetic approaches to characterize the metabolic requirements of the K. marxianus trait syndrome. We then generated a population-genomic resource for K. marxianus and harnessed it in molecular-evolution analyses, which found hundreds of housekeeping genes with evidence for adaptive protein variation unique to this species. Our data support a model in which, in the distant past, K. marxianus underwent a vastly complex remodeling of its proteome to achieve stress resistance. Such a polygenic architecture, involving nucleotide-level allelic variation on a massive scale, is consistent with theoretical models of the mechanisms of long-term adaptation, and suggests principles of broad relevance for interspecies trait genetics.
Molecular phylogenetic tree of a group of species with distant genetic distance using Orthopteran insects
Kazuhiro Satomura

Kazuhiro Satomura, Atsushi Ogura
Nagahama Institute of Bio-Science and Technology (Japan)

Molecular phylogenetic trees are important tools for inferring phylogenetic relationships within and among species, taking into account their evolutionary history. Methods for estimating molecular phylogenetic trees have been established over the past 50 years of research. When inferring phylogenetic relationships, problems such as heterotachy and long branch attraction are known to arise due to homologous sequences being under different natural selection within a single data set or due to changing mutation bias during evolution. In addition, when estimating the age of species divergence from molecular phylogenetic trees, there is the problem of different clock rates for recent and older divergences, even when calibrated using the archaeological record. Orthoptera insects, the subject of this study, are more than 200 million years old. When constructing a molecular phylogenetic tree of this group of species, even using whole genome sequences, homologous genes common to all species cannot be found. Even when a common ortholog is obtained, it is not always derived from a common ancestor over the entire length of the sequence. Even if homologous sequences derived from a common ancestor can be defined, invariant sites and multiple substituted sites cannot be used to infer the tree. Even if we could define a one-time substitution site, we could not determine whether the mutation is fixed in the population without population genomic data. We investigated the methods and problems of molecular phylogenetic inference for a group of species with a long evolutionary history using DNA sequences of Orthoptera species.
Unveiling the Functional Dynamics of Noncoding DNA: A Deep Learning Approach with Arabidopsis thaliana

Kristy S Mualim, Moises Exposito-Alonso
Berkeley (USA), Carnegie Institute of Science (USA), Howard Hughes Medical Institute (USA), Stanford University (USA), University of California

Despite most Genome-Wide Association Studies (GWAS) variants falling within the noncoding regions of the genome, the functional implications of these variants and their contribution to fitness effect within accessible chromatin regions is still poorly understood. These challenges are compounded by difficulties in experimental perturbation - most cis-regulatory elements have small effect sizes that are hard to detect and most studies lack power, sample size or scale. By leveraging deep learning techniques, we describe a deep-learning framework to decode the intricate molecular relationships of how natural sequence variation within Arabidopsis thaliana accessions impacts regulatory function. This framework aims to open new avenues for understanding evolutionary dynamics of noncoding DNA.
Large genomic rearrangements, such as chromosomal inversions, often underlie karyotype variation, but the mechanisms by which these rearrangements arise remain poorly understood. To study the origins of inversions, we generated chromosome-level de novo genome assemblies for four subspecies of deer mice (Peromyscus maniculatus) with known inversion polymorphisms and identified ~8,000 inversions, including 47 mega-base scale inversions. Analysis of inversion breakpoints suggests that while most small (<1 Mb) inversions arise via ectopic recombination between retrotransposons, large (>1 Mb) inversions are primarily associated with segmental duplications (SDs). Large inversion breakpoints frequently occur near centromeres, which may be explained by an accumulation of transposable elements in pericentromeric regions driving SD formation. Additionally, multiple massive inversions likely arose from ectopic recombination between near-identical centromeric satellite arrays located megabases apart, a previously uncharacterized mechanism of inversion formation. Together, our results illuminate how repeats give rise to chromosomal rearrangements, shaping the propensity for massive shifts in chromosome architecture.
Seven new reference-quality genomes illuminate the genomic evolution of genetic sex determination in turtles

Leon Hilgers

Leon Hilgers, Michail Rovatsos, Thomas Brown, Martin Pippel, Chetan Munegowda, Peter Praschag, Sylke Winkler, Joana Damas, Harris Lewin, Eugene Myers, Uwe Fritz, Michael Hiller

Charles University (Czech Republic), Goethe University (Germany), LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG) (Germany), Leibniz Institute for Zoo and Wildlife Research (Germany), Max Planck Institute of Molecular Cell Biology and Genetics (Germany), Senckenberg Research Institute (Germany), Turtle Island (Austria), University of California (USA), Uppsala University (Sweden)

Turtles are ideal to study sex chromosome evolution. They repeatedly evolved genetic sex determination (GSD), with male (XY) and female (ZW) heterogamety and micro- as well as macro-sex chromosomes. Additionally, karyotype-based studies indicate that GSD evolution drastically changed genome architecture increasing chromosome numbers in GSD turtles. However, turtle sex chromosome origins, their gene contents and local as well as genome-wide consequences of GSD evolution remain unknown. Here, we generated seven reference-quality turtle genomes, re-sequenced males and females of GSD turtles, and use a haplotype-resolved assembly including X and Y chromosomes to locate an elusive sex locus. Next, we identified gene losses as well as genes under positive selection and reconstruct chromosome evolution across all turtles. This allows us to explore genome evolution linked to GSD and trace the evolution of micro- and macro-sex chromosomes. In contrast to previous hypotheses, GSD does not coevolve with chromosome numbers in turtles. Instead, correlation of GSD and chromosome numbers is driven by dynamic chromosome evolution with numerous fusions in a sister clade with temperature dependent sex determination. Our data show that chelid turtle sex chromosomes originated once from a micro-chromosome >80 mya, but remained highly homomorphic. Macro-sex chromosomes evolved by subsequent chromosome fusion. In combination, this study illuminates sex chromosome evolution in turtles, adds to mounting evidence that not all sex chromosomes are “born to die” and paves the way to establish turtles as a model to study sex chromosome evolution.
Pseudogenization-driven gene loss shapes genome evolution in Hanseniaspora

Luis Fernando García-Ortega

Luis F. García-Ortega, Angela María García-Acero, Alexander DeLuna, Lucía Morales, Eugenio Mancera, Luis Delaye
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Mexico), Departamento de Ingeniería Genética, Departamento de Ingeniería Química y Ambiental, Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), Unidad Irapuato, Unidad de Genómica Avanzada (Langebio), Universidad Nacional Autónoma de México (Mexico), Universidad Nacional de Colombia (Colombia)

Gene family contractions and gene loss have played significant roles in the evolution of fungal genomes. Among the mechanisms driving gene loss, pseudogenization stands out as a major contributor; high mutation rates during this process lead to rapid genome reduction. Fungi from the Hanseniaspora genus are a great model to understand gene loss since some spices are characterized by the smallest genomes among the budding yeasts. Furthermore, Hanseniaspora can be separated into two lineages, a slow evolution lineage and a fast evolution lineage with reduced genomes. In this study, we isolated and sequenced 16 Hanseniaspora isolates from different geographic regions across Mexico. Among them, ten were identified as H. lachancei, two as H. pseudoguilliermondii, two as H. guilliermondii, and one as H. opuntiae. Comparative analysis of these genomes revealed the dynamic nature of gene family sizes and compositions, with metabolic, DNA repair, and cell cycle control functions being the most affected. Interestingly, the distribution of pseudogenes does not correlate with genome size or gene count, but instead shows lineage specificity, indicating that distinct remodeling processes operate during the evolution of each organism\'s genome. Overall, we observed a wide range of pseudogene integrity, suggesting that each lineage may exhibit varying degrees of partial activity. This hypothesis will be further tested by phenotyping the isolates under a variety of growing conditions. Our findings highlight the importance of pseudogenization-driven gene loss as a mechanism that shapes genome evolution in autochthonous Hanseniaspora populations, providing insights into their adaptation strategies.
Phylogenetic modeling of gene expression shifts in the mole-rat clade

Maëlle Daunesse, Élise Parey, Diego Villar, Camille Berthelot
EMBL-EBI (United Kingdom), IBENS (France), Institut Pasteur (France), Queen Mary University of London (United Kingdom), University College London (United Kingdom)

Understanding the genetic basis of phenotypic adaptations poses a significant challenge in evolutionary genomics. Despite the morphological and physiological diversity in mammalian traits, their coding genomes exhibit a high degree of conservation, implying that changes in gene expression and regulation are pivotal in driving phenotype evolution. This study aims to identify and establish a connection between shifts in gene expression and cis-regulatory elements, and their potential impact on phenotypic adaptations. Using African mole-rats as a model, renowned for their unique phenotypic adaptation traits like cancer resistance and hypoxia tolerance, we aimed to elucidate the genome-wide gene expression patterns underlying these traits that have been mainly characterised at the level of candidate genes and in individual species. Profiling gene expression in heart and liver tissues across two mole-rat species and two rodent outgroups, we used a phylogenetic comparative approach to identify genes with expression shifts within the mole-rat clade and in specific genera. These shifted genes are associated with functions pertinent to known adaptations in naked mole-rats, such as myogenesis and glycolysis in the heart. Furthermore, our analysis revealed concordant changes in the regulatory landscape of these genes. By employing a phylogenetic comparative approach, we offer new insights into the interplay between gene expression, regulation, and phenotypic evolution in mammals. Our findings shed light on the molecular mechanisms driving the evolution of unique traits in mole-rats and potentially other mammalian species.
A reformulation of measures of linkage disequilibrium and population structure under drift, inbreeding, and tight linkage: An approach through probabilities of identity by state

Marcy Uyenoyama

Marcy Uyenoyama
Duke University (USA)

Observations of genome- and planet-wide patterns in linkage disequilibrium (LD) and measures of population structure (especially FST) have served as the basis for inferring demographic history. Most definitions of FST that involve probabilities of identity-by-descent (IBD) or variance components take an approach through pedigrees, with reference to an ancestral base population. The dependence of classical measures of FST on allele frequencies can cloud their interpretation as indicators of population structure. A reformulation of Nei’s GST in terms of identity-by-state (IBS) probabilities clarifies the meaning of Wright’s F-statistics. In place of references to allele frequencies in the ancestral base population, this approach entails a locus-specific accounting of the mutation process itself. Similarly, accounting for both the existence of variation as well as its pattern can simplify the inference framework for the analysis of associations among distinct loci. Under this approach, a compact method for deriving the probabilities of the joint observation IBS from a pair of loci provides diversity-based measures of LD from unphased samples. Of particular interest are associations that arise among sites separated by just tens of kilobases, implying rates of crossing-over on the order of $10^{-5}$. Interactions among genetic drift, mutation, crossing-over, and regular inbreeding determine levels of within-locus genetic diversity and between-locus identity-by-state.
Molecular Footprints on Osmoregulation? Related Genes Associated with Freshwater Colonization by Cetaceans and Sirenians

Mariana F. Nery

Mariana Nery, Elisa Ramos
University of Campinas (Brazil)

This research investigates the genetic foundations of adaptive physiological mechanisms in mammals that transitioned to aquatic lifestyles, with a particular emphasis on species moving from marine to freshwater habitats. Such transitions pose osmotic challenges, necessitating adaptations in the osmoregulatory system. We focused on the selective pressures acting on coding and regulatory regions of 20 genes related to osmoregulation in cetaceans and sirenians, which include species in both marine and freshwater environments. Our findings highlight positive selection in the vasopressin (AVP) gene in lineages that have adapted to freshwater habitats and in aquaporins across both types of environments. Dolphins displayed a higher number of positively selected sites compared to the Amazonian manatee, with AQP5 and AVP genes showing selection signals across several independent lineages. Vasopressin gene sequences in river dolphins were surprisingly similar across their independent evolutionary paths. The study also examined Transcription Factors in the promoter regions of these genes, revealing phylogenetic conservation among sister species. Genes such as ACE, AQP1, AQP5, AQP7, AVP, NPP4, and NPR1, essential for osmotic regulation, showed signs of accelerated evolution in freshwater mammals. This research enriches our understanding of the genetic evolution underpinning osmoregulation in aquatic mammals, shedding light on the complex genetic adaptations to environmental changes.
Host sex can affect pathogen-induced immune responses. Evidence suggests that men have a higher risk of both colonization and infection with Staphylococcus aureus (MRSA), with an odds ratio of ~2.4 compared to women. The mechanisms underlying increased male predisposition to S. aureus infection remain unclear. To examine the genetic basis of this bias, we conducted a 2x2 factorial experiment and analyzed RNA-seq data in S. aureus-induced dermo-necrosis (SSTI) in mice. We examined differentially expressed (DE) genes between infected and control male and female mice. The RNA-seq read counts of each gene were modeled as a function of an effect of Sex (males vs. females), Condition (infected vs. control) and Sex-by-Condition interaction, while accounting for overdispersion using the negative binomial distribution employing DESeq2 R package. The Sex-by-Condition interaction unveiled 4034 DE genes indicating that the effect of infection on gene expression differs between males and females. Among these, 2486 were upregulated and 1548 were downregulated in infected males compared to infected females. These genes underwent enrichment analysis of GO and KEGG. Over-representation and GSEA analyses revealed 26 significantly enriched immune response GO biological processes. While the mechanisms behind these sex-specific DE mouse genes remain elusive, our analysis suggests that this bias stems from sex-specific genetic architecture, characterized by gene-sex interactions. Tailored approaches that consider sexual dimorphism - such as differences in susceptibility to infectious diseases, vaccination efficacy, and treatment responses - could significantly improve patient outcomes.
European amphioxus’ enormous genetic diversity explained by asymmetric migration and long-standing high effective population size

Marina Brasó-Vives

Marina Brasó-Vives, Diego A Hartasánchez, Stéphanie Bertrand, Hector Escriva, Marc Robinson-Rechavi
CNRS-Sorbonne University (France), Department of Computational Biology, Department of Ecology and Evolution, Oceanological Observatory of Banyuls-sur-Mer (OOB), Swiss Institute of Bioinformatics (SIB) (Switzerland), University of Lausanne (Switzerland)

Amphioxus are marine filter-feeding chordates that reproduce by external fertilization. The European amphioxus (Branchiostoma lanceolatum) has an ecological range extending at least from the northeastern Atlantic Ocean to the Mediterranean Sea. The amphioxus lineage has had a slow phenotypic evolution and a low substitution rate since their split with the other chordate lineages around 500 Mya. Curiously, previous estimates extracted from single individual genomes suggest amphioxus have one of the most diverse genomes ever measured. Here, we use genomic data from two European amphioxus wild populations (Mediterranean and Atlantic) to accurately quantify and understand their genomic diversity in the context of their geographic history. We confirm both the low substitution rate of the amphioxus lineage and the extremely high levels of genomic diversity in amphioxus wild populations. We confidently estimate that at least 3% of sites are heterozygous in each individual genome. Additionally, our results show two distinct populations of amphioxus in the Mediterranean and the Atlantic and suggest that the former one was originally founded by the latter probably during the refill of the Mediterranean Sea around 5.3 Mya. We also find evidence for maintained unidirectional gene flow from the Atlantic to the Mediterranean populations. These results point towards a long-standing high effective population size as responsible for the enormous genomic diversity of amphioxus. This study highlights the exceptionality of amphioxus’ evolution and contributes to the understanding of the relation between historical geographical range distribution and genomic diversity.
Characterising the detectable and invisible fractions of genomic loci under balancing selection

Max Reuter

Débora Brandt, Filip Ruzicka, Aida Andrés, Tim Connallon, Max Reuter

Institute of Science and Technology Austria (Austria), Monash University (Australia), University College London (United Kingdom)

Balancing selection refers to a suite of population genetic processes that push allele frequencies away from the boundaries and towards a polymorphic equilibrium state, thereby helping to maintain heritable variation in fitness and its components. Empirical studies have provided iconic examples of balanced polymorphisms, yet the general prevalence of balancing selection within genomes remains a topic of longstanding debate. Genome-wide scans for signals of balancing selection suggest that it is rare. However, current inference methods are notoriously conservative and are thought to identify only a subset of loci evolving under balancing selection. What is unclear is what proportion of loci under balancing selection is essentially invisible to detection and how our appreciation of the biological properties of balanced allelic variation (stability and equilibrium frequency, or selection and dominance coefficients) is biased by limitations in the power of inference. Here, we address these questions using a combination of analytical modelling and population genetic simulations. We begin by reviewing classic population genetics theory for the maintenance of polymorphism under three scenarios of balancing selection, overdominance, antagonistic pleiotropy and sexual antagonism. We then extend this theory to make predictions about the fraction and properties of loci under each balancing selection scenario that remain stably polymorphic for extended periods of time—and hence most visible to selection scans. Finally, we determine the portion of the parameter space in which current methods are able to detect balancing selection. We discuss the implications of our predictions for the interpretation of empirical data from genome scans.
Population studies of pathogens are vital for understanding the distribution of infectious diseases. When coupled with DNA biobanks, antibody response can be correlated with genetic and clinical factors, leading to potential public health interventions. Here, we analyze serology data for 41 antigens specific to 20 pathogens in 5,798 individuals from the Mexican Biobank. Samples were distributed nationwide across Mexico and were collected as part of the 2000 National Health Survey including socio-demographic data. Serology measurements were obtained using a multiplex panel based on Luminex beads technology previously utilized to measure a number of pathogens similar to that of the UK Biobank. We used antibody titers cutoffs for each antigen to calculate the prevalence of medically relevant pathogens, including HPV, EBV, HIV, Hepatitis B and C, Helicobacter pylori, and Toxoplasma gondii. We compared our results with the UK Biobank as well other sources. We found that the pathogens with the highest prevalence in the MX Biobank also have a high prevalence globally, except for H. pylori which exhibits a higher prevalence in Mexico. In a logistic regression model, no significant differences were found in seropositivity of pathogens between sexes. However, significant differences were found when comparing speakers vs. non-speakers of indigenous languages and comparing rural vs urban areas. This work represents a valuable resource for epidemiological studies in Mexico and future studies of genetic association with immune response to certain pathogens.
Investigating the "neighbor repulsion" pattern in Patterson's f-statistics
Mehmet Somel

Gözde Ata?, Mehmet Somel
METU (Turkey), MPI-EVA (Germany)

Here we study a surprising pattern arising in population genetics analyses involving Patterson's outgroup-f3 and f4 statistics, where groups sharing a high fraction of common ancestry -often geographically neighbors- show higher apparent affinity to a more "distant" group. Such a pattern has been described for North African modern-day groups showing higher affinity to Sardinians than to their neighbors. We now present a new case involving Bronze Age East Mediterranean populations, where Anatolian groups choose populations from Greece rather than their neighbors, in seeming contradiction with patterns observed in PCA and MDS analyses. We next perform coalescent simulations involving low-level gene flow independently into "sister" groups from an external population with high within-population diversity. This recapitulates the observed f-statistic patterns, where the sister groups show higher affinity to a more distant but unadmixed population. We also show that, in the same simulations, summaries of outgroup-f3 statistics using MDS can capture overall shared ancestry patterns, similar to our empirical observations. We note that this behavior of both outgroup-f3- and f4-statistics is theoretically expected as these are single estimates representing genetic correlations arising from mixed ancestry. Our results illuminate how low-level and heterogeneous admixture can produce unexpected population genetic statistics.
Epistasis from conformational ensembles shapes evolution

Michael J Harms

Michael J Harms, Brennan A Fitzgerald, Natalie M Jaeger
Department of Chemistry & Biochemistry, Institute for Molecular Biology, University of Oregon (USA)

Macromolecules exist as ensembles of interchanging conformations. Such dynamics are important for molecular function and regulation. We hope to convince you these ensembles are also profoundly important for protein evolution. Using theoretical and computational approaches, we found that subtle changes in ensemble composition can profoundly alter the effects of future mutations. Using experiments on the lac repressor and an RNA riboswitch, we revealed that these predicted effects are detectable both in vitro and in vivo. Using phylogenetic reconstructions and experimental studies of the innate immune protein S100A9, we revealed that molecular ensembles have indeed shaped the historical evolution of natural proteins. We are currently working on computational models to predict these effects for specific proteins, simulate protein evolution when taking conformational ensembles into account, and to define bioinformatic signatures of this phenomenon in multiple sequence alignments. Together, this work is revealing an intimate connection between molecular ensembles—an inescapable biophysical feature of macromolecules—and how macromolecules evolve. Specifically: subtle, functionally invisible, changes to the energies of low-population conformations can open and close future mutational trajectories, thus dramatically altering evolutionary outcomes.
Comparing common strategies for ortholog selection used in phylogenomics

Mingzhu Yang

Mingzhu Yang, Mattia Giacomelli, Maria Eleonora Rossi, Jesus Lozano Fernandez, Philip Donoghue, Davide Pisani
University of Barcelona (Spain), University of Bristol (United Kingdom)

Phylogenomics rely on superalignments including hundreds or thousands of orthologous gene families. Many pipelines exist to identify orthologs from sets of proteomes. A strategy that has grown in popularity is extracting single-copy genes pre-identified in the BUSCO database from the species of interest (Seppey et al., 2019). Another commonly used strategy is using OrthoFinder (Emms et al., 2019) to define de novo orthologous families. Other similar strategies can be designed using different approaches to define the orthogroups (e.g. using OrthoMCL). One frequently overlooked problem is the correspondence between single-gene families inferred using different methods. While the expectation is that if different pipelines perform comparably well, they should identify the similar sets of single-gene families, available evidence suggests the opposite (Li et al., 2023). Here we identified 954 single-gene families using BUSCO for a set of nine high-quality metazoan proteomes and three outgroups. For the same genomes, OrthoFinder identified 2937 single-gene families, 365 of which were present in all 12 taxa. We functionally annotated these families and looked at the pattern of correspondence between families inferred using BUSCO and OrthoFinder. We found that the BUSCO metazoan set is depleted in translational genes (KOG functional category J). Furthermore, there is no clear correspondence between orthogroups inferred using BUSCO and OrthoFinder. We conclude that the use of alternative software for ortholog detection can potentially lead to the inference of different phylogenies, and it is thus crucial to better understand which of these approaches is more reliable.
A proposal of the Ur-RNAome

Miryam Palacios-Pérez

Miryam Palacios-Pérez, Marco V. José
Network of Researchers on the Chemical Emergence of Life, NoRCEL (United Kingdom), UNAM (Mexico), Universidad Nacional Autónoma de México, Universidad Nacional Autónoma de México (Mexico)

We portray here a plausible collection of ancient RNAs following an RNY pattern (where R indicates purine, N means any of the four bases, and Y indicates pyrimidine), since Eigen and Schuster suggested in the early 1970s, based on logical deduction of the thermodynamic properties of such polyribonucleotides in the RNA world, that this was the first genetic code. We selected 26 species (13 Bacteria and 13 Archaea), and we began by extracting an RNY RNAome from a contemporary organism and the resulting smaller RNAome was used as a query in order to obtain a list of ancient RNAs encoded by RNY codons. We found that the free energy of RNY hairpins was consistently lower than their corresponding shuffled controls. We found traces of the 3 ribosomal RNA (16S, 23S and 5S), tRNAs, 6S RNA, and the RNA moiety of RNAseP and of Signal Recognition Particle (SRP). Yet, at his stage of evolution there was not a genetic code as there was the complete absence of Peptidyl Transferase Centre, and not even vestiges of anti-Shine-Dalgarno 16S RNA. Interestingly, we detected the anticodons of both glycine (GCC) and threonine (GGU) in their hairpins proto-tRNA. It is widely accepted that the earliest RNA molecules folded into hairpins or mini-helices. Herein, we depict the 2D and 3D conformation of those earliest RNA molecules with only RNY triplets (Ur-RNAome).
Exploring fitness landscapes in laboratory yeast crosses
Misha Gupta

Misha Gupta, Artur Rego-Costa, Evgeniia Belousova, Michael M Desai
Harvard University (USA), Regeneron (USA)

Fitness landscapes are a mathematical tool used to visualize genotype-phenotype maps and have important implications for studying evolution. Populations can be thought of as ‘navigating’ these landscapes during adaptive evolution, with evolutionary paths and outcomes constrained by the structure and ruggedness of a landscape. The complexity of these landscapes can also be influenced by interactions across the genome. Studying these fitness landscapes can help predict potential evolutionary trajectories and understand the constraints and possibilities within evolutionary processes. However, these landscapes are extremely high dimensional and often challenging to explore. Previous evolution experiments have observed unexpectedly flat fitness landscapes (1). Other quantitative genetic experiments have built a large yeast cross and tried to sample a local neighborhood of the landscape densely (2), showing similarly flat structures. These flat landscapes contradict intuition, especially given complex genetic backgrounds and environmental effects. One hypothesis suggests that this results from sampling small enough local neighborhoods that they appear effectively flat. Our study aims to explore different local neighborhoods of the fitness landscape by generating individuals close to the large yeast cross, yet differing in genetic relatedness. By exploring these disjoint yet nearby neighborhoods in the fitness landscape, we might find some local curvature and complexity in the landscape. On the other hand, flat landscapes may suggest that fitness in laboratory crosses is robust to genetic changes, and such populations may evolve through drift. (1): https://doi.org/10.1111/j.1558-5646.2011.01569.x (2): https://doi.org/10.7554/eLife.73983
Mitochondrial whistleblower: Unveiling a Hidden Third Partner in Sponge Symbiosis

Momin Ahmed

Momin Ahmed, Jan Vicente, Dennis V Lavrov
Hawai‘i Institute of Marine Biology, Iowa State University (USA), University of Hawai‘i at Manoa (USA)

Sponges (Phylum Porifera) form symbiotic relationships – ranging from mutualism to parasitism – with a variety of other organisms, including other sponges. Recently, Vicente et al. described a symbiotic association between two sponge species, a demosponge Haliclona plakophila and a homoscleromorph Plakortis symbiotica. Surprisingly, our genome skimming of the H. plakophila – P. symbiotica samples revealed the presence of, not two, but three complete sponge mitochondrial genomes (mitogenomes), all with high sequence coverage. Here we report phylogenetic analysis of mitochondrial as well as 18S and 28S sequences assembled from H. plakophila – P. symbiotica association aiming to resolve the phylogenetic position of the third “Mystery” sponge. Our findings indicate that mitochondrial sequences from this species were phylogenetically uninformative, likely due to their highly unusual GC-rich nucleotide composition. By contrast, 18S and 28S genes have placed the Mystery sponge within Haplosclerida, but not in any of the recognized clades within this group. We suggest that the Mystery sponge’s atypical nucleotide composition in its mitogenome reflects its highly unusual, potentially parasitic, lifestyle. This work opens new avenues for exploring the biology of this enigmatic sponge.
Error correction of viral genomic sequences using a phylogenetic prior

Monica Arniella

Monica Arniella, Lenore Pipes, Rasmus Nielsen
University of California Berkeley (USA)

Over the course of the SARS-CoV-2 pandemic, there has been a fast-paced and widespread effort to sequence clinical and wastewater samples around the world. Analysis of these sequences has quickly revealed potentially problematic sites with high error rates, necessitating careful filtering or masking of regions in the genome. Additionally, the error rate of genomic sequencing can be highly variable between experiments, and areas of low sequencing depth make base calling even more error-prone due to ambiguity in the identity of nucleotides. In the case of viral sequences, we can use additional information from phylogenies in order to infer the most probable nucleotide at a position. Here, we use the global SARS-CoV-2 phylogeny (involving millions of sequences from around the world) and the phylogenetic method tronko by Pipes et al. to calculate prior probabilities for each nucleotide along randomly chosen genomes, and use the observation of sequencing reads to choose the bases with the highest posterior probability. To test the performance of this method, we test varied sequencing conditions using real samples downloaded from the Sequencing Read Archive (SRA), simulated reads from genomes obtained from the GISAID repository, and compare the performance of our method with standard viral base calling methods. Using this approach, we evaluate the utility of using phylogenetic information for more confident base-calling at uncertain sites and increasing the accuracy of downstream analyses.
Disentangling the Ancestral History and Population Genetic Structure of Modern Iranian Populations

Nada Mohamed Ibrahim Elsharkawy

Nada M Elsharkawy, Mark Penjueli, Aashish Jha, Carlos D. Bustamante, Roxana Daneshjou, Alexander Ioannidis
Department of Biomedical Data Science, Department of Biomolecular Engineering, New York University Abu Dhabi (UAE), Stanford University (USA), The Genetic Heritage Group, University of California Santa Cruz (USA)

Iran is considered a crucial hub for human migration throughout history, yet the Iranian populations have not been adequately represented in genomic studies. This limitation has restricted our understanding of the contributions Iranian peoples have made to modern human genomic diversity and population histories. This study aims to bridge this gap by analyzing whole genome sequences (WGS) of 73 contemporary Iranians from 24 diverse ethnic populations, in the context of WGS data from 4,000 individuals from worldwide human populations. Our findings reveal distinct admixture patterns with populations from neighboring countries, showcasing unique genetic influences and admixture patterns reflective of the region’s historical interconnections, notably with Europe, Western Asia, South Asia, and Central Asia. However, these admixture patterns appear to vary depending on ethnic affiliation. By integrating new genomic data with existing datasets, we examine three critical aspects: gene flow into Iran, gene flow out of Iran, and the temporal dynamics of these processes, and whether these dynamics differ across the different ethnic affiliations within Iran. The research contributes to a broader understanding of the pivotal role Iran has played in human prehistoric migrations and lays the groundwork for future analyses to further understand the evolutionary pathways that have shaped the population structures in this pivotal geographical area.
Creating and using open educational resources to teach biology to improve accessibility, inclusion, and social justice

Nadia Aubin-Horth

Nadia Aubin-Horth, Ariane Barrette, Alicia Pageau, Jeanne Picher-Labrie, Roxanne Turgeon
Universite Laval (Canada)

Open educational resources (OER) are teaching and learning materials belonging to the public domain or published with an intellectual property license allowing their use, adaptation, and distribution free of charge. We will present how creating and using OERs supports the process of democratization of knowledge and can improve accessibility, inclusion, and social justice. We created an OER, based on the principles of Universal Design for Learning, dedicated to teaching evolution and physiology at the university undergraduate level. This resource is in the form of a digital text interspersed with videos, images, and interactive activities. This resource has a creative commons license that allows anyone to use, copy, redistribute, transform, remix and improve it for free (CC-BY-NC-SA). It was co-created in a team including a university professor and 4 graduate students. We will present approaches used to achieve our objective of accessibility and inclusion: various modes of presentation, applied cases using published data, subtitling of videos and transcription, alternative text for all images, compatibility with screen reading software, and PDF version to work offline or in print for students with reduced access to the internet. We will present our aim to diversify the representation of scientists as they are usually portrayed in textbooks, using the creation of video interviews with colleagues from underrepresented groups in science and at different stages of their careers, in which they present their research and their scientific journey in the field. Finally, we will discuss the co-creation process and its benefits.
The genetic and structural basis of a novel iridescence phenotype in the platyfish, Xiphophorus variatus

Nadia B Haghani

Nadia B Haghani, Theresa R Gunn, Tristram O Dodge, Zihao Ou, Gabriel A Preising, Sophia Haase Cox, John J Baczenas, Cheyenne Payne, Daniel L Powell, Molly Schumer

Stanford University (USA), UC Santa Cruz (USA)

A basic goal of biology is to understand the molecular changes that underlie diverse phenotypes. The variable platyfish, Xiphophorus variatus, demonstrates a diverse array of heritable skin pigmentation patterns and has become a tractable system to study the genetic changes leading to adaptive traits. We serendipitously discovered a population of X. variatus that displays a unique iridescence trait in their scales. We confirmed this iridescence is attributed to an enrichment of reflective cells containing subcellular nanostructures. To identify loci associated with iridescence, we performed a genome wide association study (GWAS) and detected a single peak associated with the iridescence trait that spans a surprisingly wide, ~3Mb region. It contains a promising gene candidate, anaplastic lymphoma kinase ligand 2 (alkal2), required for iridophore differentiation in zebrafish. We measured transcriptional profiles of scale tissue and found higher expression levels of alkal2 in iridescent individuals. To visualize structural changes that might explain this transcriptional difference, we used HiFi PacBio reads to generate whole-genome assemblies for iridescent and non-iridescent X. variatus and aligned them at the GWAS peak region. We discovered an 18kb insertion near alkal2 in the iridescent individual that blasts to an LTR/ERV-Foamy element. Given LTR retrotransposons can act as enhancers for nearby genes, we hypothesize that this insertion may be a recently active TE that enhances alkal2 expression, suggesting a possible mechanism by which structural variation produces this novel iridescence trait.
Physiological and Transcriptomic Mechanisms of Hypoxia Acclimation in High-Altitude Deer Mice

Naim M Bautista, Nathanael Herrera, Ellen Shadowitz, Oliver Wearing, Zachary Cheviron, Graham Scott, Jason Storz
McMaster University (Canada), University of Montana (USA), University of Nebraska Lincoln (USA)

A fundamental question in evolutionary biology concerns the relative contributions of phenotypic plasticity vs. local adaptation (genotypic specialization) in enabling wide-ranging species to inhabit diverse environmental conditions. In species that are distributed across elevational gradients, plasticity in environmental tolerance can be expected to influence the ability of species to track upward shifts in temperature isoclines as a result of climate change. Here we conduct a long-term hypoxia-acclimation experiment to assess the relative roles of local adaptation and plasticity in enabling highland and lowland deer mice (Peromyscus maniculatus) to sustain aerobic thermogenesis at simulated elevations that span the species’ current range and exceed the upper limits. We measured whole animal physiological performance (thermogenic VO2max) in conjunction with physiological traits and tissue-specific transcriptomes to gain insight into underlying mechanisms of plasticity. Our results demonstrate that highland natives exhibit a superior thermogenic capacity at the most severe levels of hypoxia, suggesting that the realized niche of the species and its ability to inhabit such a broad range of elevational zones is attributable to a combination of genetically based local adaptation and evolved changes in the plastic response to hypoxia.
The extent of incomplete lineage sorting in divergences of major groups of Neoaves

Naoko Takezaki

Kagawa University (Japan)

Owing to large-scale molecular studies, it has become clear that extant birds are separated into three clades: Palaeognathes, Galloanserae and Neoaves. However, the higher-level phylogenetic relationships of Neoaves are difficult to resolve likely due to rapid diversification. Previous studies recognized ten major groups in Neoaves, “the magnificent seven” and three orphan groups. Their relationships have been differently estimated by different studies and different types of sequence data such as coding and non-coding. This study investigated the extent of incomplete lineage sorting using different types of data [coding sequences, introns, ultra-conserved elements (UCEs), and untranslated regions (UTRs)] from three previous studies using gene concordance factor (proportion that the relationship or the grouping appeared in gene trees) and site concordance factor (proportion of sites that supported them by parsimony-based criterion). Most of the relationships of the major groups appeared in estimated phylogenies with concatenated sequences and even clustering of the major groups were supported by less than few percent of gene concordance factors and 30% - 40% of site concordance factors, many of which were close to 33.3%. The values of both concordance factors were similar for all the different types of sequence data. This result indicates that in the relationships of many of the major groups the extent of incomplete lineage sorting is extreme and that their divergences are close to multifurcation.
Primordial ferredoxin reduction by hydrogen with iron at the intersection between geochemistry and biochemistry

Natalia Mrnjavac

Natalia Mrnjavac, Max Brabender, Delfina P. Heinriques Pereira, Manon L. Schlikker, Martina Preiner, William F. Martin
Heinrich Heine University (Germany), Institute of Molecular Evolution, Max Planck Institute for Terrestrial Microbiology (Germany), Microcosm Earth Center

The acetyl-CoA pathway of carbon fixation is considered to be the primordial metabolic route to carbon and energy in the first life forms. In modern cells, the direct electron donor for CO2 reduction via the acetyl-CoA pathway is ferredoxin, a small protein that is itself reduced by molecular hydrogen. This reaction is catalyzed by hydrogenases, which had to evolve a complex and fine-tuned process, flavin-based electron bifurcation, in order to overcome the thermodynamic barrier to ferredoxin reduction by hydrogen at physiological conditions. Ferredoxin is a small and simple electron carrier harboring one or two FeS clusters, and is likely to have evolved before flavin-based electron bifurcation and hydrogenases. We experimentally showed that hydrogen can reduce ferredoxin with Fe(0) non-enzymatically in a reactor, with conditions mimicking those in serpentinizing hydrothermal vents. This finding suggests that serpentinizing vent environments on the early Earth could replace the physiological function of flavin-based electron bifurcation, forging a link between geochemical processes in Earth's crust and nascent metabolism. Under the serpentinizing hydrothermal vents hypothesis for the origin of life, hydrogen-dependent ferredoxin reduction by iron would reflect a phase of biochemical evolution in which protometabolic reactions required direct physical contact of proteins with geochemical surfaces, before the origin of self-compartmentalized free-living cells.
Pilot investigation of historical yak “crusties” using paleogenomics
Natividad Lupianez Corpas
Natividad Lupianez Corpas, Meghana Natesh, Maanasa Raghavan
University of Chicago (USA)

Museum specimens offer an invaluable resource for studying species' biogeography and evolutionary histories. Coupling this resource with recent improvements in paleogenomic analyses has opened the doors to inferences on the past that were previously inaccessible, though biomolecular preservation and contextual gaps in museum records may limit the scope of such studies. Here, we present results from a pilot on “crusties”, or dried bits of tissue, scraped from historical bones curated at the Field Museum. While crusties offer a minimally-destructive source of DNA from museum samples, they may not be as rich in endogenous DNA as bone or more intact tissues. Our pilot focuses on yak (Bos grunniens), a high-altitude adapted bovid species restricted today to the Tibetan plateau and surrounding regions. Wild yak (Bos mutus) populations persist on the Tibetan Plateau, though in dwindling numbers. Museum records for these yak samples are incomplete, noting the 1960s-1970s as the time of acquisition from the Chicago Zoological Society with no reference to the original sampling locations. Moreover, it is unclear whether the sampled yak are of a wild or domestic origin. We generate shotgun sequencing data from four individuals to evaluate DNA preservation and the feasibility of the generated data for establishing their genetic affinity with published wild and domestic yak across their range. Our pilot showcases the utility of paleogenomics on museum collections to supplement historical records in addition to offering a snapshot of genetic diversity in the recent past.
Investigating Functional Convergence Across Millions Of Years Of Evolution

Navya Shukla

Navya Shukla, Laura E Cook, Maddy L Comerford, Charles Y Feigin, Andrew J Pask, Irene Gallego Romero

Center for Genomics, Evolution and Medicine, Institute of Genomics, LaTrobe University (Australia), Lawrence Berekely National Laboratory (USA), School of BioSciences and Melbourne Integrative Genomics, St. Vincent's Institute of Medical Research (Australia), University of Melbourne (Australia), University of Tartu (Estonia)

Understanding the molecular mechanisms underpinning convergent traits can help characterize forces that drive adaptation and provide insight into the predictability of the evolutionary process. The extinct thylacine, or Tasmanian tiger, has been long known to have a very similar morphology to that of the eutherian canids, particularly in the craniofacial region, despite having shared a common ancestor ~160 million years ago. Previous comparative genomics studies between the thylacine and a representative canid, the gray wolf, have found that much of underlying molecular convergence between these species is found in non-coding, regulatory regions of the genome. These analyses identified a set of 339 thylacine-wolf accelerated regions (TWARs), putative craniofacial cis-regulatory elements that have undergone accelerated evolution in both thylacine and wolf lineages. Using a Massively Parallel Reporter Assay, we functionally characterize the activity of orthologous TWAR sequences from 7 different species—the species of interest (thylacine and wolf), ancestral reconstruction from their clades (Dasyuromorphia and Carnivora), two related outgroup species with different craniofacial phenotypes (Tasmanian devil and panda) and a control species (mouse). We test their activity in two mouse cell populations involved in craniofacial development. By comparing changes in regulatory activity of orthologous elements, we assess which regions showing a convergent shift in molecular activity also underlie convergent phenotypes. Dense 10bp tiling of our sequence library allows the resolution to detect specific transcription factor motif changes driving the observed phenotypic changes.
Microbial Diversity Conservation and Characterization in the Yucatan Peninsula: Insights from Native Bee’s Honey and Water Sources in Calakmul Reserve.

Nelly Sélem-Mojica

Nelly Sélem-Mojica, Aurora Xolalpa-Aroche, David Alberto García-Estrada, Haydee Contreras-Peruyero, Karina Verdel-Aranda, Enrique Jesús Delgado Suárez

(C)INVESTAV (Mexico), Colegio de Posgraduados de Campeche (Mexico), UNAM (Mexico), UNAM CCM (Mexico), Universidad Intercultural Maya de Quintana Roo (Mexico)

The Yucatan Peninsula is part of the Mesoamerican Biological Corridor, where various conservation activities take place. Understanding our microbial diversity is crucial for its protection. This study examined microbial diversity in i) honey samples from native bee species Melipona beecheii and Scaptotrigona mexicana and ii) water sources within the Calakmul Biosphere Reserve. Meliponiculture has deep roots in Yucatan, serving as an economic activity and cultural component. Notable is the honey of two native bee species: Melipona beecheii and Scaptotrigona mexicana. Honey possesses antibiotic properties, contrasting industrial sugar syrup. This project aims to valorize traditional Yucatan meliponiculture via microbial characterization of its honey. Metagenomes show Lactobacillus presence; one is significantly more abundant in M. beecheii honey and previously isolated from vinegar. No biosynthetic gene clusters were previously known. However, the honey-derived genome shows such antibiotic-producing gene clusters warranting further investigation. Calakmul, a UNESCO World Heritage Biosphere Reserve, lacks permanent water bodies. Instead, seasonal rains fill "aguadas," temporary water formations vital for macrofauna, including endangered species. Samples were collected from three sites with varying protection levels during 2017-2019, establishing a baseline microbial diversity pre-Mayan Train megaproject. Surprisingly, semi-urban samples showed higher alpha diversity than protected area samples.
The probability of a unique gene occurrence at the tips of a phylogenetic tree in the absence of horizontal gene transfer (the last-one-out)

Nico Bremer

Gene loss is an important process in gene and genome evolution. If a gene is present at the root of a rooted binary phylogenetic tree and can be lost in one descendant lineage, it can be lost in other descendant lineages as well, and potentially can be lost in all of them, leading to extinction of the gene on the tree. In that case, just before the gene goes extinct in the rooted phylogeny, there will be one lineage that still retains the gene for some period of time, representing a ‘last-one-out’ distribution. If there are many (hundreds) of leaves in one clade of a phylogenetic tree, yet only one leaf possesses the gene, it will look like the result of a recent gene acquisition, even though the distribution at the tips was generated by loss. Here we derive the probability of observing last-one-out distributions under a Markovian loss model and a given gene loss rate \( \gamma \). We find that the probability of observing such cases can be calculated mathematically, and can be surprisingly high, depending upon the tree and the rate of gene loss. Examples from real data show that gene loss can readily account for the observed frequency of last-one-out gene distribution patterns that might otherwise be attributed to lateral gene transfer.
MiRNAs have long been associated with animal innovation. Specifically, it is known that bursts of miRNA diversity coincide with functional and morphological innovation. Indeed, placental mammals have 13 miRNA families identified thus far that are unique and conserved in this clade. However, we know from gene regulatory network biology that “newer” miRNAs have fewer targets and major transitions such as the shift in reproductive strategy in mammals from oviparity to vivipary are most likely supported by a combination of new miRNAs and older miRNAs. Here we consider the role of “older” miRNAs (i.e., miRNAs that predate the origin of mammals) in the evolution and diversification of mammal reproductive strategies. Using a set of >400 mammal genomes, we have applied machine learning approaches to identify miRNAs and their homologs across all species in our set. From our large presence absence matrix, we extracted all miRNAs that emerged prior to mammal divergence, i.e. are “older”, and we mapped their evolutionary trajectories (e.g., expansions and contractions of miRNA gene families). Then using independent methods have identified the targets, pre- and post-mammal origin. Here we explore shifts in the regulatory network through time and to identify novel patterns coincide with the origin of mammals. The likely impacts of these “older” miRNAs and their target profiles on mammal implantation and subsequent embryonic development are discussed.
**Inferring distributions of fitness effects of wild house mice from allele frequency spectra**

Olivia Fernflores, David Castellano, Emanuel Masiero da Fonseca, Travis J Struck, Ryan N Gutenkunst
University of Arizona (USA)

The distribution of fitness effects (DFE) of new mutations is a key input into the evolutionary process. We aim to infer the DFE among multiple populations of wild house mice, so that the extensive knowledge of mouse molecular biology can be leveraged to understand the biological basis of the DFE. Here we present distributions of fitness effects for pairs of populations from Iran and France, France and Germany, and Germany and Heligoland. We find that for all population pairs, the best DFEs are those that infer a high correlation between mutational fitness effects in the two populations.
Using saturation mutagenesis to map the evolutionary forces that shaped human regulatory elements

Omer Ronen, Simon Fishilevich, Yoav Navon, Georgii Bazykin, Tomas Marques-Bonet, Fumitaka Inoue, David Gokhman

Department of Biomedical Informatics, Department of Molecular Genetics, Harvard Medical School (USA), Institute for the Advanced Study of Human Biology (WPI-ASHBi), Institute of Evolutionary Biology (UPF-CSIC) (Spain), Kyoto University (Japan), Weizmann Institute of Science (Israel)

Adaptation via natural selection is the process by which the incredible fit between species and their environment has evolved. However, despite the good understanding of evolutionary adaptation at the phenotypic level, we have a limited understanding of the genetic changes that made us human. Particularly, we are still far from understanding how adaptation presents itself at the molecular level. This task is even more difficult in the noncoding parts of the genome, where most adaptations occur. Saturation mutagenesis is an approach that allows researchers to compare true evolutionary changes to the full spectrum of possible mutations. We apply it using massively parallel reporter assays (MPRAs), allowing us to generate the largest saturation mutagenesis library of regulatory regions in the human genome, consisting of 500 diverse regulatory elements. To identify signatures of selection, we leveraged and expanded classical evolutionary tests, such as the McDonald-Kreitman test, to apply them to noncoding regulatory elements. We also developed meta-analysis approaches to gain insight not only into specific regions, but also into classes of regulatory elements. We identified signals of positive and negative selection in various elements, and discovered specific variants that have likely driven human-specific phenotypes, e.g., in IRF4, a key transcription factor that affects pigmentation and skin diseases. This work allowed us to study how selection has shaped human noncoding regulatory elements, and to shed light on the role of key substitutions in human adaptations.
Diving deeper: demystifying the distribution of the ‘rare’ deep-sea amphipod, Alicella gigantea.
Paige J. Maroni

Paige J. Maroni, Yakup Niyazi, Alan J. Jamieson
Minderoo-UWA Deep-Sea Research Centre, School of Biological Sciences and Oceans Institute, The University of Western Australia (Australia)

Amphipods (Arthropoda: Crustacea) inhabit all aquatic environments worldwide and are amongst the most specious and ecologically diverse orders of crustaceans encompassing over 10,000 extant species. Their ubiquity extends to the deep-sea, particularly the abyssal (3000-6000 m) and hadal environments (6000~11000 m), yet, due to sporadic and infrequent sampling, plus a lack of physical specimens retrieved from the deep-sea, most amphipod species are classified as “rare”. One such example, Alicella gigantea Chevreux, 1899, known as the “supergiant amphipod” has historically been collected in small numbers, perhaps signifying low population densities, and providing a sense of rarity. Consequently, little is known about the demography, genetic variation, and population dynamics of A. gigantea with only six studies having published DNA sequence data. As more records emerge across the vastness of the deep-sea, and from depths that are rarely sampled, there is an ever-growing body of evidence to show that A. gigantea should be considered far from rare. In this study, we compile all records of A. gigantea and use two mitochondrial and one nuclear gene (COI, 16S and 28S) to explore the distribution patterns of A. gigantea across all oceans, further examine the phylogeographic structure across the species and discuss the species history throughout tectonic time. Specifically, we utilised time-calibrated phylogenies and habitat mapping software to estimate divergence times and discuss these speciation events in relation to global oceanographic and geological events, while also investigating haplotype and phylogenetic data to understand this species contemporary distribution.
Comparative annotation of Drosophila genomes
Pankaj Dhakad

Pankaj, Bernard, Darren Dhakad, Kim, Obbard
Stanford University (USA), University of Edinburgh (United Kingdom)

Accurate prediction of genomic features lays the foundation for evolutionary analyses. Annotation methods typically fall into two categories: ab-initio approaches, which predict gene structures using statistical approaches such as Hidden Markov Models; and sequence alignment-based methods, which employ sequence alignment of known RNAs, cDNA, and proteins to discover transcripts. In this study, we employed the Comparative Annotation Toolkit (CAT) pipeline, supplemented with Braker3, to annotate 304 Drosophila genomes. CAT offers a comprehensive approach, integrating transcript projection, transcriptome, and proteome alignments, along with simultaneous gene-finding. CAT provides a flexible way to simultaneously annotate entire clades and identify orthology relationships. Nevertheless, CAT's efficacy in identifying species-specific genes appears limited in the absence of RNAseq data or high-quality reference annotations. To address this limitation, we complemented CAT annotations with those generated by braker3, leading to substantial improvements in annotations. To assess annotation quality across these Drosophila species, we employed phylogenetic mixed models to infer phylogenetic effects on gene number/length, and effect of availability of RNAseq and distance from a reference species on gene number/length. Finally, leveraging the comprehensive annotation data, we assessed the codon usage bias patterns in the genus Drosophila. This study demonstrates the importance of advanced annotation pipelines for accurate genome annotation. The annotation dataset aids in understanding gene structures and species-specific features, while also facilitating research into evolutionary phenomena like codon usage bias (CUB) in Drosophila.
Hill-Robertson interference effects bias the inference of fitness effects of new mutations in partially-selfing populations

Parul Johri

Austin Daigle, Parul Johri

Department of Biology, Department of Genetics, University of North Carolina (USA)

The accurate estimation of the distribution of fitness effects (DFE) of new mutations is critical for population genetic inference but remains a challenging task. While various methods have been developed for DFE inference using the site frequency spectrum of putatively neutral and selected sites, their applicability in species with diverse life history traits is not well understood. Selfing is common among eukaryotic species and can lead to decreased effective recombination rates, increasing the effects of selection at linked sites, including interference between selected alleles. We employ forward simulations to investigate the limitations of current DFE estimation approaches in the presence of selfing and linked effects of selection. We find that distortions of the site frequency spectrum due to Hill-Robertson interference in highly selfing populations lead to mis-inference of the deleterious DFE of new mutations. While accounting for the decrease in the effective population size due to linked effects of selection largely accounts for the observed bias in populations with moderate levels of selfing, this correction is unable to accurately estimate the DFE in highly selfing populations. In addition, the presence of cryptic population structure and uneven sampling across subpopulations leads to the false inference of a deleterious DFE skewed towards effectively neutral/mildly deleterious mutations. Finally, the proportion of adaptive substitutions estimated at high rates of selfing is substantially overestimated. Our observations apply broadly to species and genomic regions with little/no recombination.
Insights into Human Neurotransmission through Molecular Evolution Studies in Bacteria
Paul C Taylor

Jake A Fowler, Jennifer A Miles, Andrew D Millard, David J Scanlan, Michael E Webb, Ana M Rojas-Mendoza, Paul C Taylor
Centro Andaluz de Biología del Desarrollo (Spain), University of Leeds (United Kingdom), University of Leicester (United Kingdom), University of Warwick (United Kingdom)

Neurotransmission is fundamental to human health. Disorders cause e.g. Parkinson’s disease and mental health conditions (10.1001/archneur.1986.00520100062016). While major advances in medicine resulted from study of species like E coli and yeast (10.15698/mic2022.04.773), unicellular organisms offer limited insights into neurotransmission. We research homologs of proteins involved in neurotransmission that are found in multicellular cyanobacteria (10.1016/j.bbrc.2021.09.071). We focus on the archetypal machinery of neurotransmission (glutamate receptors, calmodulin) and enzymes key to biosynthesis of neurotransmitters including dopamine and serotonin. The dominant mode of neurotransmission involves activation of ionotropic glutamate receptors (iGluR) that allow calcium ions to enter neurons. Cyanobacteria are known to have homologs (GluR0) of iGluR (10.1016/j.jmb.2007.10.081), but to date there is no comprehensive molecular evolution study of the links between iGluR and GluR0. Using a combination of phylogenetic analysis, statistical coupling analysis and modelling, we identified key features of extant iGluR and GluR0 that emerged during critical stages of their evolution and that may have as yet unremarked medical significance. Once in the neuron, calcium binds to calmodulin. We showed that the closest bacterial homologs to animal calmodulins are calmodulin-like domains in cyanobacterial proteins. We used NMR spectroscopy to show how calcium affects the conformation of one such domain from Okeania. GTP cyclohydrolase I (GCH1) is key to biosynthesis of tetrahydrobiopterin, necessary for neurotransmitter biosynthesis. Reminiscent of GluR, calmodulin and our older work (10.1016/j.bbrc.2018.12.171;10.1007/s00239-020-09965-x), phylogenetic analysis shows cyanobacterial GCH1 are the most closely-related prokaryotic enzymes to animal GCH1. We will use them to probe GCH1 kinetics.
Molecular responses of brain functions to chronic low dose ionizing radiation in wild mammals from the Chernobyl Exclusion Zone

Rebecca Saager

Rebecca S Saager, Vladimir M Jovanovic, Zbyszek Boratyński, Phillip Lehmann, Katja Nowick, Anne Hartleib

CIBIO (Portugal), Freie Universität Berlin (Germany), Stockholm University (Sweden), Universität Greifswald (Germany)

Ionizing radiation (IR) impacts cognition, brain size, and brain development in humans and other animals. Cognition is a crucial trait for animals adapting to stressors like IR contamination. Here, we analyzed RNA-sequencing data from the frontal cortex, visual cortex, hippocampus, and amygdala of bank voles (Myodes glareolus) from different areas in the Chernobyl Exclusion Zone with varying levels of IR exposure, to discover molecular determinants in the response to increased radioactive contamination in an ecological realistic context. We assessed gradual and categorical gene expression patterns in relation to internal dose rate, identifying numerous genes potentially affected by radiation. Of the 198 genes with categorical expression changes affected by IR, four were differentially expressed across all studied tissues. The hippocampus, the brain region involved in learning and memory, showed the highest number of differentially expressed genes. Gene set enrichment within the affected hippocampal genes points to a change in the astrocyte function(s) in this brain region. This further highlights the impact of IR on modulating spatial memory, reward and fear behavioral response. We also performed weighted gene co-expression network analysis to elucidate molecular pathways involved in bank vole adaptation to radiation. We further validated our results using matching tissues of the yellow-necked mice (Apodemus flavicollis) from the same study area. Overall, this research advances our understanding of the ecological and physiological consequences of nuclear accidents and provides valuable insights into the broader impacts of IR on wildlife and humans alike.
Population-level degeneration of young Y chromosomes in the mountain pine beetle.
Ryan R Bracewell

Ryan R Bracewell, Medhavi Verma, Barbara Bentz, Scott Geib
Indiana University (USA), USDA ARS (USA), USDA Forest Service (USA)

Neo-sex chromosomes provide a unique opportunity to study the cascade of genomic changes that occur when autosomes transform into heteromorphic sex chromosomes. Much of the work to date has focused on longer-term changes in gene content and gene regulation. Much less is known about natural variation within a species and how neo-sex chromosome variation might segregate in natural populations. The mountain pine beetle is a unique system to explore intraspecific variation because there appears to be three highly differentiated neo-Y chromosomes that have been associated with patterns of reproductive incompatibility between populations. To fully understand the extent of neo-XY differentiation and degeneration, we used a combination of PacBio HiFi, Hi-C, and RNA-seq data to assemble and annotate haplotype resolved male and female assemblies from three distinct neo-X/Y types of mountain pine beetle. We find that the fusion that led to the neo-XY likely involved a fusion between an ‘old’ X chromosome and three autosomes. We find an astonishing level of neo-Y chromosome degeneration and structural changes in the form of both large- and small-scale inversions. Transposable elements are enriched on the neo-Y and we find population specific patterns of degeneration and gene loss suggesting that certain neo-X/neo-Y combinations could have gene dosage issues, thereby providing a potential mechanism explaining reproductive incompatibilities.
Genomic architecture in social insects is more strongly associated with phylogeny than social behavior
Sara E Miller

Sara E Miller, Garrett Behrends, Thomas Hagan, Samir Kuinkel
University of Missouri - St. Louis (USA)

The evolution of sociality in insects has been predicted to reduce effective population sizes, in turn leading to changes in genome architecture, including higher recombination rates, increased GC-biased gene conversion (gBGC) and greater intragenomic variation in GC content to maintain castes through differential methylation. As the number of sequenced insect genomes continues to grow, it remains an open question which of these features of genome are common across all social insect genomes. A major challenge to determining such commonalities has been the lack of phylogenetically controlled analyses. Here we analyze genome assemblies for 352 species of Hymenoptera and seven species of Blattodea to test if GC content, genome size, distribution of CpG sites or codon bias repeatedly differed between social and non-social species. Overall, we found little support for predictable changes in genome architecture associated with sociality. Social species have slight differences in GC content, genome size and CpG site distributions compared to non-social species, however these differences disappeared after accounting for phylogenetic relationships. Similarly, Hymenoptera genomes had an overall AT-bias at four-fold degenerate nucleotides that were similar for social and non-social species. We performed phylogenetically controlled analyses within two well represented families, Apidae and Halictidae, but only found a significant negative relationship between sociality and GC content within Apidae. In all, these results suggest that unique origins of social behavior may produce unique trends in genomic architecture. Our study highlights the need to examine genome architecture across independent origins of social behavior.
MYB transcription factors involved in Ophrys orchids labellum development
Sara Edith Garcia-Morales

Sara E Garcia-Morales, Mattia Alessandriní, Alessia Russo, Philipp Schlüter
Department of Plant Evolutionary Biology, Department of Plant and Microbial Biology and Zürich-Basel Plant Science Centre, Department of Systematic and Evolutionary Botany and Zürich-Basel Plant Science Centre, Institute of Biology, University of Hohenheim (Germany), University of Zurich (Switzerland)

The sexually deceptive Ophrys orchids attract specific pollinator species, deceiving them via mimicry of floral traits such as odor, color, and shape, specifically at the labellum. Our study focuses on O. sphegodes, whose highly specific pollinator interaction has led to a mechanism that promotes reproductive isolation. Deciphering these traits’ genetic regulation is necessary to better understand the molecular basis of pollinator attraction. MYB transcription factors (TFs) are known to play an essential role in molecular processes. In orchids, several MYB TFs have been studied in well-known biosynthetic pathways and show precedents that pigment accumulation displays spacio-temporal specificity. To study the molecular mechanisms behind labellum pigmentation, we first identified the O. sphegodes MYB TFs. Previously published genomes had come mainly from orchids distant from the Ophrys genus. Since the recent publication of the Plathantera and Ophrys genomes, where duplication events were described across their genomes, a characterization of the MYB TFs was needed to identify which clades underwent duplication events in Ophrys sphegodes. We then analyzed gene duplication, selection, and gene expression in clades associated with labellum development. This data will contribute to our understanding of the molecular processes that shape the complex Ophrys labellum.
Himalayan Langurs (Semnopithecus schistaceus) are one of the most widely distributed colobine monkeys found in the Himalayas from Pakistan in the west to Bhutan in the east. Further, their distribution encompasses a wide range of altitudes (300-3000 meters) and is interspersed with numerous deep river valleys. In this study, we investigate the role of riverine barriers and altitudinal gradients in shaping the population genetic structure of these langurs. Previous approaches relied on mitochondrial markers, but this study employs nuclear microsatellite markers for enhanced genetic resolution. Fecal samples were non-invasively collected across 4 Indian Himalayan states: Jammu and Kashmir, Himachal Pradesh, Uttarakhand, and Sikkim based on distribution records from past studies. DNA extraction, PCR amplification of 7 microsatellite regions, and microsatellite genotyping were conducted followed by analysis using Arlequin, PCoA, STRUCTURE, Neighbor-joining tree, AMOVA, and paired Mantel test with pairwise F_ST and pairwise altitudes. The preliminary results show extensive gene flow across riverine barriers and altitudinal gradient indicating a lack of significant genetic differentiation. This implies that the Himalayan heterogeneous landscape features (altitude and rivers) might not be significantly affecting the genetic diversity of these langurs. These results do not support splitting the Himalayan langurs into multiple species/subspecies.
Archaeogenomics of maize evolution in the South American Andes
Shuya ZHANG

Shuya ZHANG, Robin Allaby
Smithsonian Institution (USA), University of Science and Technology of China (China), University of Warwick (United Kingdom)

Maize used as a staple crop evolved from a wild plant called teosinte due to human activities in Mexico approximately 9000 years ago. It then was brought through Central America by humans around 7500 years ago, and later, around 6500 years ago, it was spread into South America. While it is widely known when maize spread to various regions, there is still debate regarding whether maize had undergone complete domestication prior to its arrival in South America. Screening for the best-preserved representative samples across the collection of Andean maize specimens established 9 samples to take forward for genome analysis. Structure analysis placed the ancient genomes of the Andes, all pre-Inca, to a single group to the exclusion of modern Andean landraces. Genetic distances between ancient genomes established links between maize samples from distinct cultural groups and the occurrence of network hubs. Analysis of selection signatures showed two possible episodes of selection involved in the adaptation to altitude, the first in pre-Inca times and the second more recently. The relationships revealed between ancient genomes provide evidence for cultural interaction through maize in the Andes spreading from a hub in the Southeast, and later spreading across the continent to Brazil. The evidence also suggests the hypothesis of a second wave of maize into the Andes, possibly as late as the Inca period, which may have brought secondary adaptations to altitude.
Stickwitu: a macroevolutionary approach reveals adhesive organs associated traits in exclusive meiofauna phyla

Sónia CS Andrade

Thainá Cortez, Maikon Di Domenico, Sónia CS Andrade
Universidade Federal do Paraná (Brazil), Universidade de Sao Paulo (Brazil)

The meiofauna encompasses all animals ranging from 40 to 1000 ?m in size with representatives in 23 metazoan phyla. Some include clades with few representatives of meiofauna (as Annelida, Nemertea, Mollusca), while others are exclusively meiofaunal (Kinorhyncha, Micrognathozoa, Gnathostomulida, and Loricifera). Certain traits are well-established as potentially convergent among these groups, characterizing adaptations to the interstitial environment, such as the presence of adhesive glands for adhesion to grains. To understand the molecular mechanisms potentially associated with these adaptations, we used a comparative phylogenomic approach to identify novel and ancestral genomic regions among these species, which may be associated with the traits of interest. For this purpose, we analyzed genomic data from 162 marine invertebrates taxa, being 64 meiofaunal. After identifying 577,966 homologous groups (HG) with MCL, we employed PAPS pipeline (Phylogenetic_Aware_Parsing_Script) to infer ancestral HGs and novel-core HGs. A significant number of HGs were identified for the branches with exclusive meiofaunal species. Several gene ontologies with relevant metabolic roles were found, including some associated with body movement, such as muscle contraction, regulation of cilium movement, as well functions related to adhesion and secretion production, such as cell-cell adhesion via plasma membrane adhesion molecules, and proteoglycan biosynthetic process, pathways that might be linked to the functions of adhesive organs. Therefore, by comparing these genomic regions, we were able to assess molecular mechanisms at genomic scale, providing insights into the evolutionary processes underlying the key traits essential for the success of meiofauna in the interstitial environment.
An interdisciplinary approach to ascertaining the utility of various skeletal elements in non-human aDNA studies

Sophy Charlton

Ludwig Maximilian University (Germany), Queen Mary University of London (United Kingdom), University of Cambridge (United Kingdom), University of Oxford (United Kingdom), University of York (United Kingdom), Vilnius University (Lithuania)

All ancient DNA (aDNA) studies are limited by the degraded and fragmented nature of aDNA, as well as the low endogenous DNA content of the majority of archaeological samples. In recent years, the targeted sampling of specific skeletal elements in humans - particularly the petrous portion of the temporal bone (pars petrosa) - has been utilised to mitigate some of these issues. In contrast, the potential of similar approaches in the sampling of faunal populations is yet to be fully explored. Here, in order to empirically characterise the utility of petrous for aDNA analysis relative to other skeletal elements in non-human species, we present the first comparative aDNA study utilising paired dog (Canis lupus familiaris) skeletal elements (petrous vs. bone/tooth) (n=56). We informed our targeted sampling strategy through the use of micro-CT scanning, indicating denser areas of cortical bone. This allowed for the exploration of aDNA preservation across skeletal elements, considering both endogenous and exogenous fractions. Our results suggest that endogenous DNA yields from petrous may be up to 200x higher than other skeletal elements from the same individual. However, petrous appear to have lower microbial species richness when compared to other skeletal elements, including pathogens. We leverage the findings of our research to identify the most suitable substrate(s) for the genetic analysis of archaeological dog specimens and other mammalian fauna.
Fertilization, or the fusion of sperm and egg, is essential for most sexually reproducing organisms. The genes associated with reproduction undergo rapid evolution yielding tremendous diversity in reproductive proteins between taxa and closely related species. Despite more than a century of research, the molecular mechanisms by which sperm bind and dissolve the extracellular matrices surrounding the ovum remain unclear. Conserved across all animals, the egg coat is a species-specific barrier that restricts access to the ovum plasma membrane and limits polyspermy, which is lethal in most sexually reproducing species. As broadcast spawners that undergo external fertilization, abalone are a classic model for studying the molecular and evolutionary basis of fertilization where sperm lysin non-enzymatically dissolves the egg coat by binding its egg protein partner VERL. Here, we investigate lysin and VERL interactions in two abalone species (Haliotis rufescens and H. fulgens) to gain a mechanistic understanding of species-specific egg coat dissolution. VERL is a highly repetitive molecule with multiple tandem domains: the two most N-terminal domains (VN1, VN2) are distinct and rapidly evolving, while the remaining domains are homogenized and conserved. We found that species specificity of egg coat dissolution is mediated by weak but specific lysin–VN1 interactions. However, mechanical dissolution of the egg coat requires tight but non-specific lysin–VN3 binding, which induces a conformational change that destabilizes and restructures the egg coat. Structural modularity between species specificity and mechanistic dissolution within VERL may facilitate speciation of abalone by rapid sexual coevolution of gametic interactions.
Archaea constitute the most understudied domain of life, with different cellular structures, genetic processes, and metabolic pathways than Eubacteria or Eukaryota. This has led to unusual adaptive characteristics in archaean taxa, including their ability to thrive in extreme environments. Archaea also represent the most metabolically diverse branch of life, having great potential as bioconverters and bioremediators. Recently, Medina-Chávez & Kalambokidis et al. (in preparation for Evolution) found that Halobacterium salinarum, an extremophilic archaean, can adapt quickly to multiple co-occurring environmental stressors by expanding its niche. This pivotal result challenges the expectation that adaptation requires specialization and resultant tradeoffs. A critical follow-up question is: What are the genomic and metabolic bases for H. salinarum's rapid adaptation that enables this archaean to thrive in extreme environments? To explore this, we integrate whole genome sequencing and metabolomic analyses to compare biomolecular architectural changes between ancestral and derived populations. This has enabled us to (1) identify localized changes in H. salinarum DNA at the resolution of single-nucleotide polymorphisms; (2) characterize shifts in gene expression related to phenotypic adaptation; and (3) delineate the adaptive potential of this lineage for generating biotechnologically-relevant metabolic products. This work provides the genomic and metabolomic bases for the first-ever long-term archaean evolution experiment. This foundational insight is necessary for utilizing the most understudied domain of life as a platform for biotechnological innovation.
CRISPR is mutagenic, even in the absence of gRNA.
Stephan Baehr
Stephan Baehr, Samuel Perez, Adalyn Brown, Alyssa Cenzano, Katelyn Hancock, Lea Patrick, Michael Lynch
Arizona State University (USA), Biodesign Institute (USA), The Ohio State University (USA)

CRISPR is a gene editing platform with seemingly endless applications, from engineering to science and medicine. The excitement of CRISPR's potential has lead to hundreds of clinical trials aimed at treating human genetic diseases. However, the field is aware that CRISPR's promise is not entirely without risk; CRISPR by its nature modifies the genome, and may have off-target effects. No enzyme or machine is perfect. Researchers have diligently sought to quantify mutation risk, by testing a specific locus, or set of loci. However, to date, we are unaware of any experiment using the most sensitive detection methods of evolutionary biology, which can detect error rates that may occur one in a trillion bases per cell division. We demonstrate a mutational hazard of expressing Cas9 to an untargeted locus of yeast, S. cerevisiae, which demonstrates a dose-response relationship with increasing Cas9 expression. We find a modest 1.2-fold to 2.7-fold increase in mutation rate, as quantified by fluctuation test. We demonstrate that gRNA appears to increase the mutagenicity of CRISPR, be it Cas9, Cas14, or an Adenine Base Editor (ABE). We note that if indels are the primary means of mutation mechanism in Cas9, the indel rate may be 10-fold higher in our high-expression yeast lines relative to WT controls.
Evolution of discrete phenotypic plasticity in a gene regulatory network model

Takahiro Sakamoto

Takahiro Sakamoto, Hideki Innan
SOKENDAI (The Graduate University for Advanced Studies) (Japan)

Phenotypic plasticity is a common strategy for a species to adapt to temporally or spatially varying environments. Some cases of plasticity exhibit a discrete phenotype where the phenotype changes discontinuously along an environment gradient with no intermediate phenotypes arising in intermediate environments (discrete plasticity). It has been argued that discrete plasticity requires a developmental switch to control different phenotypes. However, since most previous theories of phenotypic plasticity are based on the quantitative genetic model and do not consider the developmental process, the evolutionary origin of the developmental switch underlying discrete plasticity remains unclear. In this study, we integrated a gene regulatory network into the population genetics framework and investigated the evolutionary dynamics of phenotypic plasticity in a temporally fluctuating environment. In our model, an individual’s fitness is determined by the expression levels of n genes, which change throughout development due to gene regulation. Optimal expression levels differ between the environments so phenotypic plasticity is favored. An environmental sensor gene is also included to enable environment-dependent gene regulation. Using simulations, we investigated whether a population evolves adaptive plasticity and under what conditions the evolved plasticity tends to be discrete. We found that plasticity evolves in most cases, with the proportion of discrete plasticity varying across different parameter sets. Notably, the proportion of discrete plasticity is high when the fluctuation period of the environment is long. Our study highlights the importance of considering development in theoretical studies of phenotypic plasticity.
Inferring the characteristics of ancient polyploidization events from modern plant genomes
Tamsen Dunn

Tamsen Dunn, Arun Sethuraman
San Diego State University (USA)

Multiple rounds of whole genome duplication (WGD) have occurred throughout angiosperm evolution. The majority of flowering plants appear to have evolved from allopolyploids (hybrids with multiple subgenomes derived from different species), while only a few genera are ancient autopolyploids (multiple subgenomes derived from the same species). Ancient WGDs are commonly detected in a modern genome by analyzing the levels of divergence (Ks) between pairs of duplicate genes. Since Ks is a proxy for time passed, modes in the histogram of these data indicate a spike in contemporaneous gene duplication. WGDs may be detected by fitting mixture models to identify peaks in Ks distributions. However, there are several limitations to this technique, one of which is that the Ks peak gives the divergence time between the duplicated subgenomes, not the actual time the subgenomes coalesce in an ancestor, which for allopolyploids can be a difference of several million years. Furthermore, allo and autopolyploids follow distinct molecular evolutionary trajectories. Thus surveys which seek to generalize the consequences of polyploidy must appropriately take into account the time and mode of speciation. To address this concern, I propose and test a novel, scalable analysis method to determine the mode of ancient speciation from the shape of the Ks histogram of a modern genome. Preliminary work so far has demonstrated that our prototype method provides a better fit to Ks curves from the 1000 plant transcriptome dataset as compared with previous methods, and works well for both allo and autopolyploid-derived species.
Rice is a critical staple crop for global food security. However, rice is highly vulnerable to pests and pathogens, resulting in major yield reductions. Pest and pathogen pressure is predicted to increase with the heightened frequency of weather extremes and anticipated decreases in water availability associated with climate change. This concern becomes even more acute with the global push for direct-seeded rice cultivation because this is not possible without resistance to pests and pathogens. Knowledge of molecular resistance mechanisms is essential for developing novel approaches for pest and pathogen management in rice production. Studies that leveraged an evolutionary systems biology approach demonstrated its ability to predict mechanisms that elucidate genotype-phenotype relationships. Our reanalysis of transcriptomic data from a field-based systems biology study revealed pathways that show evidence of selection in rice sub-populations that experience relatively high pest and pathogen pressure and that are potentially involved in rice defense responses to both root-knot nematodes (RKNs, Meloidogyne) and Fall Armyworms (FAWs, Spodoptera). We identified that differentially expressed genes in rice roots and shoots responding to RKNs and FAWs, respectively, were significantly enriched within fitness-linked modules of a gene co-expression network constructed from traditional rice varieties with genetically and geographically diverse backgrounds. Key plant defense-related pathways enriched in these modules included the diterpenoid phytoalexins. To determine the role of various classes of diterpenoids in rice responses to above- and belowground herbivory we performed functional genetic tests on rice under attack by RKNs and FAWs.
Plants, being sessile organisms, are exposed to abiotic stress (cold, drought, flooding, salinity). However, they have evolved over time, acquiring adaptive strategies to tolerate stress. Mosses have been the first plants to adapt and populate the land. Pseudocrossidium replicatum is a moss classified by our group as totally tolerant to desiccation. This moss can also recover from 10 days of exposure to high salinity (700 mM NaCl) or ultra-freezing conditions (-80oC). Our group is interested in identifying the key genes responsible for abiotic stress tolerance in this moss. In the present work, we performed a transcriptomic and bioinformatic analysis of the protonemal tissue exposed to 200 mM NaCl for 3 hours. Nine NaCl-responsive transcription factors (TFs) were found. Three TFs belong to the family APETALA/ETHYLENE RESPONSIVE FACTOR (AP2/ERF). To make function comparisons between P. replicatum, other non-vascular vascular species and vascular model plants and algae, a phylogenetic analysis of these AP2 TFs (ERF3, DREB3, and DREB2A) was carried out. The three P. replicatum AP2 genes analysed conserved more similarity to the corresponding homologous from Physcomitrella patens, followed by genes from vascular plants and algae. The P. replicatum interactome revealed that it is also more related to the predicted interactome for P. patens; it diverges with respect to the functions in vascular plants, where these genes are mainly involved in regulating seed development. This analysis could infer information about the conservation of structures and functions throughout evolution and identify possible specific evolutionary novelties of P. replicatum.
Towards Simulation Optimization: An Examination of the Impact of Scaling in Coalescent and Forward Simulations

Tessa Ferrari

Tessa Ferrari, Xinjun Zhang, Jazlyn Mooney
University of Michigan (USA), University of Southern California (USA)

Scaling is a common practice to increase computational efficiency. However, there exists a dearth of standardized guidelines for best practices. Few studies have examined the effects of scaling on diversity and whether the results are directly comparable to unscaled and empirical data. We examine the effects of scaling in two model populations, modern humans and Drosophila melanogaster. The reason is twofold: 1) human populations require moderate-to-no scaling for simulations while more dramatic scaling is required for Drosophila; and 2) model populations have empirical data for comparison. We determine whether estimates of diversity, the site frequency spectra, deleterious variation, coalescence, run time, and memory are affected by scaling, the length of the simulated segment, and burn-in time. For humans we test both additive and recessive models of dominance and in Drosophila we test the additive model only. We find that the typical 10N generation burn-in is often not sufficient for full coalescence to occur in human or Drosophila simulations. As expected, memory and run-time increase as the scaling coefficient decreases and the length of the simulated segment increases. Simulating larger segments in humans is preferable. Diversity estimates have less noise, and concatenating these segments together to generate a genome is effective. Conversely, simulating smaller segments in Drosophila is preferable for achieving levels of diversity similar to empirical data; due to larger selective forces being employed from more aggressive scaling and causing background selection to occur at an increased rate.
The human biofluid metabolome is influenced by environmental and genetic factors, including genetic ancestry-linked variation shaped by varying continental demographic histories. How these factors contribute to metabolic processes under extreme environmental exposures of prematurity remains unexplored. We aimed to identify urinary metabolites and pathways that vary with continental genetic ancestry in genetically admixed, extremely premature infants in the neonatal intensive care unit. Our study included 171 infants <29 weeks gestational age including individuals with maternal self-reported Hispanic White, non-Hispanic White, and Black/African American race/ethnicity. Untargeted metabolomics was performed using UHPLC-MS/MS on urines collected at two timepoints between days 6-14 and 23-30 postnatal age. Three-way continental genetic ancestry proportions were inferred using a reference panel including African, European, and Amerindigenous ancestry individuals. We tested for associations between proportions of genetic ancestry and quantitative measures of ~1,000 individual metabolites among all infants combined, and infants stratified by social categories of maternal race/ethnicity. At timepoint 1, we identified 119, 72, and 58 metabolites associated with African, European, and Amerindigenous genetic ancestry proportions, respectively, and 64, 33, and 39 metabolites, respectively, at time 2. Ten metabolites were associated with the same ancestry in the same direction at both timepoints, and several metabolites showed consistent effects across racial/ethnic social identities. Our study suggests that genetic variation shaped by varying continental demographic histories has an impact on metabolite variation of extremely premature infants, and highlights the importance of considering genetic ancestry in clinical studies of diverse populations.
Genomic Offset and Climate Adaptations in Grouse: Relationships with Demographic History

Theodore E Squires

Theodore E Squires, Patrik Rödin-Mörch, Jacob Höglund, Kristinn P Magnússon
Icelandic Institute of Natural History (Iceland), University of Akureyri (Iceland), Uppsala University (Sweden)

Using whole genome re-sequencing data we study the effects of climate influenced declines in effective population size on the accumulation of deleterious mutations and the response to future climate change in populations of cold-adapted avian sister species from the Holarctic: rock ptarmigan (Lagopus muta) and willow ptarmigan (Lagopus lagopus). We reconstruct the demographic histories of the populations and determine their nucleotide diversity, past and present inbreeding, and mutation load. Genomic vulnerability to future climate change scenarios (also known as offset) is predicted for the populations. We show that relatively small and isolated populations have reduced nucleotide diversity, higher signatures of past and present inbreeding, and higher estimates of mutation load. Among the studied populations, the most vulnerable to a mismatch between current and predicted future environments are rock ptarmigan populations in East Greenland, Iceland, and Svalbard, while among willow ptarmigan, subspecies residing on the British Isles are the most vulnerable.
A simulation study to examine the impact of recombination on phylogenomic inferences under the multispecies coalescent model

Tianqi Zhu

Tianqi Zhu, Ziheng Yang
Academy of Mathematics and Systems Science, CAS (China), University College London (United Kingdom)

recombination within locus and free recombination among loci. Yet, in real data sets intralocus recombination causes different sites of the same locus to have different genealogical histories so that the model is misspecified. The impact of recombination on various coalescent-based phylogenomic analyses has not been systematically examined. Here, we conduct a computer simulation to examine the impact of recombination on several Bayesian analyses of multilocus sequence data, including species tree estimation, species delimitation (by Bayesian selection of delimitation models) and estimation of evolutionary parameters such as species divergence and introgression times, population sizes for modern and extinct species, and cross-species introgression probabilities. We found that recombination, at rates comparable to estimates from the human being, has little impact on coalescent-based species tree estimation, species delimitation and estimation of population parameters. At rates 10 times higher than the human rate, recombination may affect parameter estimation, causing positive biases in introgression times and ancestral population sizes, although species divergence times and cross-species introgression probabilities are estimated with little bias. Overall, the simulation suggests that phylogenomic inferences under the multispecies coalescent model are robust to realistic amounts of intralocus recombination.

Timothy D O'Connor

Victor Borda, Omar Caceres, Cesar Sanchez, Carlos Padilla, Diego Veliz-Otani, Marla Mendes Aquino, Bradley Maron, Eduardo Tarazona-Santos, Heinner Guio, Timothy D O'Connor

INBIOMEDIC (Peru), Institute for Genome Science, Institute for Health Computing, Instituto Nacional de Salud (Peru), Instituto de Ciencias Biologicas, Laboratorio de Innovación y Desarrollo, Program in Health Equity and Population Health (USA), Pulmonary Hypertension Center University of Maryland Medical Center (USA), SickKids Institute (Canada), Universidad de Hunuco (Peru), Universidade Federal de Minas Gerais (Brazil), University of Maryland School of Medicine (USA)

Latin American populations have a recent history of extensive migration and admixture that shaped their genetic diversity. Particularly, admixed Peruvians are recognized as having predominantly Indigenous American ancestry. Phenotypically, Peruvians tend to be shorter and, as in many low and middle-income countries, Peru has shown an increasing rate of obesity. We generated genome-wide data for 336 admixed individuals combined with 515 previously described samples representing 13 regions to explore population structure, demographic history, and the relationship between Indigenous American ancestries and variation with complex phenotypes (i.e., height and waist-to-hip ratio). We identified that Indigenous American ancestries in admixed Peruvians mirror the north-south interaction known from archeology and linguistics previously described for Peruvian Natives. We detected the most recent admixture events between European and Native Peruvians dating to 8-10 generations ago which occurred with populations already admixed between distinct original populations. We observed differences between continental ancestry proportions between autosomes and chromosome X suggesting sex bias admixture with a predominant contribution of Indigenous American females. We found a correlation between small stature and a higher proportion of Indigenous American ancestry but no specific locus reached GWAS significance. We did find a significant association between locus 7p21.1 and the waist-to-hip ratio. Interestingly, the affected allele is common in African populations but rare in other 1000 Genomes populations. Our fine-scale population study provides subtle details of the genetic history and diversity, serving as one more step to understanding the broader human genetic variation landscape.
Resolving the regulatory circuitry driving gall development in response to root-knot nematode herbivory in tomato

Tyler R Inskeep, Damaris Godinez-Vidal, Simon C Groen
Center for Plant Cell Biology, Department of Botany and Plant Sciences, Department of Nematology, Institute for Integrative Genome Biology, Riverside (USA), University of California

Plants are parasitized by many species, one of the most ubiquitous being the root-knot nematode (RKN; Meloidogyne spp.). We believe that RKNs exert spatiotemporally variable selective pressure on their host populations because their cold-blooded lifestyles establish latitudinal clines in RKN activity. Despite the presence of RKN-responsive resistance genes in populations of tropical crops such as tomato (Solanum lycopersicum) and rice (Oryza sativa), substantial variation in resistance is still observed across populations lacking resistance genes. Genome-wide analyses across these populations reveal QTLs associated with resistance in regulatory regions or within linkage disequilibrium of known transcription factors, suggesting that rewiring of gene regulatory networks (GRNs) might facilitate host adaptation to RKN parasitism. To evaluate this hypothesis, we focus on the interaction between tomato and the RKN Meloidogyne incognita. We are specifically interested in how RKNs manipulate host development to establish permanent feeding structures, galls, in their hosts’ roots due to our detailed understanding of root developmental GRNs and the hypothesis that root developmental variation might facilitate adaptation to novel environments, across which RKN abundance varies. By profiling tissue-specific cell division patterns in galls, we’re gaining an understanding of the cell types contributing to this interaction. To resolve the GRNs mediating these processes, we’re performing single cell RNA-sequencing of developing tomato galls. To understand how evolution has shaped these processes, we’re comparing gall size across tomato landraces and using genome-wide mapping approaches to understand the network components required for a susceptible interaction.
Establishment of newly-formed allopolyploids through transition to selfing

Uliana Kolesnikova

Uliana Kolesnikova, Alison Scott, Polina Novikova

MPI for Plant Breeding Research (Germany)

Whole-genome duplication is a frequent event across the tree of life, and is especially common in plants. Despite this, newly formed polyploids may struggle to become established. The dramatic karyotypic change of a whole genome duplication comes with reproductive challenges, both due to improper segregation of chromosomes in meiosis and a lack of compatible mating partners. While autopolyploids are compatible with unreduced gametes produced in low amounts by diploid progenitors, allopolyploids (polyploid hybrids) are typically unable to cross with parental diploid species. In Brassicaceae, allopolyploid establishment seems to be facilitated by a transition to self-compatibility. We explain this process using allotetraploid Arabidopsis kamchatika, which inherited the ability to self-pollinate from the diploid progenitor species A. lyrata. Specifically, the self-compatible parental A. lyrata lineage arose in Siberia, and lost self-incompatibility due to degradation in the male component of the self-recognition system. We have shown that this mutation is dominant. Having inherited the dominant non-functional self-incompatibility system variant from Siberian A. lyrata, A. kamchatika transitioned to self-compatibility immediately upon allotetraploid formation. This mechanism of immediate transition to selfing in an allopolyploid due to the self-compatibility of a diploid ancestor is observed in other Brassicaceae (Arabidopsis suecica, Capsella bursa-pastoris, and Brassica napus). We propose that transition to selfing is a necessary step in the establishment of allopolyploids in Brassicaceae. In addition, our findings imply that loss of self-incompatibility in diploid species can impact the evolution of allopolyploids.
Uncovering novel Bacterial and Archaeal diversity: genomic insights from metagenome-assembled genomes in Cuatro Cienegas, Coahuila

Ulises Erick Rodriguez-Cruz

Luis E Eguiarte, Luis E Eguiarte, Valeria Souza
Instituto de Ecologia (Mexico)

A comprehensive study was conducted in the Cuatro Ciénegas Basin (CCB) in Coahuila, Mexico, known for its remarkable microbiological diversity and unique physicochemical properties. Within the CCB, the "Archaean Domes" (AD) site is remarkable due to the abundance of hypersaline, non-lithifying microbial mats. This study focused on analyzing the small domes and circular structures formed in AD by metagenome assembly genomes (MAGs) with the aim of expanding our understanding of the prokaryotic tree of life by uncovering previously unreported lineages, as well as analyzing the diversity of bacteria and archaea in the CCB. A total of 325 MAGs were identified, including 48 Archaea and 277 Bacteria. Remarkably, 22 archaea and 104 bacteria could not be classified at the genus level, highlighting the remarkable novel diversity of the CCB. Besides, AD site showed considerable diversity at the phylum level, with Proteobacteria being the most abundant, followed by Desulfobacteria, Spirochaetes, Bacteroidetes, Nanoarchaeota, Halobacteriota, Cyanobacteria, Planctomycetota, Verrucomicrobiota, Actinomycetes and Chloroflexi. In Archaea, the monophyletic groups of MAGs belonged to the Archaeoglobi, Aenigmarchaeota, Candidate Nanoarchaeota, and Halobacteriota. Among Bacteria, monophyletic groups were also identified, including Spirochaetes, Proteobacteria, Planctomycetes, Actinobacteria, Verrucomicrobia, Bacteroidetes, Candidate Bipolaricaulota, Desulfobacteria, and Cyanobacteria. These monophyletic groups were possibly influenced by geographic isolation, as well as extreme environmental conditions in the pond DA, such as stoichiometric imbalance of C:N:P of 122:42:1, fluctuating pH (5-9.8) and high salinity (5.28%).
Size matters: revealing the genes controlling shoot apical meristem gigantism in cacti

Ulises Rosas

Brenda A. López Ruíz, Michael D. Purugganan, Ulises Rosas

Center for Genomics and Systems Biology, Instituto de Biología, Jardín Botánico, New York University (USA), UNAM (Mexico)

The Cactaceae family has larger shoot apical meristems (SAMs) than other seed plants, such as Arabidopsis thaliana. Mechanisms involved in regulating SAM size in cacti may be similar to those found in model plants. Yet, larger SAM may be achieved by reconfiguring existing genetic elements, leading to novelties found in cacti. The Mammillaria genus is a good model system for studying SAM development due to its short life cycle, high diversity and ease of reproduction. We identified five developmental stages of M. san-angelensis plantlets and compared SAM size with adult plants using histological and microscopic electronic analysis. We also de novo sequenced the whole genome of M. san-angelensis using the PacBio long-read sequencing platform and identified two putative gene homologues that control SAM size: WUSCHEL (WUS) and SHOOT MERISTEMLESS (STM). The MsWUS gene consists of 2409 nucleotides with three exons, two introns, and a putative protein of 273 amino acids. We found a high level of conservation at the nucleotide and amino acid level of the homeodomain, WUS-box, and EAR-like domains compared to A. thaliana. However, the acidic region showed less conservation as compared to the A. thaliana sequence but was highly conserved among several members of the Caryophyllales. We discuss that the evolution of the SAM regulatory network is responsible for some of the novelties seen in cacti.
Population genomics reveals complex genetic history of North Asian human populations

Vadim Stepanov

Vadim Stepanov, Vladimir Kharkov, Nikita Kolesnikov, Aleksey Zarubin
Institute for Medical Genetics, Tomsk National Research Medical Center (Russian Federation)

The whole genome research on genetic diversity in human populations is accumulating significant impact for medical genetics and personalized medicine. Objective of the study is to identify population structure and reveal signals of adaptive evolution in native populations of North Asia. We also aim to demonstrate the importance of incorporating population genetic diversity and ancestry information into present biomedical research. Genome-wide analysis of genetic diversity using own WGS and SNPs array data was performed on a total sample over 2000 individuals representing 42 populations. PCA demonstrates general pattern of genetic diversity distributed according geographic locations of the populations. Major ancestral components can be distinguished as general Western- or Eastern Eurasian, and specific such as proto-Uralic, paleo-Asian, and Beringian. Uneven distribution of ROH among populations demonstrated founder effect in most North Asian populations. Analysis of genetic ancestry with ancient populations reveals significant contribution of Upper Paleolithic North Asian cultures to South and North Siberian people. Paleolithic European hunters-gatherers and Neolithic farmers contributed to modern native South Siberian and North European populations. We demonstrate the population-specific distributions of medically relevant variants in diverse human populations. In conclusion, complex genetic history shaped by early migrations, isolation by distance, founder effects and natural selection has led to current genomic diversity patterns in human populations of North Asia. Future genetic studies and medical programs should take into account the genetic architecture of North Asian peoples. The study was supported by the Russian Science Foundation grant No.22-64-00060.
Traits are not independently evolved and are often constrained by each other. For example, during the artificial selection for desired traits in agriculture, the evolution of reduced survivorship or reproduction is frequently observed. While cases of constrained evolution have been observed, the underlying mechanisms are usually unclear. While the anticorrelation of many traits can be explained by first principles regarding physics or chemistry, the underlying genetic architecture may also play an unequivocal role, if the mutations contributing to the artificial selection lead to the fitness reduction through another trait. To study these questions, the artificial selection for larger cell sizes in budding yeast, Saccharomyces cerevisiae, was used as an efficient way of forward genetic screening. In particular, we used flow cytometry to sort cells with top 10% sizes as the inoculum for the next growth cycle every two days. Throughout the artificial selection for about 20 cycles, we tracked the evolutionary dynamics of cell sizes. The growth curves of evolved populations were also recorded to study their correlation with cell size evolution. Using the lineage-tracking technique with DNA barcodes, we will also dissect the genetic composition of the evolved population to identify the genetic basis of cell size variations and test whether or not the mutations affecting cell sizes simultaneously affect the growth and survival.
Firefly toxin lucibufagins evolved after the origin of bioluminescence

Ying Zhen

Fireflies were believed to originally evolve their bioluminescence as warning signals to advertise their toxicity to predators, which was later adopted in adult mating. Although evolution of bioluminescence has been investigated extensively, the warning signal hypothesis of its origin has not been tested. We test this hypothesis by systematically determining the presence or absence of firefly toxin Lucibufagins (LBGs) across firefly species and inferring the time of origin of LBGs. We confirmed the presence of LBGs in subfamily Lampyrinae, but more importantly, we revealed the absence of LBGs in other lineages including subfamilies of Luciolinae, Ototretinae and Psilocladinae, two incertae sedis lineages, and Rhagophthalmidae family. Ancestral state reconstructions for LBGs based on firefly phylogeny constructed using genomic data suggested that the presence of LBGs in the common ancestor of Lampyrinae subfamily was highly supported, but unsupported in more ancient nodes including firefly common ancestors. Our results suggest that firefly LBGs probably evolved much later than the evolution of bioluminescence. We thus conclude that firefly bioluminescence did not originally evolve as direct warning signals to toxic LBGs and advise that future studies should focus on other hypotheses. Moreover, LBG toxins are known to directly target and inhibit the α subunit of Na+,K+-ATPase (ATPα). We further examined effects of amino acid substitutions in firefly ATPα on its interactions with LBGs. We found that ATPα in LBGs-containing fireflies are relatively insensitive to LBGs, which suggested that target site insensitivity contributes to LBGs-containing fireflies' ability to deal with their own toxins.
Telomere length estimation in the blood of the cynomolgus monkey using the sequencing data

YUN JUNG LEE

Yun-Jung Lee, Se-Hee Choe, Hyeon-Mu Cho, Hye-Ri Park, Min-Gyeong Ko, Sang-Je Park, Young-Hyun Kim, Jae-Won Huh
Korea Research Institute of Bioscience and Biotechnology (KРИBB) (Korea, Korea University of Science & Technology (UST) (Korea, Republic of)

Telomere is a structure with TTAGGG repeat sequences located at the ends of chromosomes, maintaining chromosome stability and protecting chromosomes from degradation. Due to the end replication problem where the ends of linear DNA cannot be fully replicated, telomeres gradually shorten. It is believed that the telomere length (TL) is a representative hallmark of aging. Initially, experimental methods were used to estimate TL, but advances in whole genome next-generation sequencing (NGS) technology and programming tools now allow the length to be estimated using computational methods. Macaca fascicularis is an experimental animal model suitable for aging research because it is evolutionarily close to humans. This study was conducted because there was no paper measuring the TL of monkeys using a computational method. We selected a total of eight individuals, two of the same age each, and performed paired-end sequencing for two years. We used Computel and Telomerecat to measure TL using sequencing data. The sequencing result files were manipulated for using different tools: Computel is an R-based program that takes the fastq file as input, while Telomereat, a Python-based program, uses the bam file as input. The results have shown that both tools showed shorter telomere lengths compared to previous years. When calculate the correlation coefficient of TL by age, Telomereat\’s is more negative than Computel\’s. Due to small sample sizes, neither analysis tools revealed significance for telomere length changes with age; however, we identified a decline in TL with age using Macaca WGS data.
Phylogenomics provides insights into the evolution of dragonflies and damselflies
Zifeng Li

Zifeng Li
Chinese Academy Of Science (Kunming Animal Research Institution) (China)

Odonata is a group with high species diversity in the Insecta, but its phylogenetic relationship with Ephemeroptera and Neoptera has been controversial. Although previous studies have attempted to provide a strong argument for this dispute based on morphological classification and mitochondrial genes, powerful evidence from the genome-wide level is still lacking. Dragonflies are a group with a very high success rate of predation among flying insects which is related to the visual ability of dragonflies. How the dragonfly’s visual ability evolved to cope with this specialized hunting strategy and various environmental clues in an aerial space remains unknown. Here, for the first time, our study systematically analyzed the phylogeny and compound eye evolution of dragonflies by combining long-read genome sequencing, bulk transcriptome sequencing and single-cell transcriptome sequencing. Our phylogenetic analysis revealed that Ephemeroptera was sister to Odonat+Neoptera. Gene family analysis showed that the gene families of rhodopsin gene Rh6 had significant gene family expansion in the ancestor of the Odonata. Furthermore, we analyzed the transcriptome data of 26 tissues of the Chlorogomphus papilio and found that Rh6_Cp1, Rh6_Cp2, Rh6_Cp3 and Rh6_Cp4 were all tissue-specific highly expressed genes in compound eyes. The single cell transcriptome results showed that there were 16 cell types in the compound eye of Chlorogomphus papilio, among which Rh6_Cp1, Rh6_Cp2 and Rh6_Cp3 were highly expressed in the photoreceptor cell types. Our results provide key sources for understanding the association between gene duplication in Odonata genome and the evolution of visual ability.
S39 - Associate Editor's Symposium
Alternative Splicing (AS) is a ubiquitous process in eukaryotes by which different combinations of introns are spliced to produce distinct transcript isoforms from a gene. It has garnered interest in the last few decades because of its significant role in gene function diversification, as well as in cancer. Recent studies have highlighted that the transcript isoforms of human genes are often conserved in orthologous genes from various species. The conserved transcripts are referred to as transcript orthologs, and the identification of transcript ortholog groups provides valuable insights for studying their functions. Exploring the evolutionary history of homologous transcripts enhances our understanding of their protein functions and their origins. It allows us to better understand the role of alternative splicing in transcript evolution. It also provides additional information on gene evolution, such as the history of transcript isoform gains and losses in the evolution of a gene family, and the evolution of alternative splicing. In this presentation, we will describe a computational method and a database we have developed to study transcript evolution. The method first infers orthology relations between transcripts of homologous genes [Ouedraogo-Ouangraoua 2023, https://doi.org/10.1007/978-3-031-36911-7_2], and then uses this information to estimate a transcript phylogeny based on precomputed clusters of orthologous transcripts [Ouedraogo-Ouangraoua 2024, to appear in RECOMB-CG 2024 proceedings]. The TranscriptDB database provides access to the transcript phylogenies and transcript orthology relations inferred for all eukaryotic gene families of the Ensembl Compara database [Ouedraogo-Ouangraoua 2023, https://transcriptdb.cobius.usherbrooke.ca/].
The interplay between mutation, genetic drift, directional and balancing selection shapes the diversity of populations, however, it can be highly complex and challenging to disentangle. To address this, sophisticated models with a high degree of flexibility and the ability to handle multiindividual data are required. We have developed a set of polymorphism-aware phylogenetic models called PoMos. These models are based on the Moran model and have recently demonstrated their effectiveness in inferring species trees, as well as capturing mutational effects, fixation biases, and GC-bias rates. To enhance accessibility, we have implemented these models in the open-source Bayesian inference framework RevBayes. In this study, we further expanded the capabilities of PoMos to investigate neutral, GC-bias, and for the first time, balancing selection. The novel aspect of our approach lies in PoMos’ ability to account for long standing ancestral polymorphisms and incorporate parameters that measure frequency-dependent selection. We implemented validation tests and assessed the model on the data simulated with SLiM and a custom Moran model simulator. We examined real sequences from Drosophila populations to gain insights into the evolutionary dynamics of regions subject to frequency-dependent balancing selection, particularly in the context of sex-limited colour dimorphism.
Evidence for the role of selection for reproductively advantageous alleles in human aging

Jianzhi George Zhang

Jianzhi Zhang
University of Michigan (USA)

The antagonistic pleiotropy hypothesis posits that natural selection for pleiotropic mutations that confer earlier or more reproduction but impair the post-reproductive life causes aging. This hypothesis of the evolutionary origin of aging is supported by case studies but lacks unambiguous genomic evidence. Here we genomically test this hypothesis using the genotypes, reproductive phenotypes, and death registry of 276,406 UK Biobank participants. We observe a strong, negative genetic correlation between reproductive traits and lifespan. Individuals with higher polygenic scores for reproduction (PGS_R) have lower survivorships to age 76 (SV_76), and PGS_R increased over birth cohorts from 1940 to 1969. Similar trends are seen from individual genetic variants examined. The antagonistically pleiotropic variants are often associated with cis-regulatory effects across multiple tissues or on multiple target genes. These and other findings support the antagonistic pleiotropy hypothesis of aging in humans and point to potential molecular mechanisms of the reproduction-lifespan antagonistic pleiotropy.
Hibernation in bears involves a series of seasonally reversible changes in metabolism and physiology. Understanding the mechanisms behind the reversible nature of hibernation has implications for both animal and human health. One of the factors driving changes in hibernation is the significant alteration in gene expression across various tissues. The extent to which hibernation-related shifts in gene expression are shared across hibernating mammals is not well understood. Moreover, the evolutionary origins of hibernation remain unclear. In this study, we utilized RNA-sequencing datasets within various hibernating mammals and performed comparative transcriptomic analysis of multiple tissues to identify hibernation-related genes in each species. By comparing orthologous genes across species, we found that the sharing of hibernation-related differential gene expression is limited. Functional analysis showed that many of the shared genes are involved in metabolic processes, while genes unique to certain hibernating rodents were associated with functions such as immunity and blood coagulation. Additionally, we observed that many groups of orthologous genes contain genes that are upregulated during hibernation in certain species, while downregulated in others, highlighting the complexity and potential rewiring of hibernation regulatory pathways throughout evolution. We couple these analyses with analyses of population-specific selection and broader scale molecular evolution to determine the selective-regimes underlying hibernation-related genes. We delve into the evolution of regulatory regions in brown bears to uncover population-specific selection. Investigating the similarities and differences between hibernation phenotypes at the molecular level can contribute to understanding the evolutionary origins of hibernation.
The distribution of fitness effects varies with phylogeny across animals
Kirk E. Lohmueller

California State University Northridge (USA), National Laboratory of Genomics for Biodiversity (Mexico), Pennsylvania State University (USA), Princeton University (USA), UC Berkeley (USA), UCLA (USA), UCSF (USA), University of Washington (USA)

The distribution of fitness effects (DFE) describes the selection coefficients \( (s) \) of newly arising mutations and fundamentally influences population genetics processes. Despite its importance, the extent of DFE variation in natural populations has not been systematically investigated across species with divergent phylogenetic histories and distinct ecological roles. Here, we inferred the DFE in natural populations of eleven animal (sub)species across mammals, birds, and insects using the site frequency spectrum (SFS). We find a significant phylogenetic signal in the DFE parameters, with the expected selection coefficient \( (E[s]) \) being more similar in more closely related species (Pagel’s lambda = 0.84, \( P = 0.04 \)). Additionally, mammals have a higher proportion of strongly deleterious mutations (22.4% to 47.4% in mammals; 0.0% to 16.0% in insects and birds) and a lower proportion of weakly deleterious mutations than insects and birds. Population size is negatively correlated with the average \( s \) of new deleterious mutations \( (P = 0.03) \). This result may imply a drift barrier to \( s \), similar to that for mutation rates. Specifically, larger populations experience more efficient positive selection for improved protein stability, allowing subsequent deleterious mutations to be only mildly deleterious. In smaller populations, however, selection is less efficient at increasing protein stability, and so subsequent mutations are more strongly deleterious. Several of our results are also consistent with Fisher’s geometric model. Overall, our study provides insights into the evolution of a fundamental population genetics parameter and suggests the evolutionary stability of the DFE across animal species.
Detecting barriers to gene flow from genomic patterns: application to maize and teosintes

Maud I Tenaillon

Maud I Tenaillon, Ewen Burban, Arthur Wojcik, Stella Huyn, Agnès Rousselet, Marie Couderc, Harry Belcram, Yves Vigouroux, Sylvain Glémin
AgroParisTech (France), CIRAD (France), CNRS, Department of Ecology and Genetics, ECOBIO-UMR 6553, Evolutionary Biology Centre (Sweden), GQE-Le Moulon, INRAE, IRD, UMR DIADE, UMR DIADE (France), UR1-CNRS (France), Univ Montpellier, Université Montpellier, Université Paris-Saclay

Domestication is the process of divergent selection between wild forms undergoing natural selection in their habitats, and domesticates evolving under combined natural and human-mediated selection. Domestication has been considered as a choice example to study adaptation. It also offers an excellent opportunity to catch the very-first processes at work in ecological speciation, where adaptive divergence between nascent lineages triggers the onset of reproductive isolation (RI). Here, we present RIDGE – Reproductive Isolation Detection using Genomic polymorphisms – a tool tailored for quantifying gene flow barrier proportion and identifying the relevant genomic regions. RIDGE relies on an Approximate Bayesian Computation with a model-averaging approach to accommodate diverse scenarios of lineage divergence. It captures heterogeneity in effective migration rate (m) along the genome while accounting for variation in linked selection and recombination, estimates barrier proportion (Q) and provides a test at the locus scale to detect gene flow barriers. Simulations and analyses of published datasets indicate that RIDGE is highly effective at both detecting ongoing migration and identifying barrier loci, even for recent divergence times (<0.1 2Ne generations). We applied RIDGE to 40 resequenced genomes from maize and teosintes. We estimated Q to 1.7% of the maize genome (49Mb). Interestingly, one locus known to be a barrier between maize and teosintes was detected. RIDGE also demonstrated its ability to distinguish barrier loci from domestication loci.
Single-cell Patterns and Evolutionary Mechanisms of Subgenome Expression Differentiation in Common Wheat
Wenfeng Qian

Institute of Genetics and Developmental Biology Chinese Academy of Sciences (China)

Polyplidization enhances plant adaptability and plasticity beyond their diploid ancestors, largely through subgenome expression diversification. In this study, we developed a single-cell expression atlas for common wheat seedlings (hexaploid) at the subgenomic level. We categorized genes into four groups based on their expression variations across cell types and subgenomes, observing varying evolutionary pressures. We explored the evolutionary drivers behind subgenome expression diversity by comparing our data with the single-cell transcriptomes of diploid or tetraploid ancestors. The expression patterns of homoeologous genes in common wheat closely mirrored those in its tetraploid ancestor. This study showcases the use of single-cell technology to elucidate the patterns and evolutionary processes of plant polyploidization.